Check for
updates

Structural Changes in Nonlocal Denoising Models Arising Through Bi-Level Parameter Learning

Elisa Davoli[1] · **Rita Ferreira[2]** · **Carolin Kreisbeck[3]** ·
Hidde Schönberger[3]

Abstract

We introduce a unified framework based on bi-level optimization schemes to deal with parameter learning in the context of image processing. The goal is to identify the optimal regularizer within a family depending on a parameter in a general topological space. Our focus lies on the situation with non-compact parameter domains, which is, for example, relevant when the commonly used box constraints are disposed of. To overcome this lack of compactness, we propose a natural extension of the upper-level functional to the closure of the parameter domain via Gamma-convergence, which captures possible structural changes in the reconstruction model at the edge of the domain. Under two main assumptions, namely, Mosco-convergence of the regularizers and uniqueness of minimizers of the lower-level problem, we prove that the extension coincides with the relaxation, thus admitting minimizers that relate to the parameter optimization problem of interest. We apply our abstract framework to investigate a quartet of practically relevant models in image denoising, all featuring nonlocality. The associated families of regularizers exhibit qualitatively different parameter dependence, describing a weight factor, an amount of nonlocality, an integrability exponent,

✉ Carolin Kreisbeck
   carolin.kreisbeck@ku.de

   Elisa Davoli
   elisa.davoli@tuwien.ac.at

   Rita Ferreira
   rita.ferreira@kaust.edu.sa

   Hidde Schönberger
   hidde.schoenberger@ku.de

[1] Institute of Analysis and Scientific Computing, TU Wien, Wiedner Hauptstrasse 8-10, 1040 Vienna, Austria

[2] CEMSE Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

[3] Mathematisch-Geographische Fakultät, KU Eichstätt-Ingolstadt, Ostenstrasse 26, 85072 Eichstätt, Germany

 Springer

and a fractional order, respectively. After the asymptotic analysis that determines the relaxation in each of the four settings, we finally establish theoretical conditions on the data that guarantee structural stability of the models and give examples of when stability is lost.

Keywords Bi-level learning scheme · Parameter optimization · Γ-convergence · Nonlocal regularizers · Image denoising models

Mathematics Subject Classification 49J21 · 49J45

1 Introduction

One of the most widely used methods to solve image restoration problems is the variational regularization approach. This variational approach consists of minimizing a reconstruction functional that decomposes into a fidelity and a regularization term, which give rise to competing effects. While the fidelity term ensures that the reconstructed image is close to the (noisy) data, the regularization term is designed to remove the noise by incorporating prior information on the clean image. In the case of a simple L^2-fidelity term, the reconstruction functional is given by

$$\mathcal{J}(u) = \|u - u^\eta\|^2_{L^2(\Omega)} + \mathcal{R}(u), \qquad \text{for } u \in L^2(\Omega),$$

where $\Omega \subset \mathbb{R}^n$ is the image domain, $u^\eta \in L^2(\Omega)$ the noisy image, and $\mathcal{R} : L^2(\Omega) \to [0, \infty]$ the regularizer.

A common choice for $\mathcal{R}$ is the total variation (TV) regularization proposed by Rudin, Osher, & Fatemi [52], which penalizes sharp oscillations, but does not exclude edge discontinuities, as they appear in most images. Since its introduction, the TV-model has inspired a variety of more advanced regularization terms, like the infimal-convolution total variation $(ICTV)$ [19], the total generalized variation (TGV) [14], and many more, cf. [10] and the references therein. Due to the versatility of the variational formulation, regularizers of a completely different type can be used as well. Recently, a lot of attention has been directed towards regularizers incorporating nonlocal effects, such as those induced by difference quotients [5, 11, 15, 38] and fractional operators [1, 3, 4]. Nonlocal regularizers have the advantage of not requiring the existence of (full) derivatives, allowing to work with functions that are less regular than those in the local counterpart.

