

MIT Open Access Articles

Consistency of Probability Measure Quantization by Means of Power Repulsion–Attraction Potentials

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Fornasier, Massimo, and Jan-Christian Hütter. “Consistency of Probability Measure Quantization by Means of Power Repulsion–Attraction Potentials.” *Journal of Fourier Analysis and Applications* 22.3 (2016): 694–749.

As Published: <http://dx.doi.org/10.1007/s00041-015-9432-z>

Publisher: Springer US

Persistent URL: <http://hdl.handle.net/1721.1/104779>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



Consistency of Probability Measure Quantization by Means of Power Repulsion–Attraction Potentials

Massimo Fornasier¹ · Jan-Christian Hütter²

Received: 18 February 2015 / Revised: 25 July 2015 / Published online: 20 October 2015
© Springer Science+Business Media New York 2015

Abstract This paper is concerned with the study of the consistency of a variational method for probability measure quantization, deterministically realized by means of a minimizing principle, balancing power repulsion and attraction potentials. The proof of consistency is based on the construction of a target energy functional whose unique minimizer is actually the given probability measure ω to be quantized. Then we show that the discrete functionals, defining the discrete quantizers as their minimizers, actually Γ -converge to the target energy with respect to the narrow topology on the space of probability measures. A key ingredient is the reformulation of the target functional by means of a Fourier representation, which extends the characterization of conditionally positive semi-definite functions from points in generic position to probability measures. As a byproduct of the Fourier representation, we also obtain compactness of sublevels of the target energy in terms of uniform moment bounds, which already found applications in the asymptotic analysis of corresponding gradient flows. To model situations where the given probability is affected by noise, we further consider a modified energy, with the addition of a regularizing total variation term and we investigate again its point mass approximations in terms of Γ -convergence. We show that such a discrete measure representation of the total variation can be interpreted

Communicated by Hans G. Feichtinger.

✉ Jan-Christian Hütter
huetter@math.mit.edu

Massimo Fornasier
massimo.fornasier@ma.tum.de

¹ Fakultät Mathematik, Technische Universität München, Boltzmannstrasse 3,
85748 Garching bei München, Germany

² Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue,
Cambridge, MA 02139, USA

as an additional nonlinear potential, repulsive at a short range, attractive at a medium range, and at a long range not having effect, promoting a uniform distribution of the point masses.

Keywords Variational measure quantization · Fourier–Stieltjes transform · Total variation regularization · Gamma convergence

Mathematics Subject Classification 28A33 · 42B10 · 49J45 · 49M25 · 60E10 · 65D30 · 90C26

1 Introduction

1.1 Variational Measure Quantization and Main Results of the Paper

Quantization of d -dimensional probability measures deals with constructive methods to define atomic probability measures supported on a finite number of discrete points, which best approximate a given (diffuse) probability measure [24, 26]. Here the space of all probability measures is endowed with the Wasserstein or Kantorovich metric, which is usually the measure of the distortion of the approximation. The main motivations come from two classical relevant applications. The first we mention is information theory. In fact the problem of the quantization of a d -dimensional measure ω can be re-interpreted as the best approximation of a random d -dimensional vector X with distribution ω by means of a random vector Y which has at most N possible values in its image. This is a classical way of considering the digitalization of an analog signal, for the purpose of optimal data storage or parsimonious transmission of impulses via a channel. As we shall recall in more detail below, image dithering [33, 35] is a modern example of such an application in signal processing. The second classical application is numerical integration [30], where integrals with respect to certain probability measures need to be well-approximated by corresponding quadrature rules defined on the possibly optimal quantization points with respect to classes of continuous integrand functions. Numerical integration belongs to the standard problems of numerical analysis with numerous applications. It is often needed as a relevant subtask for solving more involved problems, for instance, the numerical approximation of solutions of partial differential equations. Additionally a number of problems in physics, e.g., in quantum physics, as well as any expectation in a variety of stochastic models require the computation of high-dimensional integrals as main (observable) quantities of interest. However, let us stress that the range of applications of measure quantization has nowadays become more far reaching, including mathematical models in economics (optimal location of service centers) or biology (optimal foraging and population distributions).

In absence of special structures of the underlying probability measure, for instance being well-approximated by finite sums of tensor products of lower dimensional measures, the problem of optimal quantization of measures, especially when defined on high-dimensional domains, can be hardly solved explicitly by deterministic methods. In fact, one may need to define optimal tiling of the space into Voronoi cells, based

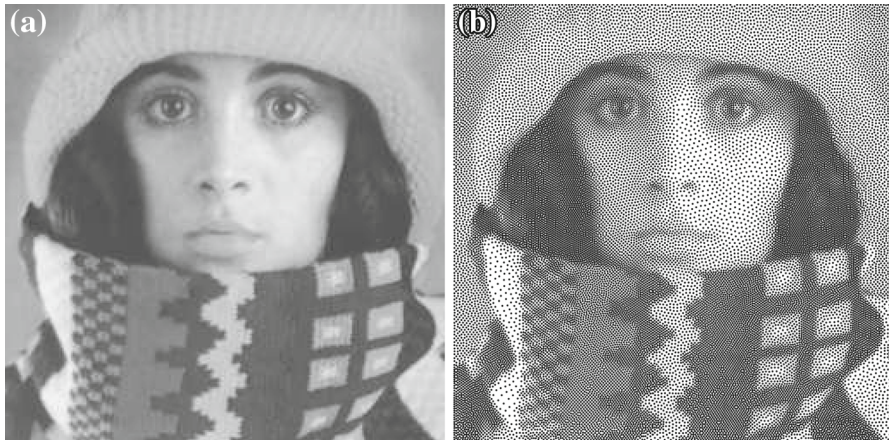


Fig. 1 Un-dithered and dithered image from [35] with the kind permission of the authors. **a** Original image. **b** Dithered image

again on testing the space by suitable high-dimensional integrations, see Sect. 5.2.1 below for an explicit deterministic construction of high-dimensional tilings for approximating probability measures by discrete measures. When the probability distribution can be empirically “tested”, by being able to draw at random samples from it, measure quantization can be realized by means of the empirical distribution. As a direct consequence of the Glivenko–Cantelli theorem, this way of generating empirical quantization points leads to the consistency of the approximation, in the sense of almost sure convergence of the empirical distribution to the original probability measure, as the number of draws goes to infinity, see Lemma 3.3 below recalling an explicit statement of this well-known result. Other results address also the approximation rate of such a randomized quantization, measuring the expected value of the Wasserstein distance between the empirical distributions and original probability measure, see for instance [13] and references therein. Unfortunately, in those situations where the probability distribution is given but it is too expensive or even impossible to be sampled, also the use of simple empirical distributions might not be viable. A concrete example of this situation is image dithering,¹ see Fig. 1. In this case the image represents the given probability distribution, which we actually can access, but it is evidently impossible to sample random draws from it, unless one designs actually a quantization of the image by means of deterministic methods, which again may lead us to tilings, and eventually making use of pseudorandom number generators.² To overcome this difficulty, a variational approach has been proposed in a series of papers [33, 35].

While there are many ways to determine the proximity of two probability measures (for a brief summary over some relevant alternatives, see [8]), the interesting idea orig-

¹ <http://en.wikipedia.org/wiki/Dither>.

² http://en.wikipedia.org/wiki/Pseudorandom_number_generator One practical way to sample randomly an image would be first to generate (pseudo-)randomly a finite number of points according to the uniform distribution from which one eliminates points which do not realize locally an integral over a prescribed threshold.

inally proposed in [33] consists in employing variational principles for that purpose. Namely, we consider the points $x = (x_i)_{i=1, \dots, N}$ to be attracted by the high-intensity locations of the black-and-white image, which represents our probability distribution ω , by introducing an attraction potential

$$\mathcal{V}_N(x) := \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^d} |x_i - y| d\omega(y)$$

which is to be minimized ($d = 2$ in case of image processing). If left as it is, the minimization of this term will most certainly not suffice to force the points into an intuitively good position, as the minimizer would consist of points accumulating on low-dimensional sets, for instance, in one dimension ($d = 1$), being all placed at the median of ω . Therefore, we shall enforce the spread of the points by adding a pairwise repulsion term

$$\mathcal{W}_N(x) := -\frac{1}{2N^2} \sum_{i,j=1}^N |x_i - x_j|,$$

leading to the minimization of the functional

$$\mathcal{E}_N(x) := \mathcal{V}_N(x) + \mathcal{W}_N(x), \tag{1.1}$$

which produces visually appealing results as in Fig. 1, see also [18] for alternative approaches towards image halftoning based on kernel estimations. By considering more general kernels $K_a(x_i, y)$ and $K_r(x_i, x_j)$ in the attraction and repulsion terms

$$\mathcal{E}_N^{K_a, K_r}(x) = \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^2} K_a(x_i, y) d\omega(y) - \frac{1}{2N^2} \sum_{i,j=1}^N K_r(x_i, x_j), \tag{1.2}$$

as already mentioned above, an attraction-repulsion functional of this type can easily be prone to other interesting interpretations. For instance, one could also consider the particles as a population subjected to attraction to a nourishment source ω , modeled by the attraction term, while at the same time being repulsed by internal competition. As one can see in the numerical experiments reported in [20, Sect. 4], the interplay of different powers of attraction and repulsion forces can lead some individuals of the population to fall out of the domain of the resource (food), which can be interpreted as an interesting mathematical model of social exclusion. The relationship between functionals of the type (1.2) and optimal numerical integration in reproducing kernel Hilbert spaces has been highlighted in [23], also showing once again the relevance of (deterministic) measure quantization towards designing efficient quadrature rules for numerical integration.

However, the generation of optimal quantization points by the minimization of functionals of the type (1.2) might also be subjected to criticism. First of all the functionals are in general nonconvex, rendering their global optimization, especially in high-dimension, a problem of high computational complexity, although, being the

functional the difference of two convex terms, numerical methods based on the alternation of gradient descent and gradient ascent iterations proved to be rather efficient in practice, see [35] for details. Especially one has to notice that for kernels generated by radial symmetric functions applied on the Euclidean distance of their arguments, the evaluation of the functional and of its subgradients results in the computation of convolutions which can be rapidly executed by non-equispaced fast Fourier transforms [19,31]. Hence, this technical advantage makes it definitively a promising alternative (for moderate dimensions d) with respect to deterministic methods towards optimal space tiling, based on local integrations and greedy placing, as it is for instance done in the strategy proposed in Sect. 5.2.1 below. Nevertheless, while for both empirical distributions and deterministic constructions consistency results are available, see for instance Lemmas 3.3 and 5.7 below, and the broad literature on these techniques [24], so far no similar results have been provided for discrete measures supported on optimal points generated as minimizers of functionals of the type (1.2), which leads us to the scope of this paper.

We shall prove that, for a certain type of kernels $K_a(x, y) = \psi_a(x - y)$ and $K_r(x, y) = \psi_r(x - y)$, where $\psi_a: \mathbb{R}^d \rightarrow \mathbb{R}_+$ and $\psi_r: \mathbb{R}^d \rightarrow \mathbb{R}_+$ are radially symmetric functions, the empirical measure $\mu_N = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$ constructed over points $x = (x_i)_{i=1, \dots, N}$ minimizing (1.2) converges narrowly to the given probability measure ω , showing the consistency of this quantization method. The strategy we intend to apply to achieve this result makes use of the so-called Γ -convergence [12], which is a type of variational convergence of sequences of functionals over separable metrizable spaces, which allows for simultaneous convergence of their respective minimizers. The idea is to construct a “target functional” \mathcal{E} whose unique minimizer is actually the given probability measure ω . Then one needs to show that the functionals $\mathcal{E}_N^{K_a, K_r}$ actually Γ -converge to \mathcal{E} for $N \rightarrow \infty$ with respect to the narrow topology on the space of probability measures, leading eventually to the convergence of the corresponding minimizers to ω . We immediately reveal that the candidate target functional for this purpose is, in the first instance, given by

$$\mathcal{E}[\mu] := \int_{\Omega \times \Omega} \psi_a(x - y) d\omega(x) d\mu(y) - \frac{1}{2} \int_{\Omega \times \Omega} \psi_r(x - y) d\mu(x) d\mu(y), \quad (1.3)$$

where we consider from now on a more general domain $\Omega \subset \mathbb{R}^d$ as well as measures $\mu, \omega \in \mathcal{P}(\Omega)$, where $\mathcal{P}(\Omega)$ is the space of probability measures. The reason for this natural choice comes immediately by observing that

$$\mathcal{E}[\mu_N] = \mathcal{E}_N^{K_a, K_r}(x), \text{ where } \mu_N = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}, \text{ and } x = (x_1, \dots, x_N).$$

For later use we denote

$$\mathcal{V}[\mu] := \int_{\Omega \times \Omega} \psi_a(x - y) d\omega(x) d\mu(y) \text{ and } \mathcal{W}[\mu] := -\frac{1}{2} \int_{\Omega \times \Omega} \psi_r(x - y) d\mu(x) d\mu(y).$$

However, this natural choice for a target limit functional poses several mathematical issues. First of all, as the functional is composed by the difference of two positive terms which might be simultaneously not finite over the set of probability measures, its well-posedness has to be justified. This will be done by restricting the class of radial symmetric functions ψ_a and ψ_r to those with at most quadratic polynomial growth and the domain of the functional to probability measures with bounded second moment. This solution, however, conflicts with the natural topology of the problem, which is the one induced by the narrow convergence. In fact, the resulting functional will not be necessarily lower semi-continuous with respect to the narrow convergence and this property is well-known to be necessary for a target functional to be a Γ -limit [12]. Thus, we need to extend the functional \mathcal{E} from the probability measures with bounded second moment $\mathcal{P}_2(\Omega)$ to the entire $\mathcal{P}(\Omega)$, by means of a functional $\widehat{\mathcal{E}}$ which is also lower semi-continuous with respect to the narrow topology. The first relevant result of this paper is to prove that such a lower semi-continuous relaxation $\widehat{\mathcal{E}}$ can be explicitly expressed, up to an additive constant term, for $\psi(\cdot) = \psi_a(\cdot) = \psi_r(\cdot) = |\cdot|^q$, and $1 \leq q < 2$, by means of the Fourier formula

$$\widehat{\mathcal{E}}[\mu] = -2^{-1}(2\pi)^{-d} \int_{\mathbb{R}^d} |\widehat{\mu}(\xi) - \widehat{\omega}(\xi)|^2 \widehat{\psi}(\xi) \, d\xi, \tag{1.4}$$

where for any $\mu \in \mathcal{P}(\mathbb{R}^d)$, $\widehat{\mu}$ denotes its Fourier–Stieltjes transform, and $\widehat{\psi}$ is the *generalized Fourier-transform of ψ* , i.e., a Fourier transform with respect to a certain duality, which allows to cancel the singularities of the Fourier transform of kernel ψ at 0. We have gathered most of the important facts about it in Appendix, recalling concisely the theory of conditionally positive semi-definite functions from [36]. The connection between functionals composed of repulsive and attractive power terms and Fourier type formulas (1.4) is novel and required to us to extend the theory of conditionally positive semi-definite functions from discrete points to probability measures. This crucial result is fundamental for proving as a consequence the well-posedness of the minimization of \mathcal{E} in $\mathcal{P}(\Omega)$ and the uniqueness of the minimizer ω , as it is now evident by the form (1.4), and eventually the Γ -convergence of the particle approximations. Another relevant result of this paper, which follows again from the Fourier representation, is the uniform r -th-moment bound for $r < \frac{q}{2}$ of the sublevels of $\widehat{\mathcal{E}}$ leading to their compactness in certain Wasserstein distances. This result plays a major role, for instance, in the analysis of the convergence to steady states of corresponding gradient flows (in dimension $d = 1$), which are studied in our follow up paper [15].

Another useful consequence of the Fourier representation is to allow us to add regularizations to the optimization problem. While for other quantization methods mentioned above, such as deterministic tiling and random draw of empirical distributions, it may be hard to filter the possible noise on the probability distribution, the variational approach based on the minimization of particle functionals of the type (1.1) is amenable to easy mechanisms of regularization. Differently from the path followed in the reasoning above, where we developed a limit from discrete to continuous functionals, here we proceed in the opposite direction, defining first the expected continuous regularized functional and then designing candidate discrete functional approximations, proving then the consistency again by Γ -convergence. One effective

way of filtering noise and still preserving the structure of the underlying measure ω is the addition to the discrepancy functional $\widehat{\mathcal{E}}$ of a term of *total variation*. This technique was introduced by Rudin, Osher, and Fatemi in the seminal paper [32] for the denoising of digital images, leading to a broad literature on variational methods over functions of bounded variations. We refer to [9, Chapter 4] for an introduction to the subject and to the references therein for a broad overview. Inspired by this well-established theory, we shall consider a regularization of \mathcal{E} by a total variation term,

$$\mathcal{E}^\lambda[\mu] := \widehat{\mathcal{E}}[\mu] + \lambda |D\mu|(\Omega), \quad (1.5)$$

where $\lambda > 0$ is a regularization parameter and μ is assumed to be in $L^1(\Omega)$, having distributional derivative $D\mu$ which is a finite Radon measure with total variation $|D\mu|$ [1]. Beside providing existence of minimizers of \mathcal{E}^λ in $\mathcal{P}(\Omega) \cap BV(\Omega)$ (here $BV(\Omega)$ denotes the space of bounded variation functions on Ω), and its Γ -convergence to $\widehat{\mathcal{E}}$ for $\lambda \rightarrow 0$, we also formulate particle approximations to \mathcal{E}^λ . While the approximation to the first term $\widehat{\mathcal{E}}$ is already given by its restriction to atomic measures, the consistent discretization in terms of point masses of the total variation term $|D\mu|(\Omega)$ is our last result of the present paper. By means of kernel estimators [37], we show in arbitrary dimensions that the total variation can be interpreted at the level of point masses as an additional attraction-repulsion potential, actually repulsive at a short range, attractive at a medium range, and at a long range not having effect, eventually tending to place the point masses into uniformly distributed configurations. We conclude with the proof of consistency of such a discretization by means of Γ -convergence. To our knowledge this interpretation of the total variation in terms of point masses has never been pointed out before in the literature.

1.2 Further Relevance to Other Work

Besides the aforementioned relationship to measure quantization in information theory, numerical integration, and the theory of conditionally positive semi-definite functions, energy functionals such as (1.3), being composed of a quadratic and a linear integral term, arise as well in a variety of mathematical models in biology and physics, describing the limit of corresponding particle descriptions. In particular the quadratic term, in our case denoted by \mathcal{W} , corresponding to the self-interaction between particles, has emerged in modeling biological aggregation. We refer to the survey paper [7] and the references therein for a summary on the mathematical results related to the mean-field limit of large ensembles of interacting particles with applications in swarming models, with particular emphasis on existence and uniqueness of aggregation gradient flow equations. We also mention that in direct connection to (1.3), in the follow up paper [15] we review the global well-posedness of gradient flow equations associated to the energy \mathcal{E} in one dimension, providing a simplified proof of existence and uniqueness, and we address the difficult problem of describing the asymptotic behavior of their solutions. In this respect we stress once more that the moment bounds derived in Sect. 4 of the present paper play a fundamental role for that analysis.

Although here derived as a model of regularization of the approximation process to a probability measure, also functionals like (1.5) with other kernels than polynomial growing ones appear in the literature in various contexts. The existence and characterization of their minimizers are in fact of great independent interest. When restricted to characteristic functions of finite perimeter sets, a functional of the type (1.5) with Coulombic-like repulsive interaction models the so-called non-local isoperimetric problem studied in [10,28,29]. Non-local Ginzburg-Landau energies modeling diblock polymer systems with kernels given by the Neumann Green's function of the Laplacian are studied in [21,22]. The power potential model studied in the present paper is contributing to this interesting constellation.

1.3 Structure of the Paper

In Sect. 2, we start by recalling a few theoretical preliminaries, followed by examples and counterexamples of the existence of minimizers of \mathcal{E} in the case of power potentials, depending on the powers and on the domain Ω , where elementary estimates of the behavior of the power functions are used in conjunction with appropriate notions of compactness for probability measures, i.e., uniform integrability of moments and moment bounds.

Starting from Sect. 3, we study the limiting case of coinciding powers of attraction and repulsion, where there is no longer an obvious confinement property given by the attraction term. To regain compactness and lower semi-continuity, we consider the lower semi-continuous envelope of the functional \mathcal{E} , which can be proven to coincide, up to an additive constant, with the Fourier representation (1.4), see Theorem 3.10 in Sect. 3.2, which is at first derived on $\mathcal{P}_2(\mathbb{R}^d)$ in Sect. 3.1. The main ingredient to find this representation is the generalized Fourier transform in the context of the theory of conditionally positive semi-definite functions, which we briefly recapitulated in Appendix.

Having thus established a problem which is well-posed for our purposes, we proceed to prove one of our main results, namely the convergence of the minimizers of the discrete functionals to ω , Theorem 3.13 in Sect. 3.3. This convergence will follow in a standard way from the Γ -convergence of the corresponding functionals. Furthermore, again applying the techniques of Appendix used to prove the Fourier representation allows us to derive compactness of the sublevels of \mathcal{E} in terms of a uniform moment bound in Sect. 4.

Afterwards, in Sect. 5, we shall introduce the total variation regularization of \mathcal{E} . Firstly, we prove consistency in terms of Γ -convergence for vanishing regularization parameter in Sect. 5.1. Then, in Sect. 5.2, we propose two ways of computing a version of it on particle approximations and again prove consistency for $N \rightarrow \infty$. One version consists of employing kernel density estimators, while, in the other one, each point mass is replaced by an indicator function extending up to the next point mass with the purpose of computing explicitly the total variation. In Sect. 6, we exemplify the Γ -limits of the first approach by numerical experiments.

1.4 More Insights for a More Efficient Reading

The present paper is a conciser version of the preprint appearing with the same title in arXiv at <http://arxiv.org/abs/1310.1120>, extracted from the first part of the Master thesis of Jan-Christian Hütter [27]. The second part of the thesis has been already published in the paper [15]. For the sake of self-containedness, in the preprint one can find included also the proofs of most of the auxiliary lemmas recalled from other papers or books, which are instead left here just as references in the present paper. Hence, the reader has the possibility either to read the references or just to access our preprint for additional insights. We kept within the present paper almost exclusively our original findings. Nevertheless, as we used several and diverse techniques from topological spaces of probability measures, harmonic analysis, and variational calculus, we made an effort to provide enough information for the paper to be read also by nonspecialists.

2 Preliminary Observations

2.1 Narrow Convergence and Wasserstein-Convergence

We begin with a brief summary of well-known measure theoretical results which will be needed in the following. Let $\Omega \subset \mathbb{R}^d$ be fixed and $\mathcal{P}_p(\Omega)$ denote the set of probability measures μ with finite p th-moment

$$\int_{\Omega} |x|^p d\mu(x) < \infty.$$

For an introduction to the narrow topology in spaces of probability measures $\mathcal{P}(\Omega)$, see [2, Chapter 5.1]. Let us only briefly recall a few relevant facts, which will turn out to be useful later on. First of all let us recall the definition of narrow convergence. A sequence of probability measures $(\mu_n)_{n \in \mathbb{N}}$ narrowly converges to $\mu \in \mathcal{P}(\Omega)$ if

$$\lim_{n \rightarrow \infty} \left| \int_{\Omega} g(x) d\mu_n(x) - \int_{\Omega} g(x) d\mu(x) \right| = 0, \text{ for all } g \in C_b(\Omega),$$

where $C_b(\Omega)$ is the space of bounded continuous functions on Ω . It is immediate to show that L^1 convergence of absolutely continuous probability measures in $\mathcal{P}(\Omega)$ implies narrow convergence. Moreover, as recalled in [2, Remark 5.1.1], there is a sequence of continuous functions $(f_k)_{k \in \mathbb{N}}$ on Ω and $\sup_{x \in \Omega} |f_k(x)| \leq 1$ such that the narrow convergence in $\mathcal{P}(\Omega)$ can be metrized by

$$\delta(\mu, \nu) := \sum_{k=1}^{\infty} 2^{-k} \left| \int_{\Omega} f_k(x) d\mu(x) - \int_{\Omega} f_k(x) d\nu(x) \right|. \quad (2.1)$$

The metrizability of $\mathcal{P}(\Omega)$ endowed with the narrow topology is an important technical ingredient, which we will need at several places along the paper, in particular, in the use of Γ -convergence.

It will turn out to be useful also to observe that narrow convergences extends to tensor products. From [5, Theorem 2.8] it follows that if $(\mu_n)_n, (\nu_n)_n$ are two sequences in $\mathcal{P}(\Omega)$ and $\mu, \nu \in \mathcal{P}(\Omega)$, then

$$\mu_n \otimes \nu_n \rightarrow \mu \otimes \nu \text{ narrowly if and only if } \mu_n \rightarrow \mu \text{ and } \nu_n \rightarrow \nu \text{ narrowly.} \quad (2.2)$$

Finally, we include some results about the continuity of integral functionals with respect to Wasserstein-convergence.

Definition 2.1 (*Wasserstein distance*) Let $\Omega \subset \mathbb{R}^d, p \in [1, \infty)$ as well as $\mu_1, \mu_2 \in \mathcal{P}_p(\Omega)$ be two probability measures with finite p th moment. Denoting by $\Gamma(\mu_1, \mu_2)$ the probability measures on $\Omega \times \Omega$ with marginals μ_1 and μ_2 , then we define

$$W_p^p(\mu_1, \mu_2) := \min \left\{ \int_{\Omega^2} |x_1 - x_2|^p d\mu(x_1, x_2) : \mu \in \Gamma(\mu_1, \mu_2) \right\}, \quad (2.3)$$

the *Wasserstein- p* distance between μ_1 and μ_2 .

Definition 2.2 (*Uniform integrability*) A measurable function $f : \Omega \rightarrow [0, \infty]$ is *uniformly integrable* with respect to a family of finite measures $\{\mu_i : i \in I\}$, if

$$\lim_{M \rightarrow \infty} \sup_{i \in I} \int_{\{f(x) \geq M\}} f(x) d\mu_i(x) = 0.$$

In particular, we say that a set $\mathcal{K} \subset \mathcal{P}_p(\Omega)$ has uniformly integrable p th-moments if

$$\lim_{M \rightarrow \infty} \sup_{\mu \in \mathcal{K}} \int_{|x| \geq M} |x|^p d\mu(x) = \infty.$$

Lemma 2.3 (Topology of Wasserstein spaces) [2, Proposition 7.1.5] For $p \geq 1$ and a subset $\Omega \subset \mathbb{R}^d, \mathcal{P}_p(\Omega)$ endowed with the Wasserstein- p distance is a separable metric space which is complete if Ω is closed. A set $\mathcal{K} \subset \mathcal{P}_p(\Omega)$ is relatively compact if and only if it has uniformly integrable p th-moments (and hence tight by Lemma 2.5 just below). In particular, for a sequence $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}_p(\Omega)$, the following properties are equivalent:

- (i) $\lim_{n \rightarrow \infty} W_p(\mu_n, \mu) = 0$;
- (ii) $\mu_n \rightarrow \mu$ narrowly and $(\mu_n)_n$ has uniformly integrable p th-moments.

Lemma 2.4 (Continuity of integral functionals) [2, Lemma 5.1.7] Let $(\mu_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{P}(\Omega)$ converging narrowly to $\mu \in \mathcal{P}(\Omega)$, $g : \Omega \rightarrow \mathbb{R}$ lower semi-continuous and $f : \Omega \rightarrow \mathbb{R}$ continuous. If $|f|, g^- := -\min\{g, 0\}$ are uniformly integrable with respect to $\{\mu_n : n \in \mathbb{N}\}$, then

$$\begin{aligned} \liminf_{n \rightarrow \infty} \int_{\Omega} g(x) d\mu_n(x) &\geq \int_{\Omega} g(x) d\mu(x) \\ \lim_{n \rightarrow \infty} \int_{\Omega} f(x) d\mu_n(x) &= \int_{\Omega} f(x) d\mu(x) \end{aligned}$$

Lemma 2.5 (Uniform integrability of moments) [6, Corollary to Theorem 25.12] Given $r > 0$ and a family $\{\mu_i : i \in I\}$ of probability measures in $\mathcal{P}(\Omega)$ with

$$\sup_{i \in I} \int_{\Omega} |x|^r \, d\mu_i(x) < \infty,$$

then the family $\{\mu_i : i \in I\}$ is tight, i.e.,

$$\lim_{M \rightarrow \infty} \sup_{i \in I} \mu_i(\{|x| \geq M\}) = 0.$$

and $x \mapsto |x|^q$ is uniformly integrable with respect to $\{\mu_i : i \in I\}$ for all $0 < q < r$.

2.2 Examples and Counterexamples to Existence of Minimizers for Discordant Powers $q_a \neq q_r$

We recall the definition of \mathcal{E} :

$$\mathcal{E}[\mu] := \int_{\Omega \times \Omega} \psi_a(x - y) \, d\omega(x) \, d\mu(y) - \frac{1}{2} \int_{\Omega \times \Omega} \psi_r(x - y) \, d\mu(x) \, d\mu(y),$$

for $\omega, \mu \in \mathcal{P}_2(\Omega)$ (at least for now) and

$$\psi_a(x) := |x|^{q_a}, \quad \psi_r(x) := |x|^{q_r}, \quad x \in \mathbb{R}^d,$$

where $q_a, q_r \in [1, 2]$. Furthermore, denote for a vector-valued measure ν its total variation by $|\nu|$ and by $BV(\Omega)$ the space of functions $f \in L^1_{\text{loc}}(\Omega)$ whose distributional derivatives Df are finite Radon measures. With some abuse of terminology, we call $|Df|(\Omega)$ the total variation of f as well. Now, we define the total variation regularization of \mathcal{E} by

$$\mathcal{E}^\lambda[\mu] := \mathcal{E}[\mu] + \lambda |D\mu|(\Omega),$$

where $\mu \in \mathcal{P}_2(\Omega) \cap BV(\Omega)$.

We shall briefly state some results which are in particular related to the asymmetric case of q_a and q_r not necessarily being equal.

2.2.1 Situation on a Compact Set

Proposition 2.6 Let $q_a \geq 1, q_r \geq 1$ and Ω be a compact subset of \mathbb{R}^d . The functionals \mathcal{E} and \mathcal{E}^λ are well-defined on $\mathcal{P}(\Omega)$ and $\mathcal{P}(\Omega) \cap BV(\Omega)$, respectively. Moreover \mathcal{E} has minimizers, and, if additionally Ω is an extension domain (see [1] for a precise definition of extension domain for bounded variation functions; e.g., a domain with Lipschitz boundary suffices), then \mathcal{E}^λ also admits a minimizer.

Proof Note that since the mapping

$$(x, y) \mapsto |y - x|^q, \quad x, y \in \mathbb{R}^d, \tag{2.4}$$

is jointly continuous in x and y , it attains its maximum on the compact set $\Omega \times \Omega$. Hence, the kernel (2.4) is a bounded continuous function, which, on the one hand, implies that the functional \mathcal{E} is bounded (and in particular well-defined) on $L^1(\Omega)$ and on the other hand that it is continuous with respect to the narrow topology. Together with the compactness of $\mathcal{P}(\Omega)$, this implies existence of a minimizer for \mathcal{E} .

The situation for \mathcal{E}^λ is similar. Due to the boundedness of Ω and the regularity of its boundary, sub-levels of $|D \cdot |(\Omega)$ are relatively compact in $L^1(\Omega) \cap \mathcal{P}(\Omega)$ by [17, Chapter 5.2, Theorem 4]. As the total variation is lower semi-continuous with respect to L^1 -convergence by [17, Chapter 5.2, Theorem 1] and L^1 -convergence implies narrow convergence, we get lower semi-continuity of \mathcal{E}^λ and therefore again existence of a minimizer. \square

2.2.2 Existence of Minimizers for Stronger Attraction on Arbitrary Domains

Note that from here on, the constants C and c are generic and may change in each line of a calculation. In the following we shall make use of the following elementary inequalities: for $q \geq 1$ and $x, y \in \mathbb{R}^d$, there exist $C, c > 0$ such that

$$|x + y|^q \leq C (|x|^q + |y|^q), \tag{2.5}$$

and

$$|x - y|^q \geq (|x|^q - c|y|^q). \tag{2.6}$$

Theorem 2.7 *Let $q_a, q_r \in [1, 2]$, $\Omega \subset \mathbb{R}^d$ closed and $q_a > q_r$. If $\omega \in \mathcal{P}_{q_a}(\Omega)$, then the sub-levels of \mathcal{E} have uniformly bounded q_a th moments and \mathcal{E} admits a minimizer in $\mathcal{P}_{q_r}(\Omega)$.*

Proof Moment bound: Let $\mu \in \mathcal{P}_{q_r}(\Omega)$. By estimate (2.6), we have

$$\begin{aligned} \mathcal{V}[\mu] &= \int_{\Omega \times \Omega} |x - y|^{q_a} d\mu(x)d\omega(y) \geq \int_{\Omega \times \Omega} (|x|^{q_a} - c|y|^{q_a}) d\mu(x)d\omega(x) \\ &= \int_{\Omega} |x|^{q_a} d\mu(x) - c \int_{\Omega} |y|^{q_a} d\omega(y). \end{aligned} \tag{2.7}$$

On the other hand, by estimate (2.5) we obtain

$$\mathcal{W}[\mu] = -\frac{1}{2} \int_{\Omega \times \Omega} |x - y|^{q_r} d\mu(x)d\mu(y) \geq -C \int_{\Omega} |x|^{q_r} d\mu(x). \tag{2.8}$$

Combining (2.7) and (2.8), we obtain

$$\mathcal{E}[\mu] + c \int_{\Omega} |x|^{q_a} d\omega(x) \geq \int_{\Omega} (1 - C|x|^{q_r - q_a}) |x|^{q_a} d\mu(x).$$

Since $q_a > q_r$, there is an $M > 0$ such that $1 - C |x|^{q_r - q_a} \geq \frac{1}{2}$, $|x| \geq M$, and hence

$$\begin{aligned} \int_{\Omega} |x|^{q_a} d\mu(x) &= \left[\int_{B_M(0)} |x|^{q_a} d\mu(x) + \int_{\Omega \setminus B_M(0)} |x|^{q_a} d\mu(x) \right] \\ &\leq M^{q_a} + 2 \left[\mathcal{E}[\mu] + c \int_{\Omega} |x|^{q_a} d\omega(x) \right]. \end{aligned} \tag{2.9}$$

These estimates show that the sub-levels of \mathcal{E} have a uniformly bounded q_a th moment, so that they are also Wasserstein- q compact for any $q < q_a$ by Lemmas 2.3 and 2.5. Given a minimizing sequence, we can extract a narrowly converging subsequence $(\mu_n)_n$ with uniformly integrable q_r th moments. With respect to that convergence, which also implies the narrow convergence of $(\mu_n \otimes \mu_n)_n$ and $(\mu_n \otimes \omega)_n$ by (2.2), the functional \mathcal{W} is continuous and the functional \mathcal{V} is lower semi-continuous by Lemma 2.4, so we shall be able to apply the direct method of calculus of variations to show existence of a minimizer in $\mathcal{P}_{q_r}(\Omega)$. \square

2.2.3 Counterexample to the Existence of Minimizers for Stronger Repulsion

Now, let $q_a, q_r \in [1, 2]$ with $q_r > q_a$. On $\Omega = \mathbb{R}^d$, the minimization of \mathcal{E} and \mathcal{E}^λ need not have a solution.

Example 2.8 (Nonexistence of minimizers for stronger repulsion) Let $\Omega = \mathbb{R}$, $q_r > q_a$, $\omega = \mathcal{L}^1 \llcorner [-1, 0]$ be the one dimensional Lebesgue measure restricted to $[-1, 0]$ and consider the sequence $\mu_n := n^{-1} \mathcal{L}^1 \llcorner [0, n]$. Computing the values of the functionals used to define \mathcal{E} and \mathcal{E}^λ yields

$$\mathcal{V}[\mu_n] = \frac{1}{n} \int_{-1}^0 \int_0^n |y - x|^{q_a} dx dy \leq \frac{1}{n} \int_0^n (y + 1)^{q_a} dy \leq \frac{(n + 1)^{q_a}}{q_a + 1}.$$

Again by direct integration we have

$$\mathcal{W}[\mu_n] = -\frac{1}{2n^2} \int_0^n \int_0^n |y - x|^{q_r} dx dy = \frac{n^{q_r}}{(q_r + 1)(q_r + 2)},$$

and

$$\|D\mu_n\| = -\frac{2}{n}.$$

By considering the limit of the corresponding sums, we obtain

$$\mathcal{E}[\mu_n] \rightarrow -\infty, \quad \mathcal{E}^\lambda[\mu_n] \rightarrow -\infty \quad \text{for } n \rightarrow \infty,$$

meaning that there are no minimizers in this case.

3 Properties of the Functional on \mathbb{R}^d

Now, let us consider $\Omega = \mathbb{R}^d$ and

$$q := q_a = q_r, \quad \psi(x) := \psi_a(x) = \psi_r(x) = |x|^q, \quad x \in \mathbb{R}^d, \tag{3.1}$$

for $1 \leq q < 2$.

Here, neither the well-definedness of $\mathcal{E}[\mu]$ for all $\mu \in \mathcal{P}(\mathbb{R}^d)$ nor the narrow compactness of the sub-levels as in the case of a compact Ω in Sect. 2.2.1 are clear, necessitating additional conditions on μ and ω . For example, if we assume the finiteness of the second moments, i.e., $\mu, \omega \in \mathcal{P}_2(\mathbb{R}^d)$, we can a priori see that both $\mathcal{V}[\mu]$ and $\mathcal{W}[\mu]$ are finite.

Under this restriction, we shall show a formula for \mathcal{E} involving the Fourier–Stieltjes transform of the measures μ and ω . Namely, there is a constant $C = C(q, \omega) \in \mathbb{R}$ such that

$$\mathcal{E}[\mu] + C = -2^{-1}(2\pi)^{-d} \int_{\mathbb{R}^d} |\widehat{\mu}(\xi) - \widehat{\omega}(\xi)|^2 \widehat{\psi}(\xi) \, d\xi =: \widehat{\mathcal{E}}[\mu], \tag{3.2}$$

where for any $\mu \in \mathcal{P}(\mathbb{R}^d)$, $\widehat{\mu}$ denotes its Fourier–Stieltjes transform,

$$\widehat{\mu}(\xi) = \int_{\mathbb{R}^d} \exp(-ix^T \xi) \, d\mu(x), \tag{3.3}$$

and $\widehat{\psi}$ is the *generalized Fourier-transform of ψ* , i.e., a Fourier transform with respect to a certain duality, which allows to cancel the singularities of the Fourier transform of the kernel ψ at 0. We have gathered most of the important facts about it in Appendix. In the case of ψ as in (3.1), such Fourier transform can be explicitly computed to be

$$\widehat{\psi}(\xi) := -2 \cdot (2\pi)^d D_q |\xi|^{-d-q}, \tag{3.4}$$

where

$$D_q := -(2\pi)^{-d/2} \frac{2^{q+d/2} \Gamma((d+q)/2)}{2\Gamma(-q/2)} > 0, \tag{3.5}$$

so that

$$\widehat{\mathcal{E}}[\mu] = D_q \int_{\mathbb{R}^d} |\widehat{\mu}(\xi) - \widehat{\omega}(\xi)|^2 |\xi|^{-d-q} \, d\xi, \tag{3.6}$$

which will be proved in Sect. 3.1.

Notice that, while \mathcal{E} might not be well-defined on $\mathcal{P}(\mathbb{R}^d)$, formula (3.6) makes sense on the whole space $\mathcal{P}(\mathbb{R}^d)$. The sub-levels of $\widehat{\mathcal{E}}$ can be proved to be narrowly compact as well as lower semi-continuous with respect to the narrow topology (see Proposition 3.8), motivating the proof in Sect. 3.2 that up to a constant, this formula is exactly the lower semi-continuous envelope of \mathcal{E} on $\mathcal{P}(\mathbb{R}^d)$ endowed with the narrow topology.

3.1 Fourier Formula in $\mathcal{P}_2(\mathbb{R}^d)$

Assume that $\mu, \omega \in \mathcal{P}_2(\mathbb{R}^d)$ and observe that by using the symmetry of ψ , $\mathcal{E}[\mu]$ can be written as

$$\mathcal{E}[\mu] = -\frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} \psi(y - x) d[\mu - \omega](x) d[\mu - \omega](y) + C, \tag{3.7}$$

where

$$C = \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} \psi(y - x) d\omega(x) d\omega(y). \tag{3.8}$$

In the following, we shall mostly work with this symmetrized variant, resulting by neglecting the additive constant, and denoted it by

$$\tilde{\mathcal{E}}[\mu] := -\frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} \psi(y - x) d[\mu - \omega](x) d[\mu - \omega](y). \tag{3.9}$$

3.1.1 Representation for Point-Measures

Our starting point is a Fourier-type representation of $\tilde{\mathcal{E}}$, whenever μ and ω are atomic measures. Such formula follows from Theorem 7.4 together with the explicit computation of the generalized Fourier transform of ψ in Theorem 7.6, as concisely recalled from [36] in the Appendix.

Lemma 3.1 *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\omega \in \mathcal{P}(\mathbb{R}^d)$ be linear combinations of Dirac measures so that*

$$\mu - \omega = \sum_{j=1}^N \alpha_j \delta_{x_j},$$

for a suitable $N \in \mathbb{N}$, $\alpha_j \in \mathbb{R}$, and pairwise distinct $x_j \in \mathbb{R}^d$ for all $j = 1, \dots, N$. Then

$$\tilde{\mathcal{E}}[\mu] = -2^{-1} (2\pi)^{-d} \int_{\mathbb{R}^d} \left| \sum_{j=1}^N \alpha_j \exp(ix_j^T \xi) \right|^2 \widehat{\psi}(\xi) d\xi, \tag{3.10}$$

where

$$\widehat{\psi}(\xi) := -2 \cdot (2\pi)^d D_q |\xi|^{-d-q}, \quad \text{with a } D_q > 0.$$

Remark 3.2 By $\overline{\exp(ix)} = \exp(-ix)$, for $x \in \mathbb{R}$, we can also write the above formula (3.10) as

$$\tilde{\mathcal{E}}[\mu] = D_q \int_{\mathbb{R}^d} |\widehat{\mu}(\xi) - \widehat{\omega}(\xi)|^2 |\xi|^{-d-q} d\xi, \quad \xi \in \mathbb{R}^d.$$

3.1.2 Point Approximation of Probability Measures by the Empirical Distribution

Lemma 3.3 (Consistency of empirical distribution) *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $(X_i)_{i \in \mathbb{N}}$ be a sequence of i.i.d. random variables with $X_i \sim \mu$ for all $i \in \mathbb{N}$. Then the empirical distribution*

$$\mu_N := \frac{1}{N} \sum_{i=1}^N \delta_{X_i},$$

where δ_X is the Dirac delta supported on the point $X \in \mathbb{R}^d$, converges with probability 1 narrowly to μ , i.e.,

$$P(\{\mu_N \rightarrow \mu \text{ narrowly}\}) = 1.$$

Additionally, if for a $p \in [1, \infty)$, $\int_{\mathbb{R}^d} |x|^p \, d\mu < \infty$, then $x \mapsto |x|^p$ is almost surely uniformly integrable with respect to $\{\mu_N : N \in \mathbb{N}\}$, which by Lemma 2.3 implies almost sure convergence of $\mu_N \rightarrow \mu$ in the p -Wasserstein topology.

Proof The almost sure narrow convergence of the empirical distribution is a consequence of the well-known Glivenko–Cantelli theorem [34]. As the second claim is less known, we recall a short proof of it. We apply the strong law of large numbers to the functions $f_M(x) := |x|^p \cdot 1_{\{|x|^p \geq M\}}$ for $M > 0$ to get the desired uniform integrability: for a given $\varepsilon > 0$, choose $M > 0$ large enough such that $\int_{\mathbb{R}^d} f_M(x) \, d\mu(x) < \frac{\varepsilon}{2}$, and then $N_0 \in \mathbb{N}$ large enough such that

$$\left| \int_{\mathbb{R}^d} f_M(x) \, d\mu_N(x) - \int_{\mathbb{R}^d} f_M(x) \, d\mu(x) \right| < \frac{\varepsilon}{2}, \quad N \geq N_0, \text{ almost surely.}$$

Now we choose $M' \geq M$ sufficiently large so to ensure that $|X_i|^p < M'$ almost surely for all $i < N_0$. By the monotonicity of $\int_{\mathbb{R}^d} f_M(x) \, d\mu(x)$ in M , this ensures

$$\begin{aligned} \sup_{N \in \mathbb{N}} \int_{\mathbb{R}^d} f_{M'}(x) \, d\mu_N &= \sup_{N \geq N_0} \int_{\mathbb{R}^d} f_{M'}(x) \, d\mu_N \leq \sup_{N \geq N_0} \int_{\mathbb{R}^d} f_M(x) \, d\mu_N(x) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

□

3.1.3 The Fourier Representation of $\tilde{\mathcal{E}}$ on $\mathcal{P}_2(\mathbb{R}^d)$

Now we establish continuity in both sides of (3.10) with respect to the 2-Wasserstein-convergence to obtain (3.2) in $\mathcal{P}_2(\mathbb{R}^d)$.