With an abundance of available choices, finding a suitable regularization term for a specific application is paramount for obtaining accurate reconstructions. This is often done by fixing a parameter-dependent family of regularizers and tuning the parameter in accordance with the noise and data. Carrying out this process via trial and error can be hard and inefficient, which led to the development of a more structured approach in the form of bi-level optimization. We refer, e.g., to [30, 31] (see also [21, 22, 35, 53]) and to the references therein, as well as to [33] for a detailed overview. The idea behind bi-level optimization is to employ a supervised learning scheme based on a set of training data consisting of noisy images and their corresponding clean

versions. To determine an optimal parameter, one minimizes a selected cost functional which quantifies the error with respect to the training data. Overall, this results in a nested variational problem with upper- and lower-level optimization steps related to the cost and reconstruction functional, respectively. Key aspects of the mathematical study of these bi-level learning schemes include establishing the existence of solutions and deriving optimality conditions, which lay the foundation for devising reliable numerical solution methods.

In recent years, there has been a rapid growth in the literature devoted to addressing the above questions. To mention but a few examples, we first refer the paper [41] dealing with learning real-valued weight parameters in front of the regularization terms for a rather general class of inverse problems; in [2, 6], the authors optimize the fractional parameter of a regularizer depending on the spectral fractional Laplacian; spatially dependent weights are determined through training via other nonlocal bi-level schemes (e.g., inside the Gagliardo semi-norm [40] or in a type of fractional gradient [32]), and in classical TV-models [23, 39, 47]; as done in [29], one can also learn the fidelity term instead of the regularizer.

A common denominator in the above references is the presence of certain a priori compactness constraints on the set of admissible parameters, such as box constraints like in [41], where the weights are assumed to lie in some compact interval away from 0 and infinity. These conditions make it possible to prove stability of the lower-level problem and obtain existence of optimal parameters within a class of structurally equivalent regularizers. However, imposing artificial restrictions to the parameter range like these may lead to suboptimal results depending on the given training data.

It is then substantial to consider removing such constraints in order to work on maximal domains naturally associated with the parameters, which is also our focus in this paper. An inherent effect of this approach is that qualitative changes in the structure of the regularizer may occur at the edges of the domain. If optimal parameters are attained at the boundary, this indicates that the chosen class of regularization terms is not well-suited to the training data. To exclude these degenerate cases, it is of interest to provide analytic conditions to guarantee that the optimal parameters are attained in the interior of the domain, thereby preserving the structure of the regularizer. The first work to address the aforementioned tasks is [30] by De Los Reyes, Schönlieb, & Valkonen, where optimization is carried out for weighted sums of local regularizers of different type with each weight factor allowed to take any value in $[0, \infty]$. As such, their bi-level scheme is able to encompass multiple regularization structures at once, like TV and TV^2 and their interpolation TGV. Similarly, the authors in [44] vary the weight factor in the whole range $[0, \infty]$ as well as the underlying finite-dimensional norm of the total variation regularizer. We also mention [28], where the order of a newly introduced nonlocal counterpart of the TGV-regularizer is tuned, and [27], which studies a bi-level scheme covering the cases of TV, TGV^2, and $NsTGV^2$ in a comprehensive way.

In this paper, we introduce a unified framework to deal with parameter learning beyond structural stability in the context of bi-level optimization schemes. In contrast to the above references, where the analysis is tailored to a specifically chosen type of parameter dependence, our regularizers can exhibit a general dependence on parameters in a topological space. Precisely, we consider a parametrized family of

regularizers $\mathcal{R}_\lambda : L^2(\Omega) \to [0, \infty]$ with λ ranging over a subset Λ of a topological space X, which is assumed to be first countable. If we focus for brevity on a single data point $(u^c, u^\eta) \in L^2(\Omega) \times L^2(\Omega)$, with u^c and u^η the clean and noisy images (see Sect. 2 for larger data sets), the bi-level optimization problem reads:

(Upper−level)   Minimize $\mathcal{I}(\lambda) := \inf_{w \in K_\lambda} \|w - u^c\|^2_{L^2(\Omega)}$ over $\lambda \in \Lambda$,

(Lower−level) $K_\lambda := \underset{u \in L^2(\Omega)}{\arg\min} \, \mathcal{J}_\lambda(u)$,

where $\mathcal{J}_\lambda(u) := \|u - u^\eta\|^2_{L^2(\Omega)} + \mathcal{R}_\lambda(u)$ is the reconstruction functional.