Lemma 3.4 (Continuity of $\tilde{\mathcal{E}}$) *Let*

$$\mu_k \rightarrow \mu, \quad \omega_k \rightarrow \omega \text{ for } k \rightarrow \infty \text{ in } \mathcal{P}_2(\mathbb{R}^d),$$

with respect to the 2-Wasserstein-convergence. Then,

$$\begin{aligned} &\int_{\mathbb{R}^d \times \mathbb{R}^d} \psi(y-x) \, d[\mu_k - \omega_k](x) \, d[\mu_k - \omega_k](y) \\ &\rightarrow \int_{\mathbb{R}^d \times \mathbb{R}^d} \psi(y-x) \, d[\mu - \omega](x) \, d[\mu - \omega](y), \text{ for } k \rightarrow \infty. \end{aligned} \tag{3.11}$$

Proof By the particular choice of ψ , we have the estimate

$$|\psi(y - x)| \leq C(1 + |y - x|^2) \leq 2C(1 + |x|^2 + |y|^2).$$

After expanding the expression to the left of (3.11) so that we only have to deal with integrals with respect to probability measures, we can use this estimate to get the uniform integrability of the second moments of μ and ω by Lemma 2.3 and are then able to apply Lemma 2.4 to obtain convergence. \square

Lemma 3.5 (Continuity of $\widehat{\mathcal{E}}$) *Let*

$$\mu_k \rightarrow \mu, \quad \omega_k \rightarrow \omega \quad \text{for } k \rightarrow \infty \text{ in } \mathcal{P}_2(\mathbb{R}^d),$$

with respect to the 2-Wasserstein-convergence, such that

$$\mu_k - \omega_k = \sum_{j=1}^{N_k} \alpha_j^k \delta_{x_j^k}$$

for suitable $\alpha_j^k \in \mathbb{R}$ and pairwise distinct $x_j^k \in \mathbb{R}^d$. Then,

$$\begin{aligned} & \int_{\mathbb{R}^d} \left| \sum_{j=1}^{N_k} \alpha_j^k \exp(i\xi \cdot x_j^k) \right|^2 \widehat{\psi}(\xi) \, d\xi \\ & \rightarrow \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \exp(i\xi \cdot x) \, d[\mu - \omega](x) \right|^2 \widehat{\psi}(\xi) \, d\xi \quad \text{for } k \rightarrow \infty. \end{aligned}$$

Proof By the narrow convergence of μ_k and ω_k , we get pointwise convergence of the Fourier transforms, i.e.,

$$\sum_{j=1}^{N_k} \alpha_j^k \exp(i\xi \cdot x_j^k) \rightarrow \int_{\mathbb{R}^d} \exp(i\xi \cdot x) \, d[\mu - \omega](x) \quad \text{for all } \xi \in \mathbb{R}^d \text{ and } k \rightarrow \infty.$$

For the convergence of the integrals we want to use now the dominated convergence: The Fourier transform of $\mu - \omega$ is bounded in ξ , so that the case $\xi \rightarrow \infty$ poses no problem due to the integrability of $\widehat{\psi}(\xi) = C |\xi|^{-d-q}$ away from 0. In order to justify the necessary decay at 0, we use the control of the first moments (since we even control the second moments by the \mathcal{P}_2 assumption): Inserting the Taylor expansion of the exponential function of order 0,

$$\exp(i\xi \cdot x) = 1 + i\xi \cdot x \int_0^1 \exp(i\xi \cdot tx) \, dt,$$

into the integrand and using the fact that μ_k and ω_k are probability measures results in

$$\begin{aligned}
 & \left| \int_{\mathbb{R}^d} \exp(i\xi \cdot x) \, d[\mu_k - \omega_k](x) \right| \\
 &= \left| \int_{\mathbb{R}^d} \left(1 + i\xi \cdot x \int_0^1 \exp(i\xi \cdot tx) \, dt \right) \, d[\mu_k - \omega_k](x) \right| \\
 &= \left| \int_{\mathbb{R}^d} \left(i\xi \cdot x \int_0^1 \exp(i\xi \cdot tx) \, dt \right) \, d[\mu_k - \omega_k](x) \right| \\
 &\leq |\xi| \underbrace{\left(\int_{\mathbb{R}^d} |x| \, d\mu_k(x) + \int_{\mathbb{R}^d} |x| \, d\omega_k(x) \right)}_{:=C}.
 \end{aligned}$$

Therefore, we have uniform bound C with respect to k , so that

$$\left| \sum_{j=1}^{N_k} \alpha_j^k \exp(i\xi \cdot x_j^k) \right|^2 \leq C |\xi|^2,$$

compensating the singularity of $\widehat{\psi}$ at the origin. Hence together with the dominated convergence theorem proving the claim. \square

The combination of the two Lemmata above with the approximation provided by the empirical distribution of Lemma 3.3 directly yields the extension of (3.10) to $\mathcal{P}_2(\mathbb{R}^d)$.

Theorem 3.6 (Fourier-representation for $\widetilde{\mathcal{E}}$ on $\mathcal{P}_2(\mathbb{R}^d)$)

$$\widetilde{\mathcal{E}}[\mu] = \widehat{\mathcal{E}}[\mu], \quad \mu \in \mathcal{P}_2(\mathbb{R}^d).$$

3.2 Extension to $\mathcal{P}(\mathbb{R}^d)$

While the well-definedness of $\mathcal{E}[\mu]$ is not clear for all $\mu \in \mathcal{P}(\mathbb{R}^d)$, since the difference of two integrals with values $+\infty$ may occur instead, for each such μ we can certainly assign a value in $\mathbb{R} \cup \{\infty\}$ to $\widehat{\mathcal{E}}[\mu]$. In the following, we want to justify in which sense it is possible to consider $\widehat{\mathcal{E}}$ as an extension of \mathcal{E} , namely that $\widehat{\mathcal{E}}$ is, up to an additive constant, the *lower semi-continuous envelope* of \mathcal{E} .

Firstly, we prove that $\widehat{\mathcal{E}}$ has compact sub-levels in $\mathcal{P}(\mathbb{R}^d)$ endowed with the narrow topology, using the following lemma as a main ingredient.

Lemma 3.7 *Given a probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ with Fourier transform $\widehat{\mu}: \mathbb{R}^d \rightarrow \mathbb{C}$, there are $C_1 = C_1(d) > 0$ and $C_2 = C_2(d) > 0$ such that for all $u > 0$,*

$$\mu \left(\left\{ x : |x| \geq u^{-1} \right\} \right) \leq \frac{C_1}{u^d} \int_{|\xi| \leq C_2 u} (1 - \operatorname{Re} \widehat{\mu}(\xi)) \, d\xi. \tag{3.12}$$

Proof The proof for the case $d = 1$ can be found in [16, Theorem 3.3.6] and we generalize it below to any $d \geq 1$. Let $u > 0$. Firstly, note that

$$1 - \operatorname{Re} \widehat{\mu}(\xi) = \int_{\mathbb{R}^d} (1 - \cos(\xi \cdot x)) \, d\mu(x) \geq 0 \quad \text{for all } \xi \in \mathbb{R}^d.$$

By starting with the integral on the right-hand side of (3.12) (up to a constant in the integration domain) and using Fubini-Tonelli as well as integration in spherical coordinates, we get

$$\begin{aligned} & \int_{|\xi| \leq u} (1 - \operatorname{Re} \widehat{\mu}(\xi)) \, d\xi \\ &= \int_{\mathbb{R}^d} \int_{|\xi| \leq u} (1 - \cos(\xi \cdot x)) \, d\xi \, d\mu(x) \\ &= \int_{\mathbb{R}^d} \int_{|\tilde{\xi}|=1} \int_0^u (1 - \cos(r\tilde{\xi} \cdot x)) r^{d-1} \, dr \, d\sigma(\tilde{\xi}) \, d\mu(x) \end{aligned} \tag{3.13}$$

$$= \int_{\mathbb{R}^d} \int_{|\tilde{\xi}|=1} \left[\frac{u^d}{d} - \int_0^u \cos(r\tilde{\xi} \cdot x) r^{d-1} \, dr \right] \, d\sigma(\tilde{\xi}) \, d\mu(x) \tag{3.14}$$

If $d \geq 2$, integrating the integral over $\cos(r\tilde{\xi} \cdot x) r^{d-1}$ in (3.14) by parts yields

$$\int_0^u \cos(r\tilde{\xi} \cdot x) r^{d-1} \, dr = \frac{\sin(u\tilde{\xi} \cdot x) u^{d-1}}{\tilde{\xi} \cdot x} - (d-1) \int_0^u \frac{\sin(r\tilde{\xi} \cdot x)}{\tilde{\xi} \cdot x} r^{d-2} \, dr,$$

which can also be considered true for $d = 1$ if the second part is assumed to be zero because of the factor $(d - 1)$.

We now prove (3.12) by estimating the integrand in (3.14) suitably from below. Using $|\sin(x)| \leq 1$ for all $x \in \mathbb{R}$ and dividing by u^d , we get

$$\begin{aligned} & d^{-1} - u^{-d} \int_0^u \cos(r\tilde{\xi} \cdot x) r^{d-1} \, dr \\ &= d^{-1} - \frac{\sin(u\tilde{\xi} \cdot x)}{u\tilde{\xi} \cdot x} + \frac{(d-1)}{u^d} \int_0^u \frac{\sin(r\tilde{\xi} \cdot x)}{\tilde{\xi} \cdot x} r^{d-2} \, dr \\ &\geq d^{-1} - \frac{1}{u|\tilde{\xi} \cdot x|} - \frac{(d-1)}{u^d} \int_0^u \frac{1}{|\tilde{\xi} \cdot x|} r^{d-2} \, dr \\ &= d^{-1} - \frac{2}{u|\tilde{\xi} \cdot x|}. \end{aligned}$$

As we want to achieve an estimate from below, by the non-negativity of the integrand $1 - \cos(\xi \cdot x)$, we can restrict the integration domain in (3.13) to

$$\widetilde{S}(x) := \left\{ \tilde{\xi} \in S^{d-1} : |\tilde{\xi} \cdot x| \geq \frac{1}{2} |x| \right\} \quad \text{and} \quad D(u) := \left\{ x : |x| \geq \frac{8d}{u} \right\},$$

yielding

$$\frac{1}{d} - \frac{1}{u^d} \int_0^u \cos(r\tilde{\xi} \cdot x)r^{d-1} dr \geq \frac{1}{2d}, \quad x \in D(u), \tilde{\xi} \in \tilde{S}(x). \tag{3.15}$$

Combining (3.15) with (3.14) gives us

$$\frac{1}{u^d} \int_{|\xi| \leq u} (1 - \operatorname{Re}\widehat{\mu}(\xi)) d\xi \geq \frac{1}{C_3} \mu \left(\left\{ |x| \geq 8D's^{-1} \right\} \right)$$

with

$$C_3 := \frac{1}{2d} \operatorname{vol}(\tilde{S}(x)),$$

where $\operatorname{vol}(\tilde{S}(x))$ is independent of x . Finally, we substitute $\tilde{u} := (8d)^{-1}u$ to get

$$\mu \left(\left\{ x : |x| \geq \tilde{u}^{-1} \right\} \right) \leq \frac{C_1}{\tilde{u}^d} \int_{|\xi| \leq C_2\tilde{u}} (1 - \operatorname{Re}\widehat{\mu}(\xi)) d\xi$$

with

$$C_1 := \frac{C_3}{(8d)^d} \quad \text{and} \quad C_2 := 8d.$$

□

Proposition 3.8 *The functional $\widehat{\mathcal{E}}: \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}_{\geq 0} \cup \{\infty\}$ is lower semi-continuous with respect to the narrow convergence and its sub-levels are narrowly compact.*

Proof Lower semi-continuity and thence closedness of the sub-levels follows from Fatou’s lemma, because narrow convergence corresponds to pointwise convergence of the Fourier transform and the integrand in the definition of $\widehat{\mathcal{E}}$ is non-negative.

Now, assume we have fixed a constant $K > 0$ and

$$\mu \in N_K(\widehat{\mathcal{E}}) := \{\mu \in \mathcal{P}(\mathbb{R}^d) : \widehat{\mathcal{E}}[\mu] \leq K\}.$$

We show the tightness of the family of probability measures $N_K(\widehat{\mathcal{E}})$ using Lemma 3.7. Let $0 < u \leq 1$. Then,

$$\begin{aligned} & \frac{1}{u^d} \int_{|\xi| \leq C_2u} (1 - \operatorname{Re}\widehat{\mu}(\xi)) d\xi \\ & \leq C_2^d \int_{|\xi| \leq C_2u} |\xi|^{-d} (1 - \operatorname{Re}\widehat{\mu}(\xi)) d\xi \\ & \leq C_2^d \int_{|\xi| \leq C_2u} |\xi|^{-d} (|1 - \operatorname{Re}\widehat{\omega}(\xi)| + |\operatorname{Re}\widehat{\omega}(\xi) - \operatorname{Re}\widehat{\mu}(\xi)|) d\xi \\ & \leq C_2^d \int_{|\xi| \leq C_2u} |\xi|^{-d} (|1 - \widehat{\omega}(\xi)| + |\widehat{\omega}(\xi) - \widehat{\mu}(\xi)|) d\xi \\ & = C_2^d \int_{|\xi| \leq C_2u} |\xi|^{(-d-q)/2} \cdot |\xi|^{(-d+q)/2} (|1 - \widehat{\omega}(\xi)| + |\widehat{\omega}(\xi) - \widehat{\mu}(\xi)|) d\xi \end{aligned} \tag{3.16}$$

$$\leq C_2^d \underbrace{\left(\int_{|\xi| \leq C_2 u} |\xi|^{-d+q} \, d\xi \right)^{1/2}}_{=:f(u)} \cdot \underbrace{\left[\left(\int_{|\xi| \leq C_2 u} |\xi|^{-d-q} |1 - \widehat{\omega}(\xi)|^2 \, d\xi \right)^{1/2} \right]}_{=C \cdot \widehat{\mathcal{E}}[\delta_0] < \infty} \tag{3.17}$$

$$+ \underbrace{\left(\int_{|\xi| \leq C_2 u} |\xi|^{-d-q} |\widehat{\omega}(\xi) - \widehat{\mu}(\xi)|^2 \, d\xi \right)^{1/2}}_{\leq D_q^{-1} K} \tag{Holder's inequality} \tag{3.18}$$

$$\leq C_2^d (f(u))^{1/2} \left(C^{1/2} + (D_q^{-1} K)^{1/2} \right),$$

where in equations (3.17) and (3.18) we used the boundedness of the first summand in (3.17) by a constant $C > 0$, which is justified because ω has bounded second moment. But

$$f(u) = \int_{|\xi| \leq C_2 u} |\xi|^{-d+q} \, d\xi = O(u^q) \text{ for } u \rightarrow 0,$$

giving a uniform control of the convergence to zero of the left-hand side of (3.16). Together with Lemma 3.7, this yields tightness of $N_K(\widehat{\mathcal{E}})$, hence relative compactness with respect to narrow convergence. Compactness then follows from the aforementioned lower semi-continuity of $\widehat{\mathcal{E}}$. \square

From this proof, we cannot deduce a stronger compactness, so that the limit of a minimizing sequence for the original functional $\widetilde{\mathcal{E}}$ (which coincides with $\widehat{\mathcal{E}}$ on $\mathcal{P}_2(\mathbb{R}^d)$ by Theorem 3.6) need not lie in the set $\mathcal{P}_2(\mathbb{R}^d)$ (actually, in Sect. 4, we shall see that we can prove a slightly stronger compactness). To apply compactness arguments, we hence need an extension of \mathcal{E} to the whole $\mathcal{P}(\mathbb{R}^d)$. For the direct method and later Γ -convergence to be applied, this extension should also be lower semi-continuous; therefore the natural candidate is the *lower semi-continuous envelope* $\widetilde{\mathcal{E}}^-$ of $\widetilde{\mathcal{E}}$, now defined on the whole $\mathcal{P}(\mathbb{R}^d)$ by

$$\widetilde{\mathcal{E}}[\mu] := \begin{cases} \widehat{\mathcal{E}}[\mu], & \mu \in \mathcal{P}_2(\mathbb{R}^d), \\ \infty, & \mu \in \mathcal{P}(\mathbb{R}^d) \setminus \mathcal{P}_2(\mathbb{R}^d), \end{cases}$$

which in our case can be defined as

$$\widetilde{\mathcal{E}}^-[\mu] := \inf_{\substack{\mu_n \rightarrow \mu \text{ narrowly} \\ \mu_n \in \mathcal{P}_2(\mathbb{R}^d)}} \liminf_{n \rightarrow \infty} \widetilde{\mathcal{E}}[\mu_n],$$

or equivalently as the largest lower semi-continuous functional smaller than $\widetilde{\mathcal{E}}$.

In order to show that actually $\widetilde{\mathcal{E}}^- \equiv \widehat{\mathcal{E}}$ on $\mathcal{P}(\mathbb{R}^d)$, which is the content of Theorem 3.10 below, we need for any $\mu \in \mathcal{P}(\mathbb{R}^d)$ a narrowly approximating sequence $(\mu_n)_{n \in \mathbb{N}}$ in $\mathcal{P}_2(\mathbb{R}^d)$, along which there is continuity of the values of $\widetilde{\mathcal{E}}$. We construct explicitly such approximating sequences by damping an arbitrary μ by dilated Gaussians.

Proposition 3.9 For $\omega \in \mathcal{P}_2(\mathbb{R}^d)$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$, there exists a sequence $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}_2(\mathbb{R}^d)$ such that

$$\begin{aligned} \mu_n &\rightarrow \mu \text{ narrowly} && \text{for } n \rightarrow \infty, \\ \widehat{\mathcal{E}}[\mu_n] &\rightarrow \widehat{\mathcal{E}}[\mu] && \text{for } n \rightarrow \infty. \end{aligned}$$

Proof 1. Definition of μ_n . Denote the Gaussian and its dilation respectively by

$$\eta(x) := (2\pi)^{-d/2} \exp\left(-\frac{1}{2}|x|^2\right), \quad \eta_\varepsilon(x) := \varepsilon^{-d} \eta(\varepsilon^{-1}x), \quad x \in \mathbb{R}^d.$$

Then $(2\pi)^{-d} \widehat{\eta}_\varepsilon = \eta_\varepsilon$ is a non-negative approximate identity with respect to the convolution and $\widehat{\eta}_\varepsilon = \exp(-\varepsilon^2|x|^2/2)$. To approximate μ , we use a smooth damping of the form

$$\mu_n := \widehat{\eta}_{n^{-1}} \cdot \mu + \left(1 - (\widehat{\eta}_{n^{-1}} \cdot \mu)(\mathbb{R}^d)\right) \delta_0,$$

such that the resulting μ_n are in \mathcal{P}_2 , with Fourier transforms

$$\widehat{\mu}_n(\xi) = (\widehat{\mu} * \eta_{n^{-1}})(\xi) - (\widehat{\mu} * \eta_{n^{-1}})(0) + 1, \quad \xi \in \mathbb{R}^d.$$

Note that because $\widehat{\mu}$ is continuous, $\widehat{\mu}_n(\xi) \rightarrow \widehat{\mu}(\xi)$ for all $\xi \in \mathbb{R}^d$. We want to use the dominated convergence theorem to deduce that

$$\widehat{\mathcal{E}}[\mu_n] = D_q \int_{\mathbb{R}^d} |\xi|^{-d-q} |\widehat{\mu}_n(\xi) - \widehat{\omega}(\xi)|^2 d\xi \rightarrow \widehat{\mathcal{E}}[\mu] \quad \text{for } n \rightarrow \infty.$$

2. Trivial case and dominating function. Firstly, note that if $\widehat{\mathcal{E}}[\mu] = \infty$, then Fatou’s lemma ensures that $\widehat{\mathcal{E}}[\mu_n] \rightarrow \infty$ as well.

Secondly, by the assumptions on ω , it is sufficient to find a dominating function for

$$\xi \mapsto |\xi|^{-d-q} |\widehat{\mu}_n(\xi) - 1|^2,$$

which will only be problematic for ξ close to 0. We can estimate the behavior of $\widehat{\mu}_n$ by that of $\widehat{\mu}$ as

$$\begin{aligned} |\widehat{\mu}_n(\xi) - 1| &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \eta_{n^{-1}}(\zeta) |\exp(i(\zeta - \xi) \cdot x) - \exp(i\zeta \cdot x)| d\mu(x) d\zeta \\ &= \int_{\mathbb{R}^d} \underbrace{\int_{\mathbb{R}^d} \eta_{n^{-1}}(\zeta) d\zeta}_{=1} |\exp(-i\xi \cdot x) - 1| d\mu(x) \\ &\leq C \left[(1 - \operatorname{Re}\widehat{\mu}(\xi)) + \underbrace{\int_{\mathbb{R}^d} |\sin(\xi \cdot x)| d\mu(x)}_{f(\xi):=} \right], \end{aligned} \tag{3.19}$$

where the right-hand side (3.19) is to serve as the dominating function. Taking the square of (3.19) yields

$$|\widehat{\mu}_n(\xi) - 1|^2 \leq C \left[(1 - \operatorname{Re}\widehat{\mu}(\xi))^2 + \left(\int_{\mathbb{R}^d} |\sin(\xi \cdot x)| \, d\mu(x) \right)^2 \right]. \tag{3.20}$$

Now, by the boundedness of the second moment of ω , we know that

$$\begin{aligned} & \int_{\mathbb{R}^d} |\xi|^{-d-q} (1 - \operatorname{Re}\widehat{\mu}(\xi))^2 \, d\xi \\ & \leq \int_{\mathbb{R}^d} |\xi|^{-d-q} |\widehat{\mu}(\xi) - 1|^2 \, d\xi \\ & \leq 2 \int_{\mathbb{R}^d} |\xi|^{-d-q} |\widehat{\mu}(\xi) - \widehat{\omega}(\xi)|^2 \, d\xi + 2 \int_{\mathbb{R}^d} |\xi|^{-d-q} |\widehat{\omega}(\xi) - 1|^2 \, d\xi < \infty \end{aligned} \tag{3.21}$$

This yields the integrability condition for the first term in equation (3.20). What remains is to show the integrability for the term f in (3.19), which will occupy the rest of the proof.