Our approach for studying this general bi-level learning scheme relies on asymptotic tools from the calculus of variations. We define a suitable notion of stability for the lower-level problems that requires the family of functionals $\{\mathcal{J}_\lambda\}_{\lambda \in \Lambda}$ to be closed under taking Γ-limits; see [13, 25] for a comprehensive introduction on Γ-convergence. Since Γ-convergence ensures the convergence of sequences of minimizers, one can conclude that, in the presence of stability, the upper-level functional $\mathcal{I}$ admits a minimizer (Theorem 2.3).

A different strategy is required to obtain the existence of solutions when stability fails. Especially relevant here is the case of real-valued parameters when box constraints are disposed of and non-closed intervals Λ are considered; clearly, stability is then lost for the simple fact that a sequence of parameters can converge to the boundary of Λ. To overcome this issue, we propose a natural extension $\overline{\mathcal{I}} : \overline{\Lambda} \to [0, \infty]$ of $\mathcal{I}$, now defined on the closure of our parameter domain, and identified via Γ-convergence of the lower-level functionals. Precisely,

(Upper−level)   Minimize $\overline{\mathcal{I}}(\lambda) := \inf_{w \in \overline{K}_\lambda} \|w - u^c\|^2_{L^2(\Omega)}$ over $\lambda \in \overline{\Lambda}$,

(Lower−level) $\overline{K}_\lambda := \underset{u \in L^2(\Omega)}{\arg\min} \, \overline{\mathcal{J}}_\lambda(u)$,

where the functionals $\overline{\mathcal{J}}_\lambda : L^2(\Omega) \to [0, \infty]$ are characterized as L^2-weak Γ-limits (if they exist) of functionals $\mathcal{J}_{\lambda'}$ with $\lambda' \to \lambda$. To justify the choice of this particular extension, we derive an intrinsic connection with relaxation theory in the calculus of variations (for an introduction, see, e.g., [24, Chapter 9] and the references therein). Explicitly, the relaxation of the upper-level functional $\mathcal{I}$ is given by its lower semi-continuous envelope (after the trivial extension to $\overline{\Lambda}$ by ∞),

$$\mathcal{I}^{\mathrm{rlx}}(\lambda) := \inf \left\{ \liminf_{k \to \infty} \mathcal{I}(\lambda_k) : (\lambda_k)_k \subset \Lambda, \lambda_k \to \lambda \text{ in } \overline{\Lambda} \right\} \text{ for } \lambda \in \overline{\Lambda}.$$

This relaxed version of $\mathcal{I}$ has the desirable property that it admits a minimizer (if $\overline{\Lambda}$ is compact) and minimizing sequences of $\mathcal{I}$ have subsequences that converge to an optimal parameter of $\mathcal{I}^{\mathrm{rlx}}$. Our main theoretical result (Theorem 2.5) shows that the extension $\overline{\mathcal{I}}$ coincides with the relaxation $\mathcal{I}^{\mathrm{rlx}}$ under suitable assumptions and therefore inherits the same properties (cf. Corollary 2.8).

Besides the generic conditions that each $\mathcal{R}_\lambda$ is weakly lower semicontinuous and has non-empty domain (see (H)), which ensure that $\mathcal{J}_\lambda$ possesses a minimizer, we work under two main assumptions:

(i) The Mosco-convergence of the regularizers, i.e., Γ-convergence with respect to the strong and weak L^2-topology, and
(ii) the uniqueness of minimizers of $\overline{\mathcal{J}}_\lambda$ for $\lambda \in \overline{\Lambda}\setminus\Lambda$.