3. Splitting f . We apply the estimate

$$|\sin(y)| \leq \min\{|y|, 1\} \quad \text{for } y \in \mathbb{R},$$

resulting in

$$f(\xi) = \int_{\mathbb{R}^d} |\sin(\xi \cdot x)| \, d\mu(x) \leq |\xi| \underbrace{\int_{|x| \leq |\xi|^{-1}} |x| \, d\mu(x)}_{:=f_1(\xi)} + \underbrace{\int_{|x| \geq |\xi|^{-1}} d\mu(x)}_{:=f_2(\xi)}.$$

4. Integrability of f_2 : By Lemma 3.7 and Hölder’s inequality, we can estimate f_2 as follows:

$$\begin{aligned} f_2(\xi) & \leq \frac{C_1}{|\xi|^d} \int_{|y| \leq C_2|\xi|} (1 - \operatorname{Re}\widehat{\mu}(y)) \, dy \leq \frac{C_1}{|\xi|^d} \underbrace{\left(\int_{|y| \leq C_2|\xi|} 1 \, dy \right)^{1/2}}_{=C|\xi|^{d/2}} \\ & \quad \times \left(\int_{|y| \leq C_2|\xi|} (1 - \operatorname{Re}\widehat{\mu}(y))^2 \, dy \right)^{1/2} \end{aligned} \tag{3.22}$$

Hence, inserting (3.22) into the integral which we want to show to be finite and applying Fubini-Tonelli yields

$$\begin{aligned} \int_{\mathbb{R}^d} |\xi|^{-d-q} f_2(\xi)^2 d\xi &\leq C \int_{\mathbb{R}^d} |\xi|^{-2d-q} \int_{|y|\leq C_2|\xi|} (1 - \operatorname{Re}\widehat{\mu}(y))^2 dy d\xi \\ &\leq C \int_{\mathbb{R}^d} (1 - \operatorname{Re}\widehat{\mu}(y))^2 \underbrace{\int_{C_2|\xi|\geq|y|} |\xi|^{-2d-q} d\xi dy}_{=C|y|^{-d-q}} \\ &\leq C \int_{\mathbb{R}^d} |y|^{-d-q} (1 - \operatorname{Re}\widehat{\mu}(y))^2 dy < \infty \end{aligned}$$

by (3.21).

5. Integrability of f_1 : We use Fubini-Tonelli to get a well-known estimate for the first moment, namely

$$\begin{aligned} f_1(\xi) &= |\xi| \int_{|x|\leq|\xi|^{-1}} |x| d\mu(x) = |\xi| \int_0^\infty \int_{\mathbb{R}^d} 1_{\{z\leq|x|\leq|\xi|^{-1}\}} d\mu(x) dz \\ &\leq |\xi| \int_0^{|\xi|^{-1}} \mu(\{z \leq |x|\}) dz. \end{aligned}$$

Next, we use Lemma 3.7 and Hölder’s inequality (twice) to obtain (remember that $1 \leq q < 2$ which ensures integrability)

$$\begin{aligned} f_1(\xi) &\leq C_1 |\xi| \int_0^{|\xi|^{-1}} z^d \int_{|\zeta|\leq C_2 z^{-1}} (1 - \operatorname{Re}\widehat{\mu}(\zeta)) d\zeta dz \\ &\leq C_1 |\xi| \int_0^{|\xi|^{-1}} z^d \underbrace{\left(\int_{|\zeta|\leq C_2 z^{-1}} 1 d\zeta \right)^{1/2}}_{=C z^{-d/2}=C z^{q/4+(-d/2-q/4)}} \left(\int_{|\zeta|\leq C_2 z^{-1}} (1 - \operatorname{Re}\widehat{\mu}(\zeta))^2 d\zeta \right)^{1/2} dz \\ &\leq C |\xi| \underbrace{\left(\int_0^{|\xi|^{-1}} z^{-q/2} dz \right)^{1/2}}_{=C|\xi|^{q/4-1/2}} \left(\int_0^{|\xi|^{-1}} \int_{|\zeta|\leq C_2 z^{-1}} z^{d+q/2} (1 - \operatorname{Re}\widehat{\mu}(\zeta))^2 d\zeta dz \right)^{1/2}. \end{aligned}$$

By squaring the expression and using Fubini-Tonelli on the second term, we obtain

$$\begin{aligned} f_1(\xi)^2 &\leq C |\xi|^{1+q/2} \int_{\mathbb{R}^d} (1 - \operatorname{Re}\widehat{\mu}(\zeta))^2 \int_0^{|\xi|^{-1}} 1_{\{z\leq C_2|\zeta|^{-1}\}} z^{d+q/2} dz d\zeta \\ &\leq C |\xi|^{1+q/2} \int_{\mathbb{R}^d} (1 - \operatorname{Re}\widehat{\mu}(\zeta))^2 \min \left\{ |\xi|^{-d-q/2-1}, |\zeta|^{-d-q/2-1} \right\} d\zeta \\ &= C |\xi|^{-d} \int_{|\zeta|\leq|\xi|} (1 - \operatorname{Re}\widehat{\mu}(\zeta))^2 d\zeta \tag{3.23} \end{aligned}$$

$$\begin{aligned} &+ C |\xi|^{1+q/2} \underbrace{\int_{|\zeta|\geq|\xi|} |\zeta|^{-d-q/2-1} (1 - \operatorname{Re}\widehat{\mu}(\zeta))^2 d\zeta}_{:=f_3(\xi)} \tag{3.24} \end{aligned}$$

The integrability against $\xi \mapsto |\xi|^{-d-q}$ of the term (3.23) can now be shown analogously to (3.22) in Step 2. Inserting the term (3.24) into the integral and again applying Fubini-Tonelli yields

$$\begin{aligned} & \int_{\mathbb{R}^d} |\xi|^{-d-q} f_3(\xi)^2 d\xi \\ & \leq C \int_{\mathbb{R}^d} |\xi|^{-d-q/2+1} \int_{|\zeta| \geq |\xi|} |\zeta|^{-d-q/2-1} (1 - \operatorname{Re}\widehat{\mu}(\zeta))^2 d\zeta d\xi \\ & = C \int_{\mathbb{R}^d} |\zeta|^{-d-q/2-1} (1 - \operatorname{Re}\widehat{\mu}(\zeta))^2 \underbrace{\int_{|\xi| \leq |\zeta|} |\xi|^{-d-q/2+1} d\xi d\zeta}_{=C|\zeta|^{-q/2+1}} \\ & = C \int_{\mathbb{R}^d} |\zeta|^{-d-q} (1 - \operatorname{Re}\widehat{\mu}(\zeta))^2 d\zeta < \infty, \end{aligned}$$

because of (3.21), which ends the proof. □

Theorem 3.10 *We have that*

$$\widetilde{\mathcal{E}}^-[\mu] = \widehat{\mathcal{E}}[\mu], \quad \mu \in \mathcal{P}(\mathbb{R}^d)$$

and that ω is the unique minimizer of $\widetilde{\mathcal{E}}^-$.

Proof For $\mu \in \mathcal{P}(\mathbb{R}^d)$ and any sequence $(\mu_n)_{n \in \mathbb{N}}$ in $\mathcal{P}_2(\mathbb{R}^d)$ with $\mu_n \rightarrow \mu$ narrowly, we have

$$\liminf_{n \rightarrow \infty} \widetilde{\mathcal{E}}[\mu_n] = \liminf_{n \rightarrow \infty} \widehat{\mathcal{E}}[\mu_n] \geq \widehat{\mathcal{E}}[\mu],$$

by the lower semi-continuity of $\widehat{\mathcal{E}}$. By taking the infimum over all the sequences converging narrowly to μ , we conclude

$$\widetilde{\mathcal{E}}^-[\mu] \geq \widehat{\mathcal{E}}[\mu] \quad \text{for all } \mu \in \mathcal{P}(\mathbb{R}^d). \tag{3.25}$$

Conversely, for $\mu \in \mathcal{P}(\mathbb{R}^d)$, employing the sequence $(\mu_n)_{n \in \mathbb{N}}$ in $\mathcal{P}_2(\mathbb{R}^d)$ of Proposition 3.9 allows us to see that

$$\widehat{\mathcal{E}}[\mu] = \lim_{n \rightarrow \infty} \widehat{\mathcal{E}}[\mu_n] = \lim_{n \rightarrow \infty} \widetilde{\mathcal{E}}[\mu_n] \geq \widetilde{\mathcal{E}}^-[\mu]. \tag{3.26}$$

Combining (3.26) with (3.25) yields the first claim, while the characterization of the minimizer follows from the form of $\widehat{\mathcal{E}}$ in (3.6). □

Having verified that $\widehat{\mathcal{E}}$ is the lower semi-continuous envelope and a natural extension of \mathcal{E} to $\mathcal{P}(\mathbb{R}^d)$, in the following we shall work with the functional $\widehat{\mathcal{E}}$ instead of \mathcal{E} or $\widetilde{\mathcal{E}}$.

Remark 3.11 Notice that the lower semi-continuous envelope and therefore $\widehat{\mathcal{E}}$ is also the Γ -limit, see Definition 3.12 below, of a regularization of $\widetilde{\mathcal{E}}$ using the second moment, i.e., by considering

$$\mathcal{I}_\varepsilon[\mu] := \widetilde{\mathcal{E}}[\mu] + \varepsilon \int_{\mathbb{R}^d} |x|^2 d\mu,$$

we have

$$\mathcal{I}_\varepsilon \xrightarrow{\Gamma} \tilde{\mathcal{E}}^- \quad \text{for } \varepsilon \rightarrow 0.$$

3.3 Consistency of the Particle Approximations

Let $N \in \mathbb{N}$ and define

$$\mathcal{P}^N(\mathbb{R}^d) := \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \text{ for some } \{x_i : i = 1, \dots, N\} \subset \mathbb{R}^d \right\}$$

and consider the restricted minimization problem

$$\widehat{\mathcal{E}}_N[\mu] := \begin{cases} \widehat{\mathcal{E}}[\mu], & \mu \in \mathcal{P}^N(\mathbb{R}^d), \\ \infty, & \text{otherwise} \end{cases} \rightarrow \min_{\mu \in \mathcal{P}(\mathbb{R}^d)}. \tag{3.27}$$

We want to prove consistency of the restriction in terms of Γ -convergence of $\widehat{\mathcal{E}}_N$ to $\widehat{\mathcal{E}}$ for $N \rightarrow \infty$. This implies that the discrete measures minimizing $\widehat{\mathcal{E}}_N$ will converge to the unique minimizer ω of $\widehat{\mathcal{E}}$ for $N \rightarrow \infty$, in other words the measure quantization of ω via the minimization of $\widehat{\mathcal{E}}_N$ is consistent.

Definition 3.12 (Γ -convergence) [12, Definition 4.1, Proposition 8.1] Let X be a separable metrizable space and $F_N : X \rightarrow (-\infty, \infty]$, $N \in \mathbb{N}$ define a sequence of functionals. Then we say that $(F_N)_{N \in \mathbb{N}}$ Γ -converges to F , written as $F_N \xrightarrow{\Gamma} F$, for an $F : X \rightarrow (-\infty, \infty]$, if

1. *lim inf-condition*: For every $x \in X$ and every sequence $x_N \rightarrow x$,

$$F(x) \leq \liminf_{N \rightarrow \infty} F_N(x_N);$$

2. *lim sup-condition*: For every $x \in X$, there exists a sequence $x_N \rightarrow x$, called *recovery sequence*, such that

$$F(x) \geq \limsup_{N \rightarrow \infty} F_N(x_N).$$

Furthermore, we call the sequence $(F_N)_{N \in \mathbb{N}}$ *equi-coercive* if for every $c \in \mathbb{R}$ there is a compact set $K \subset X$ such that $\{x : F_N(x) \leq c\} \subset K$ for all $N \in \mathbb{N}$. As a direct consequence, assuming $x_N \in \arg \min F_N \neq \emptyset$ for all $N \in \mathbb{N}$, there is a subsequence $(x_{N_k})_{k \in \mathbb{N}}$ and $x^* \in X$ such that

$$x_{N_k} \rightarrow x^* \in \arg \min F, \quad k \rightarrow \infty. \tag{3.28}$$

As we wish to consider Γ -convergence of $F_N = \widehat{\mathcal{E}}_N$ on $X = \mathcal{P}(\mathbb{R}^d)$ to $\widehat{\mathcal{E}}$, it is relevant to recall here again that $\mathcal{P}(\mathbb{R}^d)$ is metrizable when endowed with the narrow topology.

Theorem 3.13 (Consistency of particle approximations) *The functionals $(\widehat{\mathcal{E}}_N)_{N \in \mathbb{N}}$ are equi-coercive and*

$$\widehat{\mathcal{E}}_N \xrightarrow{\Gamma} \widehat{\mathcal{E}} \text{ for } N \rightarrow \infty,$$

with respect to the narrow topology. In particular,

$$\emptyset \neq \arg \min_{\mu \in \mathcal{P}(\mathbb{R}^d)} \widehat{\mathcal{E}}_N[\mu] \ni \tilde{\mu}_N \rightarrow \tilde{\mu} = \arg \min_{\mu \in \mathcal{P}(\mathbb{R}^d)} \widehat{\mathcal{E}}[\mu] = \omega,$$

for any choice of minimizers $\tilde{\mu}_N$.

Proof 1. *Equi-coercivity:* This follows from the fact that $\widehat{\mathcal{E}}$ has compact sub-levels by Proposition 3.8, together with $\widehat{\mathcal{E}}_N \geq \widehat{\mathcal{E}}$.

2. *lim inf-condition:* Let $\mu_N \in \mathcal{P}(\mathbb{R}^d)$ such that $\mu_N \rightarrow \mu$ narrowly for $N \rightarrow \infty$. Then

$$\liminf_{N \rightarrow \infty} \widehat{\mathcal{E}}_N[\mu_N] \geq \liminf_{N \rightarrow \infty} \widehat{\mathcal{E}}[\mu_N] \geq \widehat{\mathcal{E}}[\mu],$$

by the lower semi-continuity of $\widehat{\mathcal{E}}$.

3. *lim sup-condition:* Let $\mu \in \mathcal{P}(\mathbb{R}^d)$. By Proposition 3.9, we can find a sequence $(\mu^k)_{k \in \mathbb{N}} \subset \mathcal{P}_2(\mathbb{R}^d)$, for which $\widehat{\mathcal{E}}[\mu^k] \rightarrow \widehat{\mathcal{E}}[\mu]$. Furthermore, by Lemma 3.3, we can approximate each μ^k by $(\mu_N^k)_{N \in \mathbb{N}} \subset \mathcal{P}_2(\mathbb{R}^d) \cap \mathcal{P}^N(\mathbb{R}^d)$, a realization of the empirical distribution of μ^k . This has a further subsequence which converges in the 2-Wasserstein distance by Lemma 2.3, for which we have continuity of $\widehat{\mathcal{E}}$ by Lemma 3.5. A diagonal argument then yields a sequence $\mu_N \in \mathcal{P}^N(\mathbb{R}^d)$ for which

$$\widehat{\mathcal{E}}_N[\mu_N] = \widehat{\mathcal{E}}[\mu_N] \rightarrow \widehat{\mathcal{E}}[\mu] \text{ for } N \rightarrow \infty.$$

4. *Convergence of minimizers:* First of all notice that, for all $N \in \mathbb{N}$, $\mathcal{P}^N(\mathbb{R}^d)$ is closed in the narrow topology. As $(\widehat{\mathcal{E}}_N)_N$ is equi-coercive and each $\widehat{\mathcal{E}}_N$ is lower semi-continuous by Fatou’s lemma, by the direct method of calculus of variations we conclude that $\widehat{\mathcal{E}}_N$ has empirical measure minimizers for all $N \in \mathbb{N}$. The convergence of the minimizers $\tilde{\mu}_N$ to $\tilde{\mu} = \omega$ then follows by (3.28) and by being ω the unique minimizer of $\widehat{\mathcal{E}}$. □

4 An Enhanced Moment Bound

Let $q = q_a = q_r \in (1, 2)$ be strictly larger than 1 now. We want to prove that in this case, we have a stronger compactness than the one showed in Proposition 3.8, namely that the sub-levels of $\widehat{\mathcal{E}}$ have a uniformly bounded r th moment for $r < q/2$.

In the proof, we shall be using the theory recalled in Appendix in a more elaborated form than before, in particular we need to extend the generalized Fourier transform (Definition 7.3) and its explicit computation for power functions as shown in Theorem 7.6.

Theorem 4.1 *Let $\omega \in \mathcal{P}_2(\mathbb{R}^d)$. For $r < q/2$ and a given $M > 0$, there exists an $M' > 0$ such that*

$$\int_{\mathbb{R}^d} |x|^r \, d\mu(x) \leq M', \text{ for all } \mu \text{ such that } \widehat{\mathcal{E}}[\mu] \leq M.$$

Proof Let $\mu \in \mathcal{P}(\mathbb{R}^d)$. If $\widehat{\mathcal{E}}[\mu] \leq M$, then we also have

$$\begin{aligned}
 M &\geq \widehat{\mathcal{E}}[\mu] = D_q \int_{\mathbb{R}^d} |\widehat{\mu}(\xi) - \widehat{\omega}(\xi)|^2 |\xi|^{-d-q} \, d\xi \\
 &\geq \left(\int_{\mathbb{R}^d} |\widehat{\mu}(\xi) - 1|^2 |\xi|^{-d-q} \, d\xi - c \int_{\mathbb{R}^d} |\widehat{\omega}(\xi) - 1|^2 |\xi|^{-d-q} \, d\xi \right),
 \end{aligned}$$

so that there is an $M'' > 0$ such that

$$\int_{\mathbb{R}^d} |\widehat{\mu} - 1|^2 |\xi|^{-d-q} \, d\xi \leq M''.$$

Now approximate μ by the sequence of Proposition 3.9, again denoted by μ_n , and then μ_n by convolution with $\eta_{k_n^{-1}}$ to obtain the sequence $\mu'_n := \mu_n * \eta_{k_n^{-1}}$, so that we have convergence $\widehat{\mathcal{E}}[\mu'_n] \rightarrow \widehat{\mathcal{E}}[\mu]$. We set $\widehat{v}_n := (\mu'_n - \eta_{k_n^{-1}})$. By the damping of Proposition 3.9, the underlying measures μ_n and μ'_n have finite moment of any order, yielding decay of $\widehat{v}_n(x)$ of arbitrary polynomial order for $|x| \rightarrow \infty$, and the mollification implies $\widehat{v}_n \in C^\infty(\mathbb{R}^d)$. We conclude that $\widehat{v}_n \in \mathcal{S}(\mathbb{R}^d)$, where $\mathcal{S}(\mathbb{R}^d)$ is the space of Schwartz functions. Furthermore, set $v_n = \widehat{v}_n^\vee$ the inverse Fourier transform, and recall that the inverse Fourier transform is also expressed as an integral with an exponential function factor. By expanding this exponential function in its Taylor series, as it is done, for instance in Lemma 3.5, we see that for each n ,

$$v_n(\xi) = O(|\xi|) \quad \text{for } \xi \rightarrow 0.$$

Therefore, $v_n \in \mathcal{S}_1(\mathbb{R}^d)$, see Definition 7.2, and we can apply Theorem 7.6.2, to get

$$\begin{aligned}
 &\int_{\mathbb{R}^d} |x|^r \widehat{v}_n(x) \, dx \\
 &= C \int_{\mathbb{R}^d} |\xi|^{-d-r} v_n(\xi) \, d\xi \\
 &\leq C \left[\int_{|\xi| \leq 1} \underbrace{|\xi|^{-d-r}}_{=|\xi|^{-\frac{d-q+2r}{2}} |\xi|^{-\frac{d+q}{2}}} |v_n(\xi)| \, d\xi + \underbrace{\int_{|\xi| > 1} |\xi|^{-d-r} |v_n(\xi)| \, d\xi}_{\leq C < \infty} \right] \\
 &\leq C \left[\underbrace{\left(\int_{|\xi| \leq 1} |\xi|^{-d+(q-2r)} \, d\xi \right)^{1/2}}_{< \infty} \left(\int_{\mathbb{R}^d} |\xi|^{-d-q} |v_n|^2 \, d\xi \right)^{1/2} + 1 \right] \\
 &\leq C \left[\left(\int_{\mathbb{R}^d} |\xi|^{-d-q} |v_n|^2 \, d\xi \right)^{1/2} + 1 \right].
 \end{aligned}$$

Now, we recall again the continuity of $\widehat{\mathcal{E}}$ for $\omega = \delta_0$ along μ_n by Proposition 3.9, and its continuity with respect to the Gaussian mollification. The latter can be seen either by the 2-Wasserstein-convergence of the mollification for n fixed or by using

the dominated convergence theorem together with the power series expansion of exp, similarly to Lemma 5.1 below. To summarize, we see that

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} |\xi|^{-d-q} |v_n|^2 d\xi = (2\pi)^{-d} \int_{\mathbb{R}^d} |\xi|^{-d-q} |\widehat{\mu} - 1|^2 d\xi \leq (2\pi)^{-d} M'',$$

while on the other hand we have

$$\begin{aligned} \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^d} |x|^r \widehat{v}_n(x) dx &= \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^d} |x|^r d\mu_n(x) - \underbrace{\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} |x|^r \eta_{k_n^{-1}}(x) dx}_{=0} \\ &\geq \int_{\mathbb{R}^d} |x|^r d\mu(x) \end{aligned}$$

by Lemma 2.4, concluding the proof. □

5 Regularization by Using the Total Variation

We shall regularize the functional $\widehat{\mathcal{E}}$ by an additional total variation term, for example to reduce the possible effect of noise on the given datum ω . In particular, we expect the minimizer of the corresponding functional to be piecewise smooth or even piecewise constant while any sharp discontinuity in ω should be preserved, as it is the case for the regularization of a L^2 -norm data fitting term, as it is often used in image denoising, see for example [9, Chapter 4].

In the following, we begin by introducing this regularization and prove that for a vanishing regularization parameter, the minimizers of the regularizations converge to the minimizer of the original functional. The regularization allows us to consider approximating minimizers in $\mathcal{P}(\mathbb{R}^d) \cap BV(\mathbb{R}^d)$, where $BV(\mathbb{R}^d)$ is the space of bounded variation functions. Once the regularization is introduced, we consider its empirical measure approximations. In the classical literature, one finds several approaches to discrete approximations to functionals involving total variation regularization terms as well as to their BV -minimizers, e.g., by means of finite element type approximations, see for example [4]. Here however, we propose an approximation which depends on the position of (freely moving) particles in \mathbb{R}^d , which can be combined with the particle approximation of Sect. 3.3. To this end, in Sect. 5.2, we shall present two ways of embedding into L^1 the Dirac masses which are associated to particles.