We demonstrate in Example 2.7 that these assumptions are in fact optimal. Due to (i), the Γ-limits $\overline{\mathcal{J}}_\lambda$ preserve the additive decomposition into the L^2-fidelity term and a regularizer, and coincide with $\mathcal{J}_\lambda$ inside Λ. As a consequence of the latter, it follows that $\overline{\mathcal{I}} = \mathcal{I}$ in Λ, making $\overline{\mathcal{I}}$ a true extension of $\mathcal{I}$. For the parameter values at the boundary, $\lambda \in \overline{\Lambda}\setminus\Lambda$, however, the regularizers present in $\overline{\mathcal{J}}_\lambda$ can have a completely different structure from the family of regularizers $\{\mathcal{R}_\lambda\}_{\lambda\in\Lambda}$ that we initially started with. When the optimal parameter of the extended problem is attained inside Λ, one recovers instead a solution to the original training scheme, yielding structure preservation. For a discussion on related results in the context of optimal control problems [9, 16, 17], we refer to the end of Sect. 2.

To demonstrate the applicability of our abstract framework, we investigate a quartet of practically relevant scenarios with families of nonlocal regularizers that induce qualitatively different structural changes; namely, learning the optimal weight, varying the amount of nonlocality, optimizing the integrability exponent, and tuning the fractional parameter. More precisely, in all these four applications, our starting point is a non-closed real interval $\Lambda \subset [-\infty, \infty]$ and we seek to determine the extension $\overline{\mathcal{I}}$ on the closed interval $\overline{\Lambda}$, which admits a minimizer by the theory outlined above. The first step is to calculate the Mosco-limits of the regularizers, which reveals the type of structural change occurring at the boundary points. Subsequently, we study for which training sets of clean and noisy images the optimal parameters are attained either inside Λ or at the edges. In two cases, we determine explicit analytic conditions on the data that guarantee structure preservation for the optimization process.

The first setting involves a rather general nonlocal regularizer $\mathcal{R} : L^2(\Omega) \to [0, \infty]$ multiplied by a weight parameter α in $\Lambda = (0, \infty)$. Inside the domain, we observe structural stability as $\overline{\mathcal{J}}_\alpha = \mathcal{J}_\alpha$ for all $\alpha \in \Lambda$; in contrast, the regularization disappears when $\alpha = 0$ and forces the solutions to be constant when $\alpha = \infty$. Moreover, we derive sufficient conditions in terms of the data that prevent the optimal parameter from being attained at the boundary points; for a single data point (u^c, u^η), they specify to

$$\mathcal{R}(u^c) < \mathcal{R}(u^\eta) \quad \text{and} \quad \|u^\eta - u^c\|^2_{L^2(\Omega)} < \left\| \fint_\Omega u^\eta \, \mathrm{d}x - u^c \right\|^2_{L^2(\Omega)},$$

see Theorem 3.2. Notice that the first of these two conditions is comparable to the one in [30, Eq. (10)] and shows positivity of optimal weights.

Inspired by the use of different L^p-norms in image processing, such as in the form of quadratic, TV, and Lipschitz regularization [50, Sect. 4], we focus our second case on the integrability exponent of nonlocal regularizers of double-integral type;

precisely, functionals of the form

$$\mathcal{R}_p(u) = \left(\frac{1}{|\Omega \times \Omega|} \int_\Omega \int_\Omega f^p(x, y, u(x), u(y)) \, dx \, dy \right)^{1/p} \quad \text{for } p \in \Lambda = [1, \infty),$$

with a suitable $f : \Omega \times \Omega \times \mathbb{R} \times \mathbb{R} \to [0, \infty)$. Possible choices for the integrand f include bounded functions or functions of difference-quotient type. We prove stability of the lower-level problem in Λ, and determine the Mosco-limit for $p \to \infty$ via L^p-approximation techniques as in [20, 42]. In particular, we show that it is given by a double-supremal functional of the form

$$\mathcal{R}_\infty(u) = \text{ess sup}_{(x,y) \in \Omega \times \Omega} f(x, y, u(x), u(y)).$$

In order to see how this structural change affects the image reconstruction, we highlight examples of training data for which the supremal regularizer performs better or worse than the integral counterparts.

As a third application, we consider two families of nonlocal regularizers $\{\mathcal{R}_\delta\}_{\delta \in \Lambda}$ with $\Lambda = (0, \infty)$, which were introduced by Aubert & Kornprobst [5] and Brezis & Nguyen in [15], respectively, and are closely related to nonlocal filters frequently used in image processing. The parameter δ reflects the amount of nonlocality in the regularizer. It is known that the functionals $\mathcal{R}_\delta$ tend, as $\delta \to 0$, to a multiple of the total variation in the sense of Γ-convergence. Based on these results, we prove in both cases that the reconstruction functional of our bi-level optimization scheme turns into the classical TV-denoising model when $\delta = 0$, whereas the regularization vanishes at the other boundary value, $\delta = \infty$. As such, the extended bi-level schemes encode simultaneously nonlocal and total variation regularizations. We round off the discussion by presenting some instances of training data where the optimal parameters are attained either at the boundary or in the interior of Λ.

Our final bi-level optimization problem features a different type of nonlocality arising from fractional operators; to be precise, we consider, in the same spirit as in [1], the L^2-norm of the spectral fractional Laplacian as a regularizer. The parameter of interest here is the order $s/2$ of the fractional Laplacian, which is taken in the fractional range $s \in \Lambda = (0, 1)$. At the values $s = 0$ and $s = 1$, we recover local models with regularizers equal to the L^2-norm of the function and its gradient, respectively. Thus, one expects the fractional model to perform better than the two local extremes. We quantify this presumption by deriving analytic conditions in terms of the eigenfunctions and eigenvalues of the classical Laplacian on Ω ensuring the optimal parameters to be attained in the truly fractional regime. These conditions on the training data are established by proving and exploiting the differentiability of the extended upper-level functional $\overline{\mathcal{I}}$.

For completeness, we mention that practically relevant scenarios when Λ is a topological space include those in which the reconstruction parameters are space-dependent, and thus described by functions. The analysis of this class of applications is left open for future investigations.

The outline of the paper is as follows. In Sect. 2, we present the general abstract bi-level framework, and prove the results regarding the existence of optimal parameters

and the two types of extensions of bi-level optimization schemes. Sections 3–6 then deal with the four different, practically relevant applications mentioned in the previous paragraph. As a note, we point out that they are each presented in a self-contained way, allowing the readers to move directly to the sections that correspond best to their interests.

2 Establishing the Unified Framework

Let $\Omega \subset \mathbb{R}^n$ be an open bounded set, and let

$$\bigcup_{j=1}^{N} (u_j^c, u_j^\eta) \subset L^2(\Omega) \times L^2(\Omega), \quad N \in \mathbb{N},$$

be a set of available square-integrable training data, where each u_j^c represents a clean image and u_j^η a distorted version thereof, which can be obtained, for instance, by applying some noise to u_j^c. These data are collected in the vector-valued functions $u^c := (u_1^c, \ldots, u_N^c) \in L^2(\Omega; \mathbb{R}^N)$ and $u^\eta := (u_1^\eta, \ldots, u_N^\eta) \in L^2(\Omega; \mathbb{R}^N)$. As for notation, $\|v\|_{L^2(\Omega;\mathbb{R}^N)}^2 = \sum_{j=1}^{N} \|v_j\|_{L^2(\Omega)}^2$ stands for the L^2-norm of a function $v \in L^2(\Omega; \mathbb{R}^N)$.

To reconstruct each damaged image, u_j^η, we consider denoising models that consist of a simple fidelity term and a (possibly nonlocal) regularizer; precisely, we minimize functionals $\mathcal{J}_{\lambda,j} : L^2(\Omega) \to [0, \infty]$ of the form

$$\mathcal{J}_{\lambda,j}(u) = \|u - u_j^\eta\|_{L^2(\Omega)}^2 + \mathcal{R}_\lambda(u), \qquad u \in L^2(\Omega), \tag{2.1}$$

where the regularizer $\mathcal{R}_\lambda : L^2(\Omega) \to [0, \infty]$, with $\mathrm{Dom}\, \mathcal{R}_\lambda = \{v \in L^2(\Omega) : \mathcal{R}_\lambda(u) < \infty\}$, is a (possibly nonlocal) functional parametrized over $\lambda \in \Lambda$ with Λ a subset of a topological space X satisfying the first axiom of countability. Throughout the paper, we always assume that for every $\lambda \in \Lambda$, we have

$$\begin{cases} \mathrm{Dom}\, \mathcal{R}_\lambda \text{ is non-empty,} \\ \mathcal{R}_\lambda \text{ is weakly } L^2\text{-lower semicontinuous.} \end{cases} \tag{H}$$

Observe that the functionals $\mathcal{J}_{\lambda,j}$ then have a minimizer by the direct method in the calculus of variations.