5.1 Consistency of the Regularization for the Continuous Functional

For $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\lambda > 0$, define

$$\widehat{\mathcal{E}}^\lambda[\mu] := \begin{cases} \widehat{\mathcal{E}}[\mu] + \lambda |D\mu|(\mathbb{R}^d), & \mu \in \mathcal{P}(\mathbb{R}^d) \cap BV(\mathbb{R}^d), \\ \infty, & \text{otherwise,} \end{cases} \tag{5.1}$$

where $D\mu$ denotes the distributional derivative of μ , being a finite Radon-measure, and $|D\mu|$ its total variation measure [1]. We present first a useful technical result before proceeding to prove the Γ -convergence $\widehat{\mathcal{E}}^\lambda \xrightarrow{\Gamma} \widehat{\mathcal{E}}$ for $\lambda \rightarrow 0$.

Lemma 5.1 (Continuity of $\widehat{\mathcal{E}}$ with respect to Gaussian mollification) *Let $\omega \in \mathcal{P}_2(\mathbb{R}^d)$, $\mu \in \mathcal{P}(\mathbb{R}^d)$ and set again*

$$\eta(x) := (2\pi)^{-d/2} \exp\left(-\frac{1}{2}|x|^2\right), \quad \eta_\varepsilon(x) := \varepsilon^{-d} \eta(\varepsilon^{-1}x), \quad x \in \mathbb{R}^d.$$

Then,

$$\widehat{\mathcal{E}}[\eta_\varepsilon * \mu] \rightarrow \widehat{\mathcal{E}}[\mu], \quad \text{for } \varepsilon \rightarrow 0.$$

Proof If $\widehat{\mathcal{E}}[\mu] = \infty$, then the claim is true by the lower semi-continuity of $\widehat{\mathcal{E}}$ together with the fact that $\eta_\varepsilon * \mu \rightarrow \mu$ narrowly.

If $\widehat{\mathcal{E}}[\mu] < \infty$, we can estimate the difference $|\widehat{\mathcal{E}}[\eta_\varepsilon * \mu] - \widehat{\mathcal{E}}[\mu]|$ (which is well-defined, but for now may be ∞) by using

$$|a^2 - b^2| \leq |a - b| \cdot (|a| + |b|), \quad a, b \in \mathbb{C},$$

and

$$\widehat{\eta_\varepsilon * \mu}(\xi) = \exp\left(-\frac{\varepsilon^2}{2}|\xi|^2\right) \widehat{\mu}(\xi),$$

as

$$\begin{aligned} & |\widehat{\mathcal{E}}[\eta_\varepsilon * \mu] - \widehat{\mathcal{E}}[\mu]| \\ & \leq D_q \int_{\mathbb{R}^d} \left| |\widehat{\eta}_\varepsilon(\xi) \widehat{\mu}(\xi) - \widehat{\omega}(\xi)|^2 - |\widehat{\mu}(\xi) - \widehat{\omega}(\xi)|^2 \right| |\xi|^{-d-q} \, d\xi \\ & \leq D_q \int_{\mathbb{R}^d} \underbrace{(|\widehat{\eta}_\varepsilon(\xi) \widehat{\mu}(\xi) - \widehat{\omega}(\xi)| + |\widehat{\mu}(\xi) - \widehat{\omega}(\xi)|)}_{\leq 4} \underbrace{|\widehat{\eta}_\varepsilon(\xi) \widehat{\mu}(\xi) - \widehat{\mu}(\xi)|}_{=(1 - \exp(-(\varepsilon^2/2)|\xi|^2)) \widehat{\mu}(\xi)} |\xi|^{-d-q} \, d\xi \\ & \leq C \int_{\mathbb{R}^d} \left(1 - \exp\left(-\frac{\varepsilon^2}{2}|\xi|^2\right)\right) |\xi|^{-d-q} \, d\xi, \end{aligned} \tag{5.2}$$

which converges to 0 by the dominated convergence theorem: indeed, on the one hand we can estimate

$$\exp\left(-\frac{\varepsilon^2}{2}|\xi|^2\right) \geq 0, \quad \xi \in \mathbb{R}^d,$$

yielding a dominating function for the integrand in (5.2) for ξ bounded away from 0 because of the integrability of $\xi \mapsto |\xi|^{-d-q}$ there. On the other hand

$$\begin{aligned}
 1 - \exp\left(-\frac{\varepsilon^2}{2} |\xi|^2\right) &= - \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\varepsilon^2}{2} |\xi|^2\right)^n \\
 &= \frac{\varepsilon^2}{2} |\xi|^2 \sum_{n=0}^{\infty} \frac{1}{(n+1)!} \left(-\frac{\varepsilon^2}{2} |\xi|^2\right)^n,
 \end{aligned}$$

where the sum on the right is uniformly bounded for $\varepsilon \rightarrow 0$ as a convergent power-series, which, combined with $q < 2$, renders the integrand in (5.2) dominated for ξ near 0 as well. □

Proposition 5.2 (Consistency) *Let $(\lambda_N)_{N \in \mathbb{N}}$ be a vanishing sequence of positive parameters. The functionals $(\widehat{\mathcal{E}}^{\lambda_N})_{N \in \mathbb{N}}$ are equi-coercive and*

$$\widehat{\mathcal{E}}^{\lambda_N} \xrightarrow{\Gamma} \widehat{\mathcal{E}} \text{ for } N \rightarrow 0,$$

with respect to the narrow topology. In particular, the limit point of minimizers of $\widehat{\mathcal{E}}^{\lambda_N}$ coincides with the unique minimizer ω of $\widehat{\mathcal{E}}$.

Proof Firstly, observe that equi-coercivity follows from the narrow compactness of the sub-levels of $\widehat{\mathcal{E}}$ shown in Proposition 3.8, and that the *lim inf-condition* for the Γ -convergence is a consequence of the lower semi-continuity of $\widehat{\mathcal{E}}$ as in the proof of Theorem 3.13.

Existence of minimizers: We again want to apply the direct method of the calculus of variations.

Let $(\mu_k)_k$ be a minimizing sequence for $\widehat{\mathcal{E}}^\lambda$, so that the μ_k are all contained in a common sub-level of the functional. Now, for a given λ , the sub-levels of $\widehat{\mathcal{E}}^\lambda$ are relatively compact in $L^1(\mathbb{R}^d)$, which can be seen by combining the compactness of the sub-levels of the total variation in $L^1_{loc}(\mathbb{R}^d)$ with the tightness gained by $\widehat{\mathcal{E}}$: if $\widehat{\mathcal{E}}^\lambda[\mu_k] \leq M < \infty$, we can consider $(\theta_l \mu_k)_k$ for a smooth cut-off function θ_l having its support in $[-l, l]^d$. By standard arguments we have the product formula

$$D(\theta_l \mu_k) = D\theta_l \mu_k + \theta_l D\mu_k$$

and therefore

$$\begin{aligned}
 |D(\theta_l \mu_k)|(\mathbb{R}^d) &\leq \int_{\mathbb{R}^d} \mu_k(x) |D\theta_l(x)| dx + \int_{\mathbb{R}^d} \theta_l(x) d|D\mu_k|(x) \\
 &\leq C_l + |D\mu_k|(\mathbb{R}^d),
 \end{aligned}$$

so that for each l , by the compactness of the sub-levels of the total variation in L^1_{loc} , see [17, Chapter 5.2, Theorem 4], we can select an L^1 -convergent subsequence, which we denote again by $(\theta_l \mu_k)_k$. Then, by extracting further subsequences for $m \geq l$ and applying a diagonal argument we can construct a subsequence, again denoted μ_k such that

$$\|\mu_k - \mu_h\|_{L^1} \leq \|(1 - \theta_l)(\mu_k - \mu_h)\|_{L^1} + \|\theta_l \mu_k - \theta_l \mu_h\|_{L^1},$$

and

$$\|(1 - \theta_l)(\mu_k - \mu_h)\|_{L^1} \leq \frac{\varepsilon}{2},$$

for $l \geq l_0(\varepsilon)$ large enough, because of the tightness of $(\mu_k)_k$, and

$$\|\theta_l \mu_k - \theta_l \mu_h\|_{L^1} \leq \frac{\varepsilon}{2},$$

by the local convergence in L^1_{loc} for $h, k \geq k_0(l)$ large enough. From this we conclude that $(\mu_k)_k$ is a Cauchy subsequence in L^1 , hence, convergent.

The lower semi-continuity of $\widehat{\mathcal{E}}^\lambda$ follows from the lower semi-continuity of the total variation with respect to L^1 -convergence and the lower semi-continuity of $\widehat{\mathcal{E}}$ with respect to narrow convergence. Summarizing, we have compactness and lower semi-continuity, giving us that $(\mu_k)_k$ has a limit point which is a minimizer.

lim sup-condition: Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and write $\mu_\varepsilon := \widehat{\eta}_\varepsilon * \mu$ for the mollification of Lemma 5.1. Now, by Fubini’s theorem,

$$\begin{aligned} |D(\widehat{\eta}_\varepsilon * \mu)|(\mathbb{R}^d) &= \int_{\mathbb{R}^d} |(\nabla \widehat{\eta}_\varepsilon * \mu)(x)| \, dx \\ &\leq \|\nabla \widehat{\eta}_\varepsilon\|_{L^1(\mathbb{R}^d)} \mu(\mathbb{R}^d) \\ &= \varepsilon^{-d} \|\nabla \widehat{\eta}\|_{L^1(\mathbb{R}^d)}, \end{aligned}$$

so if we choose $\varepsilon(\lambda)$ such that $\lambda = o(\varepsilon^d)$, for example $\varepsilon(\lambda) := \lambda^{1/(d+1)}$, then

$$\lambda |D\mu_{\varepsilon(\lambda)}|(\mathbb{R}^d) \rightarrow 0, \quad \text{for } \lambda \rightarrow 0.$$

On the other hand, $\widehat{\mathcal{E}}[\mu_{\varepsilon(\lambda)}] \rightarrow \widehat{\mathcal{E}}[\mu]$ by Lemma 5.1, yielding the required convergence $\widehat{\mathcal{E}}^\lambda[\mu_{\varepsilon(\lambda)}] \rightarrow \widehat{\mathcal{E}}[\mu]$.

The convergence of the minimizers then follows. □

5.2 Discrete Versions of the Total Variation Regularization

As one motivation for the study of the functional \mathcal{E} was to show consistency of its particle minimizers, we shall also here consider a discretized version of the total variation regularization, for example to be able to compute the minimizers of the regularized functional directly on the level of the point approximations. We propose two techniques for this discretization.

The first technique is well-known in the non-parametric estimation of L^1 densities and consists of replacing each point with a small “bump” instead of interpreting it as a point measure. In order to get the desired convergence properties, we have to be careful when choosing the corresponding scaling of the bump. For an introduction to this topic, see [14, Chapter 3.1].

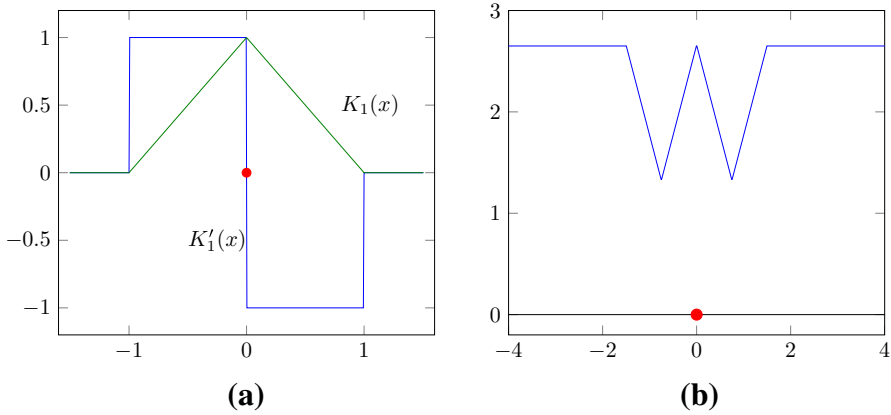


Fig. 2 Example for the discrete total variation functional. In particular, **b** shows the value of $|DQ_{0.75}[\mu_2]|(\mathbb{R})$, for $\mu_2 = \frac{1}{2}(\delta_{x_1} + \delta_{x_2})$ and for the linear kernel K_1 of **(a)**, as a function of the point $x_2 \in \mathbb{R}$ having $x_1 = 0$ fixed. **a** Linear $K_1(x)$ and corresponding $K_1'(x)$. **b** Discrete total variation for $x_1 = 0$ fixed (red), x_2 free, $h = 0.75$

The second technique replaces the Dirac deltas by indicator functions which extend from the position of one point to the next one. Unfortunately, this poses certain difficulties in generalizing it to higher dimensions, as the set on which we extend would have to be replaced by something like a Voronoi cell, well-known in the theory of optimal quantization of measures, see for example [24].

In the context of attraction-repulsion functionals, it is of importance to note that the effect of the additional particle total variation term can again be interpreted as an attractive-repulsive-term. See Fig. 2 for an example in the case of kernel density estimation with a piecewise linear estimation kernel, where it can be seen that each point is repulsive at a short range, attractive at a medium range, and at a long range does not contribute into the total variation any more. This interpretation of the action of the total variation as a potential acting on particles to promote their uniform distribution is, to our knowledge, new.

5.2.1 Discretization by Kernel Estimators and Quantization on Deterministic Tilings

Definition 5.3 (Discrete total variation via kernel estimate) For a $\mu_N = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \in \mathcal{P}^N(\mathbb{R}^d)$, a scale parameter $h = h(N)$ and a density estimation kernel $K \in W^{1,1}(\mathbb{R}^d)$ such that $\nabla K \in BV(\mathbb{R}^d, \mathbb{R}^d)$, as well as

$$K \geq 0, \quad \int_{\mathbb{R}^d} K(x)dx = 1,$$

we set

$$K_h(x) := \frac{1}{h^d} K\left(\frac{x}{h}\right)$$

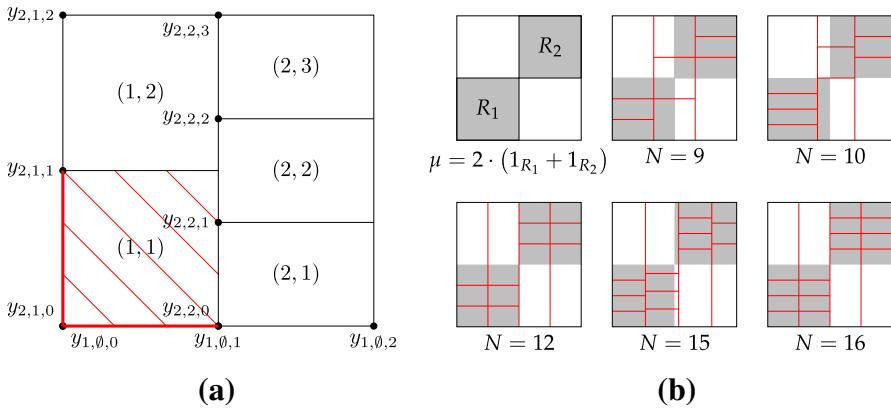


Fig. 3 Illustration of the tiling. **a** Notation of Definition 5.4 for $N = 5$. **b** Tiling as in Example 5.5 for a uniform measure on two squares in $[0, 1]^2$

and define the corresponding L^1 -density estimator by

$$Q_h[\mu_N](x) := K_h * \mu_N(x) = \frac{1}{Nh^d} \sum_{i=1}^N K\left(\frac{x - x_i}{h}\right),$$

where the definition has to be understood for almost every x . Then, we can introduce a discrete version of the regularization in (5.1) as

$$\widehat{\mathcal{E}}_N^\lambda[\mu_N] := \widehat{\mathcal{E}}[\mu_N] + \lambda |DQ_{h(N)}[\mu_N]|(\mathbb{R}^d), \quad \mu_N \in \mathcal{P}^N(\mathbb{R}^d). \tag{5.3}$$

We want to prove consistency of this approximation in terms of Γ -convergence of the functionals $\widehat{\mathcal{E}}_N^\lambda$ to $\widehat{\mathcal{E}}^\lambda$ for $N \rightarrow \infty$. For a survey on the consistency of kernel estimators in the probabilistic case under various sets of assumptions, see [37]. Here however, we want to give a proof using deterministic and explicitly constructed point approximations.

In order to find a recovery sequence for the family of functionals (5.3) (see Definition 3.12), we have to determine point approximations to a given measure with sufficiently good spatial approximation properties. For this, we suggest using a generalization of the quantile construction to higher dimensions. Let us state the properties we expect from such an approximation:

Definition 5.4 (Tiling associated to a measure) Let $\mu \in \mathcal{P}_c(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$, where $\mathcal{P}_c(\mathbb{R}^d)$ denotes the space of compactly supported probability measures, so that $\text{supp}(\mu) \subset [-R_\mu, R_\mu]^d$ for some $R_\mu > 0$, and let $N \in \mathbb{N}$. Set $\tilde{n} := \lfloor N^{1/d} \rfloor$. A good tiling (for our purposes) will be composed of an index set I and an associated tiling $(T_i)_{i \in I}$ such that (see Fig. 3 for an example of the notation):

1. I has N elements, $\#I = N$, and in each direction, we have at least \tilde{n} different indices, i.e.,

$$\{1, \dots, \tilde{n}\}^d \subset I \subset \{1, \dots, \tilde{n} + 1\}^d. \tag{5.4}$$

Additionally, for all $k \in 1, \dots, d$ and $(i_1, \dots, i_{k-1}, i_k, \dots, i_d) \in I$,

$$n_{k,i_1,\dots,i_{k-1}} := \#\{j_k : j \in I, (j_1, \dots, j_{k-1}) = (i_1, \dots, i_{k-1})\} \in \{\tilde{n}, \tilde{n} + 1\}.$$

2. There is a family of ordered real numbers only depending on the first k coordinates,

$$y_{k,i_1,\dots,i_k} \in [-R_\mu, R_\mu], \quad y_{k,i_1,\dots,i_{k-1}} < y_{k,i_1,\dots,i_k},$$

for all $k \in \{1, \dots, d\}$ and $(i_1, \dots, i_k, i_{k+1}, \dots, i_d) \in I$,

with fixed end points,

$$y_{k,i_1,\dots,i_{k-1},0} = -R_\mu, \quad y_{k,i_1,\dots,i_{k-1},n_{k,i_1,\dots,i_{k-1}}} = R_\mu,$$

associated tiles

$$T_i := \times_{k=1}^d [y_{k,i_1,\dots,(i_k-1)}, y_{k,i_1,\dots,i_k}],$$

and such that the mass of μ is equal in each of them,

$$\mu(T_i) = \frac{1}{N}, \quad \text{for all } i \in I.$$

Such a construction can always be obtained by generalizing the quantile construction. Let us show the construction explicitly for $d = 2$ as an example.

Example 5.5 (Construction in 2D) Given $N \in \mathbb{N}$, let $\tilde{n} := \lfloor \sqrt{N} \rfloor$. We can write N as

$$N = \tilde{n}^{2-m} (\tilde{n} + 1)^m + l, \tag{5.5}$$

with unique $m \in \{0, 1\}$ and $l \in \{0, \dots, \tilde{n}^{1-m} (\tilde{n} + 1)^m - 1\}$. Then we get the desired tiling by setting

$$\begin{aligned} n_{1,\emptyset} &:= \begin{cases} \tilde{n} + 1 & \text{if } m = 1, \\ \tilde{n} & \text{if } m = 0, \end{cases} \\ n_{2,i_1} &:= \begin{cases} \tilde{n} + 1 & \text{if } i_1 \leq l, \\ \tilde{n} & \text{if } i_1 \geq l + 1, \end{cases} \quad i_1 = 1, \dots, n_{1,\emptyset}, \\ w_{2,i_1,i_2} &:= \frac{1}{n_{2,i_1}}, \quad i_1 = 1, \dots, n_{1,\emptyset}, \quad i_2 = 1, \dots, n_{2,i_1}, \\ w_{1,i_1} &:= \frac{n_{2,i_1}}{\sum_{j_1} n_{2,j_1}}, \quad i_1 = 1, \dots, n_{1,\emptyset}, \end{aligned} \tag{5.6}$$

and choosing the end points of the tiles such that

$$\sum_{j_1=1}^{i_1} w_{1,j_1} = \int_{-R_\mu}^{y_{1,i_1}} \int_{-R_\mu}^{R_\mu} d\mu(x_1, x_2), \tag{5.7}$$

$$\sum_{j_1=1}^{i_1} \sum_{j_2=1}^{i_2} w_{1,j_1} w_{2,j_1,j_2} = \int_{-R_\mu}^{y_{1,i_1}} \int_{-R_\mu}^{y_{2,i_1,i_2}} d\mu(x_1, x_2).$$

Now, check that indeed $\sum_{j_1} n_{2,j_1} = N$ by (5.5) and (5.6) and that we have

$$\mu(T_{i_1,i_2}) = w_{1,i_1} w_{2,i_1,i_2} = \frac{1}{N} \text{ for all } i_1, i_2,$$

by the choice of the weights w_{1,j_1}, w_{2,j_1,j_2} as desired.

The general construction now consists of choosing a subdivision in $\tilde{n} + 1$ slices uniformly in as many dimensions as possible, while keeping in mind that in each dimension we have to subdivide in at least \tilde{n} slices. There will again be a rest l , which is filled up in the last dimension.

Proposition 5.6 (Construction for arbitrary d) *A tiling as defined in Definition 5.4 exists for all $d \in \mathbb{N}$.*

Proof Analogously to Example 5.5, let $\tilde{n} := \lfloor N^{1/d} \rfloor$ and set

$$N = \tilde{n}^{d-m} (\tilde{n} + 1)^m + l,$$

with unique $m \in \{0, \dots, d - 1\}$ and $l \in \{0, \dots, \tilde{n}^{d-1-m} (\tilde{n} + 1)^m - 1\}$. Then, we get the desired ranges by

$$n_{k,i_1,\dots,i_{k-1}} := \tilde{n} + 1, \text{ for } k \in \{1, \dots, m\} \text{ and all relevant indices;} \tag{5.8}$$

$$n_{k,i_1,\dots,i_{k-1}} := \tilde{n}, \text{ for } k \in \{m + 1, \dots, d - 1\} \text{ and all relevant indices;} \tag{5.9}$$

$$n_{d,i_1,\dots,i_{d-1}} \in \{\tilde{n}, \tilde{n} + 1\}, \text{ such that exactly } l \text{ multi-indices are } \tilde{n} + 1. \tag{5.10}$$

The weights can then be selected such that we get equal mass after multiplying by them, and the tiling is found by iteratively using a quantile construction similar to (5.7) in Example 5.5. □

Lemma 5.7 (Consistency of the approximation) *For $\mu \in \mathcal{P}_c(\mathbb{R}^d) \cap BV(\mathbb{R}^d)$, let $(T_i)_{i \in I}$ be a tiling as in Definition 5.4, and $x_i \in T_i$ for all $i \in I$ an arbitrary point in each tile. Then, $\mu_N = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$ converges narrowly to μ for $N \rightarrow \infty$. Furthermore, if*

$$h = h(N) \rightarrow 0 \text{ and } h^{2d} N \rightarrow \infty \text{ for } N \rightarrow \infty, \tag{5.11}$$

then $Q_{h(N)}[\mu_N] \rightarrow \mu$ strictly in $BV(\mathbb{R}^d)$ (strict convergence is meant here as in [1, Definition 3.14]).