The result of the reconstruction process, meaning the quality of the reconstructed image resulting as a minimizer of (2.1), is known to depend on the choice of the regularizing term $\mathcal{R}_\lambda$. Our goal is to set up a training scheme that is able to learn how to select a "good" parameter λ within a corresponding given family $\{\mathcal{R}_\lambda\}_{\lambda \in \Lambda}$ of regularizers. Here, as briefly described in the Introduction for the single data point case ($N = 1$), we follow the approach introduced in [30, 31] in the spirit of machine

learning optimization schemes, where training the regularization term means to solve the nested variational problem

(Upper−level)   Minimize $\mathcal{I}(\lambda) := \inf\limits_{w \in K_\lambda} \|w - u^c\|^2_{L^2(\Omega;\mathbb{R}^N)}$ over $\lambda \in \Lambda$,

(Lower−level) $K_\lambda := \left\{ w \in L^2(\Omega; \mathbb{R}^N) : w_j \in \arg\min\limits_{u \in L^2(\Omega)} \mathcal{J}_{\lambda,j}(u) \text{ for all } j \in \{1, \ldots, N\} \right\}$,

$$(\mathcal{T})$$

with $\mathcal{J}_{\lambda,j}$ as in (2.1). Notice that $K_\lambda \neq \emptyset$ because for all $j \in \{1, \ldots, N\}$, we have

$$K_{\lambda,j} := \mathrm{argmin}_{u \in L^2(\Omega)} \mathcal{J}_{\lambda,j}(u) \neq \emptyset \tag{2.2}$$

by Assumption (H).

   To study the training scheme ($\mathcal{T}$), we start by introducing a notion of weak L^2-stability for the family $\{\mathcal{J}_\lambda\}_{\lambda \in \Lambda}$, with

$$\mathcal{J}_\lambda := (\mathcal{J}_{\lambda,1}, \ldots, \mathcal{J}_{\lambda,N}) : L^2(\Omega) \to [0, \infty]^N \text{ for } \lambda \in \Lambda. \tag{2.3}$$

This notion relies on the concept of Γ-convergence and is related to the notion of (weak) stability as in [41, Definition 2.3], which is defined in terms of minimizers of the lower-level problem.

Definition 2.1 (*Weak L^2-stability*) The family in (2.3) is called weakly L^2-stable if for every sequence $(\lambda_k)_k \subset \Lambda$ such that $(\mathcal{J}_{\lambda_k,j})_k$ Γ-converges with respect to the weak L^2-topology for all $j \in \{1, \ldots, N\}$, there exists $\lambda \in \Lambda$ such that

$$\Gamma(w\text{-}L^2)\text{-}\lim_{k \to \infty} \mathcal{J}_{\lambda_k,j} = \mathcal{J}_{\lambda,j}$$

for all $j \in \{1, \ldots, N\}$.

   Before proceeding, we briefly recall the definition and some properties of Γ-convergence in the setting relevant to us; for more on this topic, see [13, 25] for instance.

Definition 2.2 (Γ- *and Mosco-convergence*) Let $\mathcal{F}_k : L^2(\Omega) \to [0, \infty]$ for $k \in \mathbb{N}$ and $\mathcal{F} : L^2(\Omega) \to [0, \infty]$ be functionals. The sequence $(\mathcal{F}_k)_k$ (sequentially) Γ-converges to $\mathcal{F}$ with respect to the weak L^2-topology, written $\mathcal{F} = \Gamma(w\text{-}L^2)\text{-}\lim_{k \to \infty} \mathcal{F}_k$, if:

- (Liminf inequality) For every sequence $(u_k)_k \subset L^2(\Omega)$ and $u \in L^2(\Omega)$ with $u_k \rightharpoonup u$ in $L^2(\Omega)$, it holds that

$$\mathcal{F}(u) \leq \liminf_{k \to \infty} \mathcal{F}_k(u_k).$$

- (Limsup inequality) For every $u \in L^2(\Omega)$, there exists a sequence $(u_k)_k \subset L^2(\Omega)$ such that $u_k \rightharpoonup u$ in $L^2(\Omega)$ and

$$\mathcal{F}(u) \geq \limsup_{k \to \infty} \mathcal{F}_k(u_k).$$

The sequence $(\mathcal{F}_k)_k$ converges in the sense of Mosco-convergence in $L^2(\Omega)$ to $\mathcal{F}$, written $\mathcal{F} = \text{Mosc}(L^2)\text{-}\lim_{k\to\infty}\mathcal{F}_k$, if, in addition, the limsup inequality can be realised by a sequence converging strongly in $L^2(\Omega)$.

If the liminf inequality holds, then the sequence from the limsup inequality automatically satisfies $\lim_{k\to\infty}\mathcal{F}_k(u_k) = \mathcal{F}(u)$, and is therefore often called a recovery sequence. We note that the above sequential definition of Γ-convergence coincides with the topological definition [25, Proposition 8.10] for equi-coercive sequences $(\mathcal{F}_k)_k$, i.e., $\mathcal{F}_k \geq \Psi$ for all $k \in \mathbb{N}$ and for some $\Psi : L^2(\Omega) \to [0, \infty]$ with $\Psi(u) \to \infty$ as $\|u\|_{L^2(\Omega)} \to \infty$. In particular, the theory implies that the Γ-limit $\mathcal{F}$ is (sequentially) L^2-weakly lower semicontinuous. The Γ-convergence has the key property of yielding the convergence of solutions (if they exist) to those of the limit problem, which makes it a suitable notion of variational convergence. Precisely, if u_k is a minimizer of $\mathcal{F}_k$ for all $k \in \mathbb{N}$ and u a cluster point of the sequence $(u_k)_k$, then u is a minimizer of $\mathcal{F}$ and $\min_{L^2(\Omega)}\mathcal{F}_k = \mathcal{F}_k(u_k) \to \mathcal{F}(u) = \min_{L^2(\Omega)}\mathcal{F}$, see [25, Corollary 7.20]. Notice that the existence of cluster points is implied by the assumption of equi-coercivity. In the special case when $(\mathcal{F}_k)_k$ is a constant sequence of functionals, say $\mathcal{F}_k = \mathcal{G}$ for all $k \in \mathbb{N}$, the Γ-limit corresponds to the relaxation of $\mathcal{G}$, i.e., its L^2-weakly lower semicontinuous envelope. Observe that replacing each $\mathcal{F}_k$ by its relaxation does not affect the Γ-limit of $(\mathcal{F}_k)_k$, see [25, Proposition 6.11].

As we discuss next, weak L^2-stability provides existence of solutions to the training scheme ($\mathcal{T}$). We note that the family of functionals $\{\mathcal{J}_\lambda\}_{\lambda \in \Lambda}$ as in (2.3) is equi-coercive in a componentwise sense.

Theorem 2.3 *Let $\mathcal{J}_\lambda : L^2(\Omega) \to [0, \infty]^N$ be given by (2.3) for each $\lambda \in \Lambda$. If the family $\{\mathcal{J}_\lambda\}_{\lambda \in \Lambda}$ is weakly L^2-stable, then $\mathcal{I}$ in ($\mathcal{T}$) has a minimizer.*

Proof The statement follows directly from the direct method and the classical properties of Γ-convergence.