Proof Suppose again that for some $R_\mu > 0$

$$\text{supp } \mu \subset [-R_\mu, R_\mu]^d.$$

Narrow convergence: By [16, Theorem 3.9.1], it is sufficient to test convergence for bounded, Lipschitz-continuous functions. So let $\varphi \in C_b(\mathbb{R}^d)$ be a Lipschitz function with Lipschitz constant L . Then,

$$\begin{aligned} \left| \int_{\mathbb{R}^d} \varphi(x) d\mu_N(x) - \int_{\mathbb{R}^d} \varphi(x) d\mu(x) \right| &= \left| \frac{1}{N} \sum_{i=1}^N \varphi(x_i) - \int_{\mathbb{R}^d} \varphi(x) d\mu(x) \right| \\ &\leq L \sum_{i \in I} \int_{T_i} |x - x_i| d\mu(x). \end{aligned}$$

Denote by

$$\widehat{\pi}_k(i_1, \dots, i_d) := (i_1, \dots, i_{k-1}, i_{k+1}, i_d)$$

the projection onto all coordinates except the k th one. Now, we exploit the uniformity of the tiling in all dimensions, (5.4): By using the triangle inequality and grouping the summands,

$$\begin{aligned} &\sum_{i \in I} \int_{T_i} |x - x_i| d\mu(x) \\ &\leq \sum_{i \in I} \sum_{k=1}^d \int_{T_i} |x^k - x_i^k| d\mu(x) \tag{5.12} \\ &= \sum_{k=1}^d \sum_{i \in \widehat{\pi}_k(I)} \sum_{j=1}^{n_{k,i_1, \dots, i_{k-1}}} \int_{T_i} |x^k - x_{i_1, \dots, i_{k-1}, j, i_k, \dots, i_d}^k| d\mu(x) \\ &\leq \sum_{k=1}^d \sum_{i \in \widehat{\pi}_k(I)} \underbrace{\sum_{j=1}^{n_{k,i_1, \dots, i_{k-1}}} (y_{k,i_1, \dots, i_{k-1}, (j-1)} - y_{k,i_1, \dots, i_{k-1}, j})}_{=2R_\mu} \underbrace{\int_{T_i} d\mu(x)}_{=1/N} \\ &\leq 2R_\mu d \frac{(\widetilde{n} + 1)^{d-1}}{N} \leq 2R_\mu d \frac{(\widetilde{n} + 1)^{d-1}}{\widetilde{n}^d} \leq \frac{C}{\widetilde{n}} \rightarrow 0 \text{ for } N \rightarrow \infty. \tag{5.13} \end{aligned}$$

L¹-convergence: As $K \in W^{1,1}(\mathbb{R}^d) \subset BV(\mathbb{R}^d)$, we can approximate it by C^1 functions which converge BV -strictly, so let us additionally assume $K \in C^1$ for now. Then,

$$\begin{aligned} &\int_{\mathbb{R}^d} |K_h * \mu_N(x) - \mu(x)| dx \\ &\leq \int_{\mathbb{R}^d} |K_h * \mu_N(x) - K_h * \mu(x)| dx + \int_{\mathbb{R}^d} |K_h * \mu(x) - \mu(x)| dx. \tag{5.14} \end{aligned}$$

By $h \rightarrow 0$, the second term goes to 0 (see [17, Chapter 5.2, Theorem 2]), so it is sufficient to consider

$$\begin{aligned}
 & \int_{\mathbb{R}^d} |K_h * \mu_N(x) - K_h * \mu(x)| \, dx & (5.15) \\
 & \leq \sum_{i \in I} \int_{T_i} \int_{\mathbb{R}^d} |K_h(x - x_i) - K_h(x - y)| \, dx \, d\mu(y) \\
 & = \sum_{i \in I} \int_{T_i} \int_{\mathbb{R}^d} \left| \int_0^1 \nabla K_h(x - y + t(y - x_i)) \cdot (y - x_i) \, dt \right| \, dx \, d\mu(y) \\
 & \leq \sum_{i \in I} \int_{T_i} \int_0^1 |y - x_i| \int_{\mathbb{R}^d} |\nabla K_h(x - y + t(y - x_i))| \, dx \, dt \, d\mu(y) \\
 & = \frac{1}{h} \|\nabla K\|_{L^1} \sum_{i \in I} \int_{T_i} |y - x_i| \, d\mu(y). & (5.16)
 \end{aligned}$$

Since the left-hand side (5.15) and the right-hand side (5.16) of the above estimate are continuous with respect to strict BV convergence (by Fubini-Tonelli and convergence of the total variation, respectively), this estimate extends to a general $K \in BV(\mathbb{R}^d)$ and

$$\frac{1}{h} \sum_{i \in I} \int_{T_i} |y - x_i| \, d\mu(y) \leq \frac{C}{nh} \rightarrow 0, \quad \text{for } N \rightarrow \infty,$$

by the calculation in (5.12) and condition (5.11).

Convergence of the total variation: Similarly to the estimate in (5.14), by $h \rightarrow 0$ it is sufficient to consider the L^1 distance between $\nabla K_h * \mu_N$ and $\nabla K_h * \mu$, and we approximate a general K by a smoother $K \in C^2(\mathbb{R}^d)$. By a calculation similar to (5.15)–(5.16) as well as (5.13) and using $\nabla K_h(x) = h^{-d-1} \nabla K(x/h)$, we get

$$\begin{aligned}
 & \int_{\mathbb{R}^d} |\nabla K_h * \mu_N(x) - \nabla K_h * \mu(x)| \, dx \\
 & \leq C \frac{1}{h} \sum_{i \in I} \int_{T_i} \int_{\mathbb{R}^d} |\nabla K_h(x - x_i) - \nabla K_h(x - y)| \, dx \, d\mu(y) \\
 & \leq C \left\| D^2 K \right\|_{L^1} \frac{1}{nh^2} \rightarrow 0 \quad \text{for } N \rightarrow \infty,
 \end{aligned}$$

by the condition (5.11) we imposed on h . □

Since we associate to each $\mu_N \in \mathcal{P}_N$ an L^1 -density $Q_{h(N)}[\mu_N]$ and want to analyze both the behavior of $\mathcal{E}[\mu_N]$ and $|DQ_{h(N)}[\mu_N]|(\mathbb{R}^d)$, we need to incorporate the two different topologies involved, namely the narrow convergence of μ and L^1 -convergence of $Q_{h(N)}[\mu]$, into the concept of Γ -convergence. This can be done by using a slight generalization of the variational convergence as introduced in [3], namely the $\Gamma(q, \tau^-)$ -convergence.

Definition 5.8 ($\Gamma(q, \tau^-)$ -convergence) [3, Definition 2.1] For $N \in \mathbb{N}$, let X_N be a set and $F_N : X_N \rightarrow \mathbb{R}$ a function. Furthermore, let Y be a topological space with topology τ and $q = (q_N)_{N \in \mathbb{N}}$ a sequence of embedding maps $q_N : X_N \rightarrow Y$. Then, F_N is said to $\Gamma(q, \tau^-)$ -converge to a function $F : Y \rightarrow \mathbb{R}$ at $y \in Y$, if

1. *lim inf-condition*: For each sequence made of $x_N \in X_N$ such that $q_N(x_N) \xrightarrow{\tau} y$ for $N \rightarrow \infty$,

$$F(y) \leq \liminf_{N \rightarrow \infty} F_N(x_N).$$

2. *lim sup-condition*: There is a sequence made of $x_N \in X_N$ such that $q_N(x_N) \xrightarrow{\tau} y$ for $N \rightarrow \infty$ and

$$F(y) \geq \limsup_{N \rightarrow \infty} F_N(x_N).$$

Furthermore, we say that the F_N $\Gamma(q, \tau^-)$ -converge on a set $\mathcal{D} \subset Y$ if the above is true for all $y \in \mathcal{D}$ and we call the sequence $(F_N)_{N \in \mathbb{N}}$ equi-coercive, if for every $c \in \mathbb{R}$, there is a compact set $K \subset Y$ such that $q_N(\{x : F_N(x) \leq c\}) \subset K$ uniformly with respect to $N \in \mathbb{N}$.

Remark 5.9 The main consequence of Γ -convergence, which is of interest to us, is the convergence of minimizers. This remains true also for the $\Gamma(q, \tau^-)$ -convergence, see [3, Proposition 2.4].

Here, we are going to consider

$$Y := \mathcal{P}(\mathbb{R}^d) \times BV(\mathbb{R}^d) \tag{5.17}$$

with the corresponding product topology of narrow convergence and BV weak- $*$ -convergence (actually L^1 -convergence suffices),

$$X_N := \mathcal{P}^N(\mathbb{R}^d), \quad q_N(\mu) := (\mu, Q_{h(N)}[\mu]).$$

and consider the limit $\widehat{\mathcal{E}}^\lambda$ to be defined on the diagonal

$$\mathcal{D} := \left\{ (\mu, \mu) : \mu \in \mathcal{P}(\mathbb{R}^d) \cap BV(\mathbb{R}^d) \right\}.$$

Since we will be extracting convergent subsequences of pairs $(\mu_N, Q_{h(N)}[\mu])$ in order to obtain existence of minimizers, we need the following lemma to ensure that the limit is in the diagonal set \mathcal{D} .

Lemma 5.10 (Consistency of the embedding $Q_{h(N)}$) *If $(\mu_N)_N$ is a sequence such that $\mu_N \in \mathcal{P}^N(\mathbb{R}^d)$, $\mu_N \rightarrow \mu \in \mathcal{P}(\mathbb{R}^d)$ narrowly and $Q_{h(N)}[\mu_N] \rightarrow \tilde{\mu} \in BV(\mathbb{R}^d)$ in $L^1(\mathbb{R}^d)$ as well, as $h(N) \rightarrow 0$, then $\mu = \tilde{\mu}$.*

Proof To show $\mu = \tilde{\mu}$, by the metrizability of \mathcal{P} it suffices to show that $Q_{h(N)}[\mu_N] \rightarrow \mu$ narrowly. For this, as in the proof of Lemma 5.7, we can restrict ourselves to

test convergence of the integral against bounded and Lipschitz-continuous functions. Hence, let $\varphi \in C_b(\mathbb{R}^d) \cap \text{Lip}(\mathbb{R}^d)$ with Lipschitz constant L . Then,

$$\begin{aligned} & \left| \int_{\mathbb{R}^d} \varphi(x) Q_{h(N)}[\mu_N](x) dx - \int_{\mathbb{R}^d} \varphi(x) d\mu(x) \right| \\ & \leq \left| \int_{\mathbb{R}^d} \varphi(x) K_{h(N)} * \mu_N(x) dx - \int_{\mathbb{R}^d} \varphi(x) d\mu_N(x) \right| \\ & \quad + \left| \int_{\mathbb{R}^d} \varphi(x) d\mu_N(x) - \int_{\mathbb{R}^d} \varphi(x) d\mu(x) \right|, \end{aligned}$$

where the second term goes to zero by $\mu_N \rightarrow \mu$ narrowly. For the first term, by Fubini we get that

$$\int_{\mathbb{R}^d} \varphi(x) K_{h(N)} * \mu_N(x) dx = \int_{\mathbb{R}^d} (\varphi * K_{h(N)}(-\cdot))(x) d\mu_N(x)$$

and therefore

$$\begin{aligned} & \left| \int_{\mathbb{R}^d} \varphi(x) K_{h(N)} * \mu_N(x) dx - \int_{\mathbb{R}^d} \varphi(x) \mu_N(x) dx \right| \\ & = \left| \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (\varphi(x+y) - \varphi(x)) K_{h(N)}(y) dy d\mu_N(x) \right| \\ & = \left| \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (\varphi(x+h(N)y) - \varphi(x)) K(y) dy d\mu_N(x) \right| \\ & \leq Lh \|K\|_{L^1} \mu_N(\mathbb{R}^d) \rightarrow 0, \quad N \rightarrow \infty \end{aligned}$$

by $h(N) \rightarrow 0$, proving $Q_{h(N)}[\mu_N] \rightarrow \mu$ and hence the claim. □

Theorem 5.11 (Consistency of the kernel estimate) *Under the assumption (5.11) on $h(N)$, the functionals $(\widehat{\mathcal{E}}_N^\lambda)_{N \in \mathbb{N}}$ are equi-coercive and*

$$\widehat{\mathcal{E}}_N^\lambda \xrightarrow{\Gamma(q, \tau^-)} \widehat{\mathcal{E}}^\lambda \quad \text{for } N \rightarrow \infty,$$

with respect to the topology of Y defined above, i.e., weak convergence of μ_N together with L^1 -convergence of $Q_{h(N)}[\mu_N]$. In particular, every sequence of minimizers of $\widehat{\mathcal{E}}_N^\lambda$ admits a subsequence converging to a minimizer of $\widehat{\mathcal{E}}^\lambda$.

Proof **lim inf-condition:** This follows from the lower semi-continuity of $\widehat{\mathcal{E}}$ and $\mu \mapsto |D\mu|(\mathbb{R}^d)$ with respect to narrow convergence and L^1 -convergence, respectively.

lim sup-condition: We use a diagonal argument to find the recovery sequence. An arbitrary $\mu \in BV(\mathbb{R}^d) \cap \mathcal{P}(\mathbb{R}^d)$ can be approximated by Proposition 3.9 by probability measures μ_n with existing second moment such that $\widehat{\mathcal{E}}[\mu_n] \rightarrow \widehat{\mathcal{E}}[\mu]$, namely

$$\mu_n = \widehat{\eta}_{n-1} \cdot \mu + \left(1 - \widehat{\eta}_{n-1} \cdot \mu(\mathbb{R}^d)\right) \delta_0.$$

By Lemma 5.1, we can also smooth the approximating measures by convolution with a Gaussian $\eta_{\varepsilon(n)}$ to get a narrowly convergent sequence $\mu'_n \rightarrow \mu$,

$$\mu'_n = \eta_{\varepsilon(n)} * \mu_n = \eta_{\varepsilon(n)} * (\widehat{\eta}_{n-1} \cdot \mu) + \left(1 - (\widehat{\eta}_{n-1} \cdot \mu)(\mathbb{R}^d)\right) \eta_{\varepsilon(n)},$$

while still keeping the continuity in $\widehat{\mathcal{E}}$. Since $(1 - (\widehat{\eta}_{n-1} \cdot \mu)(\mathbb{R}^d)) \rightarrow 0$, we can replace its factor $\eta_{\varepsilon(n)}$ by η_1 to get

$$\mu''_n = \eta_{\varepsilon(n)} * (\widehat{\eta}_{n-1} \cdot \mu) + \left(1 - (\widehat{\eta}_{n-1} \cdot \mu)(\mathbb{R}^d)\right) \eta_1,$$

and still have convergence and continuity in $\widehat{\mathcal{E}}$. These μ''_n can then be (strictly) cut-off by a smooth cut-off function χ_M such that

$$\begin{aligned} \chi_M(x) &= 1 \quad \text{for } |x| \leq M, \\ \chi_M(x) &\in [0, 1] \quad \text{for } M < |x| < M + 1, \\ \chi_M(x) &= 0 \quad \text{for } |x| \geq M + 1. \end{aligned}$$

Superfluous mass can then be absorbed in a normalized version of χ_1 . This process yields

$$\mu'''_n = \chi_{M(n)} \cdot \mu''_n + (1 - \chi_{M(n)} \cdot \mu''_n)(\mathbb{R}^d) \frac{\chi_1}{\|\chi_1\|_1},$$

which, for fixed n and $M(n) \rightarrow \infty$, is convergent in the 2-Wasserstein topology, hence we can keep the continuity in $\widehat{\mathcal{E}}$ by choosing $M(n)$ large enough.

Moreover, the sequence μ'''_n is also strictly convergent in BV : for the L^1 -convergence, we apply the dominated convergence theorem for $M(n) \rightarrow \infty$ when considering μ'''_n , and the approximation property of the Gaussian mollification of L^1 -functions for μ'''_n . Similarly, for the convergence of the total variation, consider

$$\begin{aligned} & \left| |D\mu'''_n|(\mathbb{R}^d) - |D\mu|(\mathbb{R}^d) \right| \\ & \leq \left| \int_{\mathbb{R}^d} \chi_{M(n)}(x) |D\mu''_n(x)| \, dx - \int_{\mathbb{R}^d} |D\mu''_n(x)| \, dx \right| \end{aligned} \tag{5.18}$$

$$+ \int_{\mathbb{R}^d} |\nabla \chi_{M(n)}(x)| \mu''_n(x) \, dx \tag{5.19}$$

$$+ \left| |D\mu''_n(x)| - |D\mu|(\mathbb{R}^d) \right| \tag{5.20}$$

$$+ (1 - \chi_{M(n)} \cdot \mu''_n)(\mathbb{R}^d) \frac{\|\nabla \chi_1\|_1}{\|\chi_1\|_1}, \tag{5.21}$$

where the terms (5.18), (5.19) and (5.21) tend to 0 for $M(n)$ large enough by dominated convergence. For the remaining term (5.20), we have

$$\begin{aligned} & \left| |D\mu_n''| - |D\mu|(\mathbb{R}^d) \right| \\ & \leq \left| |\eta_{\varepsilon(n)} * D(\widehat{\eta}_n \cdot \mu)|(\mathbb{R}^d) - |D(\widehat{\eta}_n \cdot \mu)|(\mathbb{R}^d) \right| \end{aligned} \tag{5.22}$$

$$+ \int_{\mathbb{R}^d} |\nabla \widehat{\eta}_n(x)| \, d\mu(x) \tag{5.23}$$

$$+ \int_{\mathbb{R}^d} (1 - \widehat{\eta}_n(x)) \, d|D\mu|(x) \tag{5.24}$$

$$+ \left(1 - (\widehat{\eta}_{n-1} \cdot \mu)(\mathbb{R}^d) \right) |D\eta_1|(\mathbb{R}^d). \tag{5.25}$$

Here, all terms vanish as well: (5.22) for $\varepsilon(n)$ large enough by the approximation property of the Gaussian mollification for BV -functions and (5.23), (5.24) and (5.25) by the dominated convergence theorem for $n \rightarrow \infty$. Finally, Lemma 5.7 applied to the μ_n''' yields the desired sequence of point approximations.

Equi-coercivity and existence of minimizers: Equi-coercivity and compactness strong enough to ensure the existence of minimizers follow from the coercivity and compactness of level sets of $\widehat{\mathcal{E}}$ and by $\|Q_{h(N)}(\mu_N)\|_{L^1} = 1$ together with compactness arguments in BV , similar to Proposition 5.2. Since Lemma 5.10 ensures that the limit is in the diagonal space \mathcal{D} , standard Γ -convergence arguments then yield the convergence of minimizers. □

5.2.2 Discretization by Point-Differences

In one dimension, the geometry is sufficiently simple to avoid the use of kernel density estimators and to allow us to explicitly see the intuitive effect the total variation regularization has on point masses (similar to the depiction in Fig. 2 in the previous section). In particular, formula (5.26) below shows that the total variation acts as an additional attractive-repulsive force which tends to promote equi-spacing between the points masses.

In the following, let $d = 1$ and $\lambda > 0$ fixed.

Let $N \in \mathbb{N}$, $N \geq 2$ and $\mu_N \in \mathcal{P}^N(\mathbb{R})$ with

$$\mu_N = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \quad \text{for some } x_i \in \mathbb{R}.$$

Using the ordering on \mathbb{R} , we can assume the $(x_i)_i$ to be ordered, which allows us to associate to μ_N a unique vector

$$x := x(\mu_N) := (x_1, \dots, x_N), \quad x_1 \leq \dots \leq x_N.$$

If $x_i \neq x_j$ for all $i, j \in \{1, \dots, N\}, i \neq j$, we can further define an L^1 -function which is piecewise-constant by

$$\tilde{Q}_N[\mu_N] := \frac{1}{N} \sum_{i=2}^N \frac{1}{x_i - x_{i-1}} 1_{[x_{i-1}, x_i]}$$

and compute explicitly the total variation of its weak derivative to be

$$\begin{aligned} & |D\tilde{Q}_N[\mu_N]|(\mathbb{R}) \\ &= \frac{1}{N} \left[\sum_{i=2}^{N-1} \left| \frac{1}{x_{i+1} - x_i} - \frac{1}{x_i - x_{i-1}} \right| + \frac{1}{x_2 - x_1} + \frac{1}{x_N - x_{N-1}} \right], \end{aligned} \quad (5.26)$$

if no two points are equal, and ∞ otherwise. This leads us to the following definition of the regularized functional using piecewise constant functions:

$$\begin{aligned} \mathcal{P}_\times^N(\mathbb{R}) &:= \left\{ \mu \in \mathcal{P}^N(\mathbb{R}) : \mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \text{ with } x_i \neq x_j \text{ for } i \neq j \right\}, \\ \widehat{\mathcal{E}}_{N,\text{pwc}}^\lambda[\mu] &:= \begin{cases} \widehat{\mathcal{E}}[\mu] + \lambda |D\tilde{Q}_N[\mu]|(\mathbb{R}), & \mu \in \mathcal{P}_\times^N(\mathbb{R}); \\ \infty, & \mu \in \mathcal{P}^N(\mathbb{R}) \setminus \mathcal{P}_\times^N(\mathbb{R}). \end{cases} \end{aligned}$$

Remark 5.12 The functions $\tilde{Q}_N[\mu_N]$ as defined above are not probability densities, but instead have mass $(N-1)/N$.

We shall again prove $\Gamma(q, \tau^-)$ -convergence as in Sect. 5.2.1, this time with the embeddings q_N given by \tilde{Q}_N . The following lemma yields the necessary recovery sequence.

Lemma 5.13 *If $\mu \in \mathcal{P}_c(\mathbb{R}) \cap C_c^\infty(\mathbb{R})$ is the density of a compactly supported probability measure, then there is a sequence $\mu_N \in \mathcal{P}^N(\mathbb{R})$, $N \in \mathbb{N}_{\geq 2}$ such that*

$$\mu_N \rightarrow \mu \text{ narrowly for } N \rightarrow \infty$$

and

$$\tilde{Q}_N[\mu_N] \rightarrow \mu \text{ in } L^1(\mathbb{R}), \quad |D\tilde{Q}_N[\mu_N]|(\mathbb{R}) \rightarrow \int_{\mathbb{R}} |\mu'(x)| dx \text{ for } N \rightarrow \infty.$$

Proof 1. *Definition and narrow convergence:* Let $\text{supp} \mu \subset [-R_\mu, R_\mu]$ and define the vector $x^N \in \mathbb{R}^N$ as an N th quantile of μ , i.e.,

$$\int_{x_{i-1}^N}^{x_i^N} \mu(x) dx = \frac{1}{N} \quad \text{with } x_{i-1}^N < x_i^N \text{ for all } i = 1, \dots, N-1,$$

where we set $x_0^N = -R_\mu$ and $x_N^N = R_\mu$. Narrow convergence of the corresponding measure then follows by the same arguments used in the proof of Lemma 5.7.

2. *L¹-convergence*: We want to use the dominated convergence theorem: Let $x \in \mathbb{R}$ with $\mu(x) > 0$. Then, by the continuity of μ , there are $x_{i-1}^N(x), x_i^N(x)$ such that $x \in [x_{i-1}^N(x), x_i^N(x)]$ and

$$\begin{aligned} \mu(x) - \tilde{Q}_N[\mu_N](x) &= \mu(x) - \frac{1}{N(x_i^N(x) - x_{i-1}^N(x))} \\ &= \mu(x) - \frac{1}{x_i^N(x) - x_{i-1}^N(x)} \int_{x_{i-1}^N(x)}^{x_i^N(x)} \mu(y) \, dy. \end{aligned} \tag{5.27}$$

Again by $\mu(x) > 0$ and the continuity of μ ,

$$x_i^N(x) - x_{i-1}^N(x) \rightarrow 0 \quad \text{for } N \rightarrow \infty,$$

and therefore

$$\tilde{Q}_N[\mu_N](x) \rightarrow \mu(x) \quad \text{for all } x \text{ such that } \mu(x) > 0.$$

On the other hand, if we consider an $x \in [-R_\mu, R_\mu]$ such that $x \notin \text{supp}\mu$, say $x \in [a, b]$, where the interval $[a, b]$ is such that $\mu(\xi) = 0$ for all $\xi \in [a, b]$, and again denote by $x_{i-1}^N(x), x_i^N(x)$ the two quantiles for which $x \in [x_{i-1}^N(x), x_i^N(x)]$, then $x_i^N(x) - x_{i-1}^N(x)$ stays bounded from below because $x_{i-1}^N(x) \leq a$ and $x_i^N(x) \geq b$, together with $N \rightarrow \infty$ implying that for such an x ,

$$\tilde{Q}_N[\mu_N](x) = \frac{1}{N(x_i^N - x_{i-1}^N)} \leq \frac{1}{N(b - a)} \rightarrow 0.$$

Taking into account that μ can vanish on $\text{supp}\mu$ only on a subset of measure 0, we thus have

$$\tilde{Q}_N[\mu_N](x) \rightarrow \mu(x) \quad \text{for almost every } x \in \mathbb{R}.$$

Furthermore, by (5.27) and the choice of the $(x_i^N)_i$, we can estimate the difference by

$$|\mu(x) - \tilde{Q}_N[\mu_N](x)| \leq 2 \|\mu\|_\infty \cdot 1_{[-R_\mu, R_\mu]}(x),$$

yielding an integrable dominating function for $|\mu(x) - \tilde{Q}_N[\mu_N](x)|$ and therefore justifying the L^1 -convergence

$$\int_{\mathbb{R}} |\mu(x) - \tilde{Q}_N[\mu_N](x)| \, dx \rightarrow 0, \quad N \rightarrow \infty.$$

3. *Strict BV-convergence*: For strict convergence of $\tilde{Q}_N[\mu_N]$ to μ , we additionally have to check that $\limsup_{N \rightarrow \infty} |D\tilde{Q}_N[\mu_N]|(\mathbb{R}) \leq |D\mu|(\mathbb{R})$. To this end, consider

$$\begin{aligned}
 & |D\tilde{Q}_N[\mu_N]|(\mathbb{R}) \\
 &= \sum_{i=2}^{N-1} \left| \frac{1}{N} \frac{1}{x_{i+1}^N - x_i^N} - \frac{1}{N} \frac{1}{x_i^N - x_{i-1}^N} \right| + \frac{1}{N(x_2^N - x_1^N)} + \frac{1}{N(x_N^N - x_{N-1}^N)} \\
 &= \sum_{i=2}^{N-1} \left| \frac{1}{x_{i+1}^N - x_i^N} \int_{x_i^N}^{x_{i+1}^N} \mu(x) \, dx - \frac{1}{x_i^N - x_{i-1}^N} \int_{x_{i-1}^N}^{x_i^N} \mu(x) \, dx \right| \\
 &\quad + \frac{1}{x_2^N - x_1^N} \int_{x_1^N}^{x_2^N} \mu(x) \, dx + \frac{1}{x_N^N - x_{N-1}^N} \int_{x_{N-1}^N}^{x_N^N} \mu(x) \, dx \\
 &= \sum_{i=1}^N |\mu(t_{i+1}) - \mu(t_i)|
 \end{aligned}$$

for $t_i \in [x_i^N, x_{i-1}^N], i = 2, \dots, N$ chosen by the mean value theorem (for integration) and t_1, t_{N+1} denoting $-R_\mu$ and R_μ , respectively. Hence,

$$|D\tilde{Q}_N[\mu_N]|(\mathbb{R}) \leq \sup \left\{ \sum_{i=1}^{n-1} |\mu(t_{i+1}) - \mu(t_i)| : n \geq 2, t_1 < \dots < t_n \right\} = V(\mu),$$

the pointwise variation of μ , and the claim now follows from $V(\mu) = |D\mu|(\mathbb{R})$ by [1, Theorem 3.28], because by the smoothness of μ , it is a good representative of its equivalence class in $BV(\mathbb{R})$, i.e., one for which the pointwise variation coincides with the measure theoretical one. □

As in the previous section, we have to verify that a limit point of a sequence $(\mu_N, \tilde{Q}_N[\mu_N])$ is in the diagonal \mathcal{D} .

Lemma 5.14 (Consistency of the embedding \tilde{Q}_N) *Let $(\mu_N)_N$ be a sequence where $\mu_N \in \mathcal{P}^N(\mathbb{R}), \mu_N \rightarrow \mu$ narrowly and $\tilde{Q}_N[\mu_N] \rightarrow \tilde{\mu}$ in $L^1(\mathbb{R})$. Then $\mu = \tilde{\mu}$.*

Proof Denote the cumulative distribution functions \tilde{F}_N, F_N , and F of $\tilde{Q}_N[\mu_N], \mu_N$, and μ , respectively. We can deduce $\mu = \tilde{\mu}$ if $\tilde{F}_N(x) \rightarrow F(x)$ for every $x \in \mathbb{R}$ (even if the measures $\tilde{Q}_N[\mu_N]$ have only mass $(N - 1)/N$, this is enough to show that the limit measures have to coincide, for example by rescaling the measures to have mass 1). Note that the construction of $\tilde{Q}_N[\mu_N]$ precisely consists of replacing the piecewise constant functions F_N by piecewise affine functions interpolating between the points $(x_i^N)_i$. Now, taking into account that the jump size $F_N(x_i^N) - F_N(x_{i-1}^N)$ is always $1/N$ we see that

$$\begin{aligned}
 |\tilde{F}_N(x) - F(x)| &\leq |\tilde{F}_N(x) - F_N(x)| + |F_N(x) - F(x)| \\
 &\leq \frac{1}{N} + |F_N(x) - F(x)| \rightarrow 0, \quad N \rightarrow \infty,
 \end{aligned}$$

which is the claimed convergence. □

Theorem 5.15 (Consistency of $\widehat{\mathcal{E}}_{N,\text{pwc}}^\lambda$) *Assume $d = 1$. Then for $N \rightarrow \infty$ we have $\widehat{\mathcal{E}}_{N,\text{pwc}}^\lambda \xrightarrow{\Gamma(q, \tau^-)} \widehat{\mathcal{E}}^\lambda$ with respect to the topology of Y in (5.17), i.e., the topology induced by the narrow convergence together with the L^1 -convergence of the associated densities, and the sequence $(\widehat{\mathcal{E}}_{N,\text{pwc}}^\lambda)_N$ is equi-coercive. In particular, every sequence of minimizers of $\widehat{\mathcal{E}}_{N,\text{pwc}}^\lambda$ admits a subsequence converging to a minimizer of $\widehat{\mathcal{E}}^\lambda$.*

Proof 1. *lim inf-condition:* Let $\mu_N \in \mathcal{P}_N(\mathbb{R})$ and $\mu \in BV(\mathbb{R}) \cap \mathcal{P}(\mathbb{R})$ with $\mu_N \rightarrow \mu$ narrowly and $\widetilde{Q}_N[\mu_N] \rightarrow \mu$ in L^1 . Then,

$$\liminf_{N \rightarrow \infty} \widehat{\mathcal{E}}_{N,\text{pwc}}^\lambda[\mu_N] = \liminf_{N \rightarrow \infty} [\widehat{\mathcal{E}}[\mu_N] + |D\widetilde{Q}_N[\mu_N]|(\mathbb{R})] \geq \widehat{\mathcal{E}}[\mu] + |D\mu|(\mathbb{R})$$

by the lower semi-continuity of the summands with respect to the involved topologies.

2. *lim sup-condition:* We use the same diagonal argument used in the proof of Theorem 5.11, replacing the final application of Lemma 5.7 there by Lemma 5.13, which serves the same purpose, but uses the point differences instead of the kernel estimators.
3. *Equi-coercivity and existence of minimizers:* The coercivity follows analogously to the proof of Theorem 5.11, which also justifies the existence of minimizers for each N . The convergence of minimizers to an element of \mathcal{D} then follows by standard arguments together with Lemma 5.14. □

Remark 5.16 In both cases, instead of working with two different topologies, we could also consider

$$\widehat{\mathcal{E}}_{N,\text{alt}}^\lambda := \widehat{\mathcal{E}}[Q[\mu]] + \lambda |DQ[\mu]|(\mathbb{R}^d),$$

for a given embedding Q , which in the case of point differences would have to be rescaled to keep mass 1. Then, we would obtain the same results by identical arguments, but without the need to worry about narrow convergence separately, since it is implied by the L^1 -convergence of $Q[\mu_N]$.

6 Numerical Experiments

In this section, we shall show numerical examples of approximate computations of minimizers of $\widehat{\mathcal{E}}^\lambda$ and $\widehat{\mathcal{E}}_N^\lambda$ in one dimension in order to numerically demonstrate the Γ -convergence result in Theorem 5.11.

6.1 Grid Approximation

By Theorem 5.11, we know that $\widehat{\mathcal{E}}_N^\lambda \xrightarrow{\Gamma} \widehat{\mathcal{E}}^\lambda$, telling us that the particle minimizers of $\widehat{\mathcal{E}}^\lambda$ will be close to a minimizer of the functional $\widehat{\mathcal{E}}^\lambda$, which will be a BV function. Therefore, we would like to compare the particle minimizers to minimizers which were

computed by using a more classical approximation method which in contrast maintains the underlying BV structure. One such approach is to approximate a function in BV by piecewise constant functions on an equispaced discretization of the interval $\Omega = [0, 1]$. Denoting the restriction of $\widehat{\mathcal{E}}^\lambda$ to the space of these functions on a grid with N points by $\widehat{\mathcal{E}}^\lambda_{N,\text{grid}}$, it can be seen that we have $\widehat{\mathcal{E}}^\lambda_{N,\text{grid}} \xrightarrow{\Gamma} \widehat{\mathcal{E}}^\lambda$, hence it makes sense to compare minimizers of $\widehat{\mathcal{E}}^\lambda_{N,\text{grid}}$ and $\widehat{\mathcal{E}}^\lambda_N$ for large N .

If we denote by $u \in \mathbb{R}^N$ the approximation to μ and by $w \in \mathbb{R}^N$ the one to ω , then the problem to minimize $\widehat{\mathcal{E}}^\lambda_{N,\text{grid}}$ takes the form

$$\begin{aligned} &\text{minimize} && (u - w)^T A_{q,\Omega}(u - w) + \lambda \sum_{i=1}^{N-1} |u_{i+1} - u_i| \\ &\text{subject to} && u \geq 0, \quad \sum_{i=1}^m u_i = N, \end{aligned} \tag{6.1}$$

where $A_{q,\Omega}$ is the corresponding discretization matrix of the quadratic integral functional $\widehat{\mathcal{E}}$, which is positive definite on the set $\{v : \sum v = 0\}$ by the theory of Appendix. To be more explicit, by defining $\mu = \sum_{i=1}^N u_i \chi_{[\frac{i-1}{N}, \frac{i}{N}]}$ and $\omega = \sum_{i=1}^N w_i \chi_{[\frac{i-1}{N}, \frac{i}{N}]}$, where χ_D is the characteristic function of the set D , we obtain from (3.9) and Theorem 3.6 that

$$\begin{aligned} \widehat{\mathcal{E}}(\mu) &= -\frac{1}{2} \int_0^1 \int_0^1 |x - y|^q (\mu(x) - \omega(x))(\mu(y) - \omega(y)) dx dy \\ &= \sum_{i=1}^N \sum_{j=1}^N (u_i - w_i)(u_j - w_j) \underbrace{\left(-\frac{1}{2} \int_0^1 \int_0^1 |x - y|^q \chi_{[\frac{i-1}{N}, \frac{i}{N}]}(x) \chi_{[\frac{j-1}{N}, \frac{j}{N}]}(y) dx dy \right)}_{:= (A_{q,\Omega})_{ij}} \\ &= (u - w)^T A_{q,\Omega}(u - w). \end{aligned}$$

Solving the last condition in (6.1) $\sum_{i=1}^N u_i = N$ for one coordinate of u , we get a reduced matrix $\widetilde{A}_{q,\Omega}$ which is positive definite. Together with the convex approximation term to the total variation, problem (6.1) is a convex optimization problem which can be easily solved, e.g., by the CVX package [11, 25].

As model cases to study the influence of the total variation, the following data were considered (see Fig. 4a, b) for their visual representation)

1. $\omega_1 = 4 \cdot 1_{[0,2,0,4]} + 40 \cdot 1_{[0,6,0,605]}$, the effect of the regularization being that the second bump gets smaller and more spread out with increasing parameter λ , see Fig. 5;
2. $\widetilde{\omega}_2 = \frac{1}{1 + \|\eta\|_{>0}} (\omega_2 + \eta)_{>0}$, where η is Gaussian noise affecting the reference measure $\omega_2 = 5 \cdot 1_{[0,2,0,4]}$, where we cut off the negative part and re-normalized the datum to get a probability measure. The effect of the regularization here is a filtering of the noise, see Fig. 6.

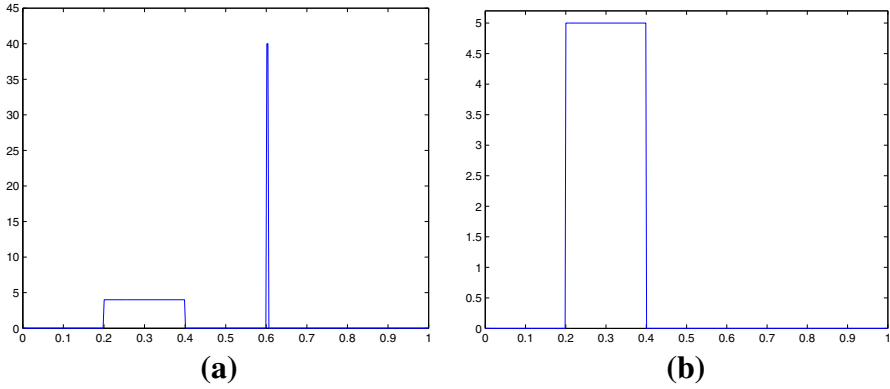


Fig. 4 The data **a** ω_1 . **b** $\omega_2, \omega_1, \omega_2$

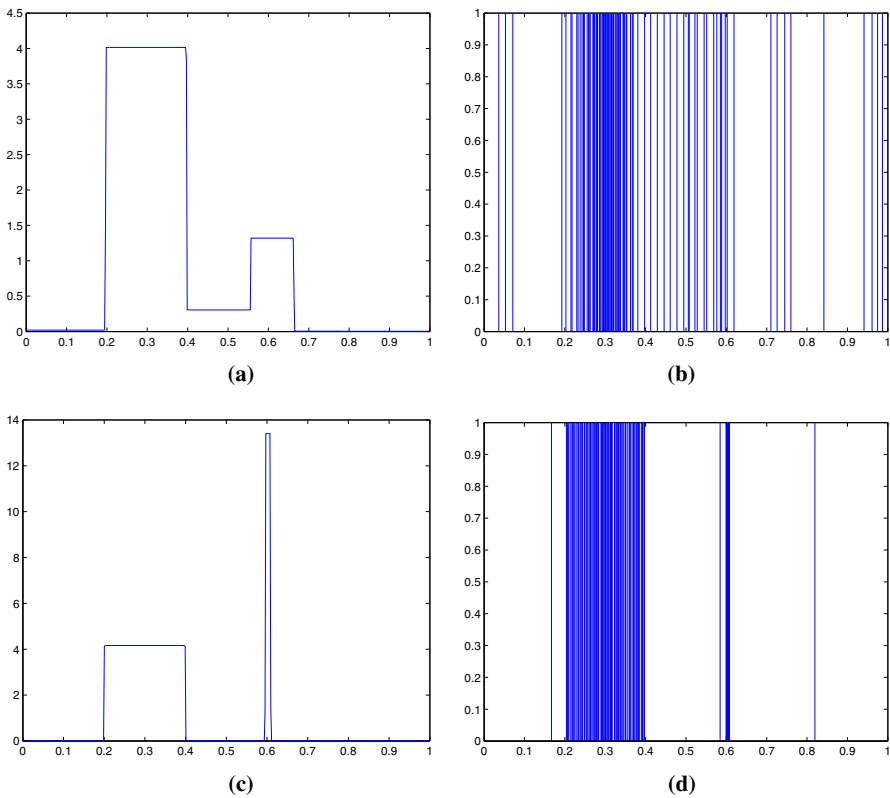


Fig. 5 Minimizers u of (6.1) and minimizers μ_N of $\widehat{\mathcal{E}}_N^\lambda$ for ω_1 as in Fig. 4a and parameters $q = 1.0, N = 100$. **a** $u \approx \mu \in L^1, \lambda = 10^{-4}$. **b** Particles supporting $\mu_N, \lambda = 10^{-4}$. **c** $u \approx \mu \in L^1, \lambda = 10^{-6}$. **d** Particles supporting $\mu_N, \lambda = 10^{-6}$

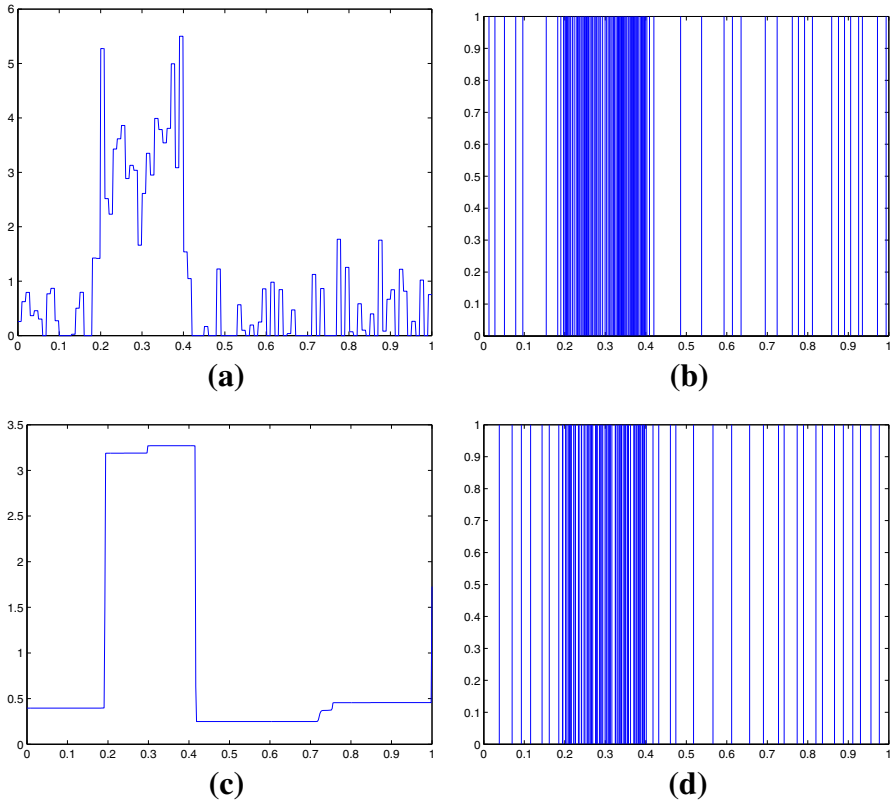


Fig. 6 Minimizers u of (6.1) and minimizers μ_N of $\widehat{\mathcal{E}}_N^\lambda$ for a projection of $\omega_2 + \eta$ as in Fig. 4b and parameters $q = 1.5$, $N = 100$. **a** $u \approx \mu \in L^1$, $\lambda = 0$. **b** Particles supporting μ_N , $\lambda = 0$. **c** $u \approx \mu \in L^1$, $\lambda = 10^{-5}$. **d** Particles supporting μ_N , $\lambda = 10^{-5}$

6.2 Particle Approximation

The solutions in the particle case were computed by the MATLAB optimization toolbox, in particular the Quasi-Newton method available via the FMINUNC command. The corresponding function evaluations were computed directly in the case of the repulsion functional and by a trapezoidal rule in the case of the attraction term. For the kernel estimator, we used the one sketched in Fig. 2,

$$K(x) = (1 - |x|) \cdot 1_{[-1,1]}(x), \quad x \in \mathbb{R}.$$

6.3 Results

As for the L^1 case, we see that the total variation regularization works well as a regularizer and allows us to recover the original profile from a datum disturbed by noise.

When it comes to the particle case, we numerically confirm the theoretical results of convergence for $N \rightarrow \infty$ of Sect. 5.2, since the minimizers of the particle system behave roughly like the quantizers of the problem in L^1 .

7 Conclusion

Beside the relatively simple results on existence for asymmetric exponents $q_a \neq q_r$ in Sect. 2, the Fourier representation of Sect. 3, building upon the theory of conditionally positive semi-definite functions concisely recalled in Appendix, proved essential to establish the well-posedness of the problem for equal exponents $1 \leq q_a = q_r < 2$, in terms of the lower semi-continuous envelope of the energy \mathcal{E} . This allowed us to use classical tools of calculus of variations, in particular the machinery of Γ -convergence, to prove statements concerning the consistency of the particle approximation, Theorem 3.13, and the moment bound, Theorem 4.1, which would be otherwise not at all obvious when just considering the original spatial definition of \mathcal{E} . Moreover, it enabled us to easily analyze the regularized version of the functional in Sect. 5, which on the particle level allowed us to present a novel interpretation of the total variation as a nonlinear attractive-repulsive potential, translating the regularizing effect of the total variation in the continuous case into an energy which promotes a configuration of the particles which is as homogeneous as possible.

Acknowledgments Massimo Fornasier is supported by the ERC-Starting Grant for the project “High-Dimensional Sparse Optimal Control”. Jan-Christian Hütter acknowledges the partial financial support of the START-Project “Sparse Approximation and Optimization in High-Dimensions” during the early preparation of this work.

Appendix: Conditionally Positive Definite Functions

In order to compute the Fourier representation of the energy functional \mathcal{E} in Sect. 3.1.3, we used the notion of *generalized Fourier transforms* and *conditionally positive definite functions* from [36], which we shall briefly recall here for the sake of completeness. In fact, the main result reported below, Theorem 7.6 is shown in a slightly modified form with respect to [36, Theorem 8.16], to allow us to prove the moment bound in Sect. 4. The representation formula (3.10) is a consequence of Theorem 7.4 below, which serves as a characterization formula in the theory of conditionally positive definite functions.

Definition 7.1 [36, Definition 8.1] Let $\mathbb{P}_k(\mathbb{R}^d)$ denote the set of polynomial functions on \mathbb{R}^d of degree less or equal than k . We call a continuous function $\Phi: \mathbb{R}^d \rightarrow \mathbb{C}$ *conditionally positive semi-definite of order m* if for all $N \in \mathbb{N}$, pairwise distinct points $x_1, \dots, x_N \in \mathbb{R}^d$, and $\alpha \in \mathbb{C}^N$ with

$$\sum_{j=1}^N \alpha_j p(x_j) = 0, \quad \text{for all } p \in \mathbb{P}_{m-1}(\mathbb{R}^d), \tag{7.1}$$

the quadratic form given by $(\Phi(x_j - x_k))_{jk}$ is non-negative, i.e.,

$$\sum_{j,k=1}^N \alpha_j \overline{\alpha_k} \Phi(x_j - x_k) \geq 0.$$

Moreover, we call Φ *conditionally positive definite of order m* if the above inequality is strict for $\alpha \neq 0$.

Generalized Fourier Transform

When working with distributional Fourier transforms, which can serve to characterize the conditionally positive definite functions defined above, it can be opportune to reduce the standard Schwartz space \mathcal{S} to functions which in addition to the polynomial decay for large arguments also exhibit a certain decay for small ones. In this way, one can elegantly neglect singularities in the Fourier transform at 0, which could otherwise arise.

Definition 7.2 (*Restricted Schwartz class \mathcal{S}_m*) [36, Definition 8.8] Let \mathcal{S} be the Schwartz space of functions in $C^\infty(\mathbb{R}^d)$ which for $|x| \rightarrow \infty$ decay faster than any fixed polynomial. Then, for $m \in \mathbb{N}$, we denote by \mathcal{S}_m the subset of those functions γ in \mathcal{S} which additionally fulfill

$$\gamma(\xi) = O(|\xi|^m) \quad \text{for } \xi \rightarrow 0. \quad (7.2)$$

Furthermore, we shall call an (otherwise arbitrary) function $\Phi: \mathbb{R}^d \rightarrow \mathbb{C}$ *slowly increasing* if there is an $m \in \mathbb{N}$ such that

$$\Phi(x) = O(|x|^m) \quad \text{for } |x| \rightarrow \infty.$$

Definition 7.3 (*Generalized Fourier transform*) [36, Definition 8.9] For $\Phi: \mathbb{R}^d \rightarrow \mathbb{C}$ continuous and slowly increasing, we call a measurable function $\widehat{\Phi} \in L^2_{\text{loc}}(\mathbb{R}^d \setminus \{0\})$ the *generalized Fourier transform* of Φ if there exists a multiple of $\frac{1}{2}$, $m = \frac{1}{2}n$, $n \in \mathbb{N}_0$ such that

$$\int_{\mathbb{R}^d} \Phi(x) \widehat{\gamma}(x) dx = \int_{\mathbb{R}^d} \widehat{\Phi}(\xi) \gamma(\xi) d\xi \quad \text{for all } \gamma \in \mathcal{S}_{2m}. \quad (7.3)$$

Then, we call m the *order* of $\widehat{\Phi}$.

Note that the order here is defined in terms of $2m$ instead of m , which is why we would like to allow for multiples of $\frac{1}{2}$.

Representation Formula for Conditionally Positive Definite Functions

Theorem 7.4 [36, Corollary 8.13] Let $\Phi: \mathbb{R}^d \rightarrow \mathbb{C}$ be a continuous and slowly increasing function with a non-negative, non-vanishing generalized Fourier transform $\widehat{\Phi}$ of order m that is continuous on $\mathbb{R}^d \setminus \{0\}$. Then, for pairwise distinct points

$x_1, \dots, x_N \in \mathbb{R}^d$ and $\alpha \in \mathbb{C}^N$ which fulfill condition (7.1), i.e.,

$$\sum_{j=1}^N \alpha_j p(x_j) = 0, \text{ for all } p \in \mathbb{P}_{m-1}(\mathbb{R}^d),$$

we have

$$\sum_{j,k=1}^N \alpha_j \bar{\alpha}_k \Phi(x_j - x_k) = \int_{\mathbb{R}^d} \left| \sum_{j=1}^N \alpha_j e^{ix_j \cdot \xi} \right|^2 \widehat{\Phi}(\xi) d\xi. \tag{7.4}$$

Computation for the Power Function

Given Theorem 7.4, in this paper we are naturally interested in the explicit formula of the generalized Fourier transform for the power function $x \mapsto |x|^q$ for $q \in [1, 2)$. It is a nice example of how to pass from an ordinary Fourier transform to the generalized Fourier transform by extending the formula by means of complex analysis methods. Our starting point will be the multiquadric $x \mapsto (c^2 + |x|^2)^\beta$ for $\beta < -d/2$, whose Fourier transform involves the modified Bessel function of the third kind:

For $\nu \in \mathbb{C}, z \in \mathbb{C}$ with $|\arg z| < \pi/2$, define

$$K_\nu(z) := \int_0^\infty \exp(-z \cosh(t)) \cosh(\nu t) dt,$$

the modified Bessel function of the third kind of order $\nu \in \mathbb{C}$.

Theorem 7.5 [36, Theorem 6.13] For $c > 0$ and $\beta < -d/2$,

$$\Phi(x) = (c^2 + |x|^2)^\beta, \quad x \in \mathbb{R}^d,$$

has (classical) Fourier transform given by

$$\widehat{\Phi}(\xi) = (2\pi)^{d/2} \frac{2^{1+\beta}}{\Gamma(-\beta)} \left(\frac{|\xi|}{c} \right)^{-\beta-d/2} K_{d/2+\beta}(c|\xi|). \tag{7.5}$$

In the following result, we have slightly changed the statement compared to the original reference [36, Theorem 8.16] in order to allow orders which are a multiple of 1/2 instead of just integers. The latter situation made sense in [36] because the definition of the order involves the space \mathcal{S}_{2m} due to its purpose in the representation formula of Theorem 7.4, where a quadratic form appears. However, in Sect. 4 we need the generalized Fourier transform in the context of a linear functional, hence a different range of orders. Fortunately, one can easily generalize the proof in [36] to this fractional case, as all integrability arguments remain true when permitting multiples of 1/2, in particular the estimates in (7.8) and (7.10).

Theorem 7.6 1. [36, Theorem 8.15] $\Phi(x) = (c^2 + |x|^2)^\beta$, $x \in \mathbb{R}^d$ for $c > 0$ and $\beta \in \mathbb{R} \setminus \mathbb{N}_0$ has the generalized Fourier transform

$$\widehat{\Phi}(\xi) = (2\pi)^{d/2} \frac{2^{1+\beta}}{\Gamma(-\beta)} \left(\frac{|\xi|}{c}\right)^{-\beta-d/2} K_{d/2+\beta}(c|\xi|), \quad \xi \neq 0 \tag{7.6}$$

of order $m = \max(0, \lfloor 2\beta + 1 \rfloor / 2)$.

2. [36, Theorem 8.16] $\Phi(x) = |x|^\beta$, $x \in \mathbb{R}^d$ with $\beta \in \mathbb{R}_+ \setminus 2\mathbb{N}$ has the generalized Fourier transform

$$\widehat{\Phi}(\xi) = (2\pi)^{d/2} \frac{2^{\beta+d/2} \Gamma((d + \beta)/2)}{\Gamma(-\beta/2)} |\xi|^{-\beta-d}, \quad \xi \neq 0.$$

of order $m = \lfloor \beta + 1 \rfloor / 2$.

Note that in the cases of interest to us, the second statement of the theorem means that the generalized Fourier transform of $\Phi(x) = |x|^\beta$ is of order $\frac{1}{2}$ for $\beta \in (0, 1)$ and 1 for $\beta \in [1, 2)$, respectively. As this statement appears in a slightly modified form with respect to [36, Theorem 8.16] we report below an explicit, although rather concise proof of it.

Proof 1. We can pass from formula (7.5) to (7.6) by analytic continuation, where the exponent m serves to give us the needed integrable dominating function, see formula (7.8) below.

Let $G = \{\lambda \in \mathbb{C} : \text{Re}(\lambda) < m\} \ni \beta$ and

$$\begin{aligned} \varphi_\lambda(\xi) &:= (2\pi)^{d/2} \frac{2^{1+\lambda}}{\Gamma(-\lambda)} \left(\frac{|\xi|}{c}\right)^{-\lambda-d/2} K_{d/2+\lambda}(c|\xi|) \\ \Phi_\lambda(\xi) &:= (c^2 + |\xi|^2)^\lambda. \end{aligned}$$

We want to show that for all $\lambda \in G$

$$\int_{\mathbb{R}^d} \Phi_\lambda(\xi) \widehat{\gamma}(\xi) d\xi = \int_{\mathbb{R}^d} \varphi_\lambda(\xi) \gamma(\xi) d\xi, \quad \text{for all } \gamma \in \mathcal{S}_{2m},$$

which is so far true for λ real and $\lambda < -d/2$ by (7.5). As the integrands $\Phi_\lambda \widehat{\gamma}$ and $\varphi_\lambda \gamma$ are analytic, they can be expressed in terms of Cauchy integral formulas. The integral functions

$$\begin{aligned} f_1(\lambda) &= \int_{\mathbb{R}^d} \Phi_\lambda(\xi) \widehat{\gamma}(\xi) d\xi = \int_{\mathbb{R}^d} \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{\Phi_z(\xi)}{z - \lambda} dz \widehat{\gamma}(\xi) d\xi \\ f_2(\lambda) &= \int_{\mathbb{R}^d} \varphi_\lambda(\xi) \gamma(\xi) d\xi = \int_{\mathbb{R}^d} \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{\varphi_z(\xi)}{z - \lambda} dz \gamma(\xi) d\xi, \end{aligned}$$

will be also analytic as soon as we can find a uniform dominating function of the integrands on an arbitrary compact curve $\mathcal{C} \subset G$, to allow the application of

Fubini-Tonelli’s theorem and derive corresponding Cauchy integral formulas for f_1 and f_2 (see the details of the proofs of [36, Theorem 8.15] and [36, Theorem 8.16]). A dominating function for the integrand of $f_1(\lambda)$ is easily obtained thanks to the decay of $\widehat{\gamma} \in \mathcal{S}$ faster of any polynomially growing function (notice that $\operatorname{Re}(\lambda) < m$). It remains to find a dominating function for the integrand of $f_2(\lambda)$. Setting $b := \operatorname{Re}(\lambda)$, for ξ close to 0 we get, by using the bound

$$|K_\nu(r)| \leq \begin{cases} 2^{|\operatorname{Re}(\nu)|-1} \Gamma(|\operatorname{Re}(\nu)|) r^{-|\operatorname{Re}(\nu)|}, & \operatorname{Re}(\nu) \neq 0, \\ \frac{1}{e} - \log \frac{r}{2}, & r < 2, \operatorname{Re}(\nu) = 0. \end{cases} \tag{7.7}$$

for $\nu \in \mathbb{C}, r > 0$, as derived in [36, Lemma 5.14], that

$$|\varphi_z(\xi)\gamma(\xi)| \leq C_\gamma \frac{2^{b+|b+d/2|} \Gamma(|b+d/2|)}{|\Gamma(-\lambda)|} c^{b+d/2-|b+d/2|} |\xi|^{-b-d/2-|b+d/2|+2m} \tag{7.8}$$

for $b \neq -d/2$ and

$$|\varphi_z(\xi)\gamma(\xi)| \leq C_z \frac{2^{1-d/2}}{|\Gamma(-\lambda)|} \left(\frac{1}{e} - \log \frac{c|\xi|}{2} \right).$$

for $b = -d/2$. Taking into account that \mathcal{C} is compact and $1/\Gamma$ is an entire function, this yields

$$|\varphi_z(\xi)\gamma(\xi)| \leq C_{m,c,\mathcal{C}} \left(1 + |\xi|^{-d+2\varepsilon} - \log \frac{c|\xi|}{2} \right),$$

with $|\xi| < \min\{1/c, 1\}$ and $\varepsilon := m - b > 0$, which is locally integrable.

For ξ large, we similarly use the estimate for large r ,

$$|K_\nu(r)| \leq \sqrt{\frac{2\pi}{r}} e^{-r} e^{|\operatorname{Re}(\mu)|^2/(2r)}, \quad r > 0, \tag{7.9}$$

from [36, Lemma 5.14] to obtain

$$|\varphi_z(\xi)\gamma(\xi)| \leq C_{\mathcal{C}} \frac{2^{1+b} \sqrt{2\pi}}{|\Gamma(-\lambda)|} c^{b+(d-1)/2} |\xi|^{-b-(d+1)/2} e^{-c|\xi|} e^{|b+d/2|^2/(2c|\xi|)}$$

and consequently

$$|\varphi_\lambda(\xi)\gamma(\xi)| \leq C_{\gamma,m,\mathcal{C},c} e^{-c|\xi|},$$

which certainly is integrable.

2. We want to pass to $c \rightarrow 0$ in formula (7.6). This can be done by applying the dominated convergence theorem in the definition of the generalized Fourier transform (7.3). Writing $\Psi_c(x) := (c^2 + |x|^2)^{\beta/2}$ for $c > 0$, we know that

$$\widehat{\Psi}_c(\xi) = \psi_c(\xi) := (2\pi)^{d/2} \frac{2^{1+\beta/2}}{|\Gamma(-\beta/2)|} |\xi|^{-\beta-d} (c|\xi|)^{(\beta+d)/2} K_{(\beta+d)/2}(c|\xi|).$$

By using the decay properties of a $\gamma \in \mathcal{S}_{2m}$ in the estimate (7.8), we get

$$|\psi_c(\xi)\gamma(\xi)| \leq C_\gamma \frac{2^{\beta+d/2}\Gamma((\beta+d)/2)}{|\Gamma(-\beta/2)|} |\xi|^{2m-\beta-d} \quad \text{for } |\xi| \rightarrow 0 \quad (7.10)$$

and

$$|\psi_c(\xi)\gamma(\xi)| \leq C_\gamma \frac{2^{\beta+d/2}\Gamma((\beta+d)/2)}{|\Gamma(-\beta/2)|} |\xi|^{-\beta-d},$$

yielding the desired uniform dominating function. The claim now follows by also taking into account that

$$\lim_{r \rightarrow 0} r^\nu K_\nu(r) = \lim_{r \rightarrow 0} 2^{\nu-1} \int_0^\infty e^{-t} e^{-r^2/(4t)} t^{\nu-1} dt = 2^{\nu-1} \Gamma(\nu).$$

□

References

1. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Oxford Mathematical Monographs. The Clarendon Press, New York (2000)
2. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures. Lectures in Mathematics ETH Zürich, 2nd edn. Birkhäuser, Basel (2008)
3. Anzellotti, G., Baldo, S., Percivale, D.: Dimension reduction in variational problems, asymptotic development in Γ -convergence and thin structures in elasticity. *Asymptotic Anal.* **9**(1), 61–100 (1994)
4. Bartels, S.: Total variation minimization with finite elements: convergence and iterative solution. *SIAM J. Numer. Anal.* **50**(3), 1162–1180 (2012)
5. Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1968)
6. Billingsley, P.: Probability and Measure. Wiley Series in Probability and Mathematical Statistics, 3rd edn. Wiley, New York (1995)
7. Carrillo, J.A., Choi, Y.P., Hauray, M.: The Derivation of Swarming Models: Mean-field limit and Wasserstein Distances. *Collective Dynamics from Bacteria to Crowds: An Excursion Through Modeling, Analysis and Simulation Series*, vol. 553, pp. 1–46. CISM International Centre for Mechanical Sciences, New York (2014)
8. Carrillo, J.A., Toscani, G.: Contractive probability metrics and asymptotic behavior of dissipative kinetic equations. *Riv. Mat. Univ. Parma* **7**(6), 75–198 (2007)
9. Chambolle, A., Caselles, V., Cremers, D., Novaga, M., Pock, T.: Theoretical Foundations and Numerical Methods for Sparse Recovery. *Radon Series on Computational and Applied Mathematics. An introduction to total variation for image analysis*, vol. 9, pp. 263–340. Walter de Gruyter, Berlin (2010)
10. Cicalese, M., Spadaro, E.: Droplet minimizers of an isoperimetric problem with long-range interactions. *Commun. Pure Appl. Math.* **66**(8), 1298–1333 (2013)
11. CVX Research, Inc. CVX: Matlab software for disciplined convex programming, version 2.0 beta. <http://cvxr.com/cvx> (2013)
12. Dal Maso, G.: An Introduction to Γ -Convergence. Progress in Nonlinear Differential Equations and their Applications, vol. 8. Birkhäuser, Boston (1993)
13. Dereich, S., Scheutzow, M., Schottstedt, R.: Constructive quantization: approximation by empirical measures. *Ann. Inst. Henri Poincaré (B)* **49**(4), 1183–1203 (2013)
14. Devroye, L., Györfi, L.: Nonparametric Density Estimation. Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics. Wiley, New York (1985)
15. Di Francesco, M., Fornasier, M., Hütter, J.-C., Matthes, D.: Asymptotic behavior of gradient flows driven by power repulsion and attraction potentials in one dimension. *SIAM J. Math. Anal.* **26**(6), 3814–3837 (2014)
16. Durrett, R.: Probability: Theory and Examples. Cambridge Series in Statistical and Probabilistic Mathematics, 4th edn. Cambridge University Press, Cambridge (2010)

17. Evans, L.C., Gariepy, R.F.: Measure Theory and Fine Properties of Functions. Studies in Advanced Mathematics. CRC Press, Boca Raton (1992)
18. Fattal, R.: Blue-noise point sampling using kernel density model. *ACM SIGGRAPH 2011 papers*, 28(3), 1–10 (2011)
19. Fenn, M., Steidl, G.: Fast NFFT based summation of radial functions. *Sampl. Theor. Sig. Image Process* **3**(1), 1–28 (2004)
20. Fornasier, M., Haškovec, J., Steidl, G.: Consistency of variational continuous-domain quantization via kinetic theory. *Appl. Anal.* **92**(6), 1283–1298 (2013)
21. Goldman, D., Muratov, C.B., Serfaty, S.: The Γ -limit of the two-dimensional Ohta-Kawasaki energy. I. Droplet density. *Arch. Ration. Mech. Anal.* **210**(2), 581–613 (2013)
22. Goldman, D., Muratov, C.B., Serfaty, S.: The Γ -limit of the two-dimensional Ohta-Kawasaki energy. Droplet arrangement via the renormalized energy. *Arch. Ration. Mech. Anal.* **212**(2), 445–501 (2014)
23. Gräf, M., Potts, D., Steidl, G.: Quadrature errors, discrepancies, and their relations to halftoning on the torus and the sphere. *SIAM J. Sci. Comput.* **34**(5), A2760–A2791 (2012)
24. Graf, S., Luschgy, H.: Foundations of Quantization for Probability Distributions. Lecture Notes in Mathematics, vol. 1730. Springer, Berlin (2000)
25. Grant, M., Boyd, S.: Graph implementations for nonsmooth convex programs. In: Blondel, V., Boyd, S., Kimura, H. (eds.) Recent Advances in Learning and Control. Lecture Notes in Control and Information Sciences, pp. 95–110. Springer (2008). http://stanford.edu/~boyd/graph_dcp.html
26. Gruber, P.M.: Optimum quantization and its applications. *Adv. Math.* **186**(2), 456–497 (2004)
27. Hütter, J.-C.: Minimizers and Gradient Flows of Attraction-Repulsion Functionals with Power Kernels and Their Total Variation Regularization. Master Thesis, Technical University of Munich (2013)
28. Knüpfer, H., Muratov, C.B.: On an isoperimetric problem with a competing nonlocal term. I: The planar case. *Commun. Pure Appl. Math.* **66**(7), 1129–1162 (2013)
29. Knüpfer, H., Muratov, C.B.: On an isoperimetric problem with a competing nonlocal term II: The general case. *Commun. Pure Appl. Math.* **67**(12), 1974–1994 (2014)
30. Pagès, G.: A space quantization method for numerical integration. *J. Comput. Appl. Math.* **89**(1), 1–38 (1998)
31. Potts, D., Steidl, G.: Fast summation at nonequispaced knots by NFFT. *SIAM J. Sci. Comput.* **24**(6), 2013–2037 (2003)
32. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Phys. D* **60**(1–4), 259–268 (1992)
33. Schmaltz, C., Gwosdek, P., Bruhn, A., Weickert, J.: Electrostatic halftoning. *Comput. Graph.* **29**, 2313–2327 (2010)
34. Shorack, G.R., Wellner, J.A.: Empirical processes with applications to statistics. Reprint of the 1986 original ed. Reprint of the 1986 original ed. edition (2009)
35. Teuber, T., Steidl, G., Gwosdek, P., Schmaltz, C., Weickert, J.: Dithering by differences of convex functions. *SIAM J. Imaging Sci.* **4**(1), 79–108 (2011)
36. Wendland, H.: cattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics, vol. 17. Cambridge University Press, Cambridge (2005)
37. Wied, D., Weißbach, R.: Consistency of the kernel density estimator: a survey. *Statist. Papers* **53**(1), 1–21 (2012)