Let $(\lambda_k)_k \subset \Lambda$ be a minimizing sequence for $\mathcal{I}$. Then, for each $k \in \mathbb{N}$, there is $w_k \in K_{\lambda_k}$ such that

$$\lim_{k\to\infty} \|w_k - u^c\|^2_{L^2(\Omega;\mathbb{R}^N)} = \inf_{\lambda \in \Lambda} \mathcal{I}(\lambda). \tag{2.4}$$

In particular, $(w_k)_k$ is uniformly bounded in $L^2(\Omega; \mathbb{R}^N)$; hence, extracting a subsequence if necessary, one may assume that $w_k \rightharpoonup w$ in $L^2(\Omega; \mathbb{R}^N)$ as $k \to \infty$ for some $w \in L^2(\Omega; \mathbb{R}^N)$. Using the equi-coercivity, we apply the compactness result for Γ-limits [25, Corollary 8.12] to find a further subsequence of $(\lambda_k)_k$ (not relabeled) such that $(\mathcal{J}_{\lambda_k,j})_k$ $\Gamma(w\text{-}L^2)$-converges for all $j \in \{1, ..., N\}$. Consequently, by the weak L^2-stability assumption and the properties of Γ-convergence on minimizing sequences, there exists $\tilde{\lambda} \in \Lambda$ such that $w \in K_{\tilde{\lambda}}$. Then, along with (2.4),

$$\mathcal{I}(\tilde{\lambda}) \leq \|w - u^c\|^2_{L^2(\Omega;\mathbb{R}^N)} \leq \liminf_{k\to\infty} \|w_k - u^c\|^2_{L^2(\Omega;\mathbb{R}^N)} = \inf_{\lambda \in \Lambda} \mathcal{I}(\lambda) \leq \mathcal{I}(\tilde{\lambda}),$$

which finishes the proof. $\qquad\square$

Remark 2.4 We give a simple counterexample to illustrate that minimizers for $\mathcal{I}$ may not exist in general. Take $\Lambda = (0, \infty) \subset \mathbb{R}$, a single data point (u^c, u^η) with $u^c = u^\eta \neq 0$, and $\mathcal{R}_\lambda(u) = \lambda \|u\|_{L^2(\Omega)}^2$ for $\lambda \in \Lambda$. Then, $\mathcal{J}_\lambda(u) = \|u - u_\eta\|_{L^2(\Omega)}^2 + \lambda \|u\|_{L^2(\Omega)}^2$ for $u \in L^2(\Omega)$ and $K_\lambda = \{u^\eta/(1+\lambda)\} = \{u^c/(1+\lambda)\}$, so that

$$\mathcal{I}(\lambda) = \left(\frac{\lambda}{1+\lambda}\right)^2 \|u^c\|_{L^2(\Omega)}^2,$$

which does not have a minimizer on $\Lambda = (0, \infty)$. By the previous theorem, the family must fail to be weakly L^2-stable. Indeed, $\Gamma(w\text{-}L^2)\text{-}\lim_{\lambda \to 0} \mathcal{J}_\lambda$ coincides with the pointwise limit and is equal to $\| \cdot - u_\eta\|_{L^2(\Omega)}^2$, which is not an element of $\{\mathcal{J}_\lambda\}_{\lambda \in (0, \infty)}$.

Theorem 2.3 is useful in many situations, including the basic case when the parameter set Λ is a compact real interval. However, weak L^2-stability is not always guaranteed, as Remark 2.4 illustrates. If, for instance, we have a sequence $(\lambda_k)_k$ converging to a point in X outside Λ, then there is no reason to expect that

$$\Gamma(w\text{-}L^2)\text{-}\lim_{k \to \infty} \mathcal{J}_{\lambda_k, j} = \mathcal{J}_{\lambda, j}$$

holds for some $\lambda \in \Lambda$.

To overcome this issue and provide a more general existence framework, we will look at a suitable replacement of the bi-level scheme. In the following, we denote by $\overline{\Lambda}$ the closure of Λ and suppose that for each $j \in \{1, \ldots, N\}$ and $\lambda \in \overline{\Lambda}$, the Γ-limits

$$\overline{\mathcal{J}}_{\lambda, j} := \Gamma(w\text{-}L^2)\text{-}\lim_{\lambda' \to \lambda} \mathcal{J}_{\lambda', j} \tag{2.5}$$

exist, where λ' takes values on an arbitrary sequence in Λ. We further set

$$\overline{\mathcal{J}}_\lambda := (\overline{\mathcal{J}}_{\lambda, 1}, \ldots, \overline{\mathcal{J}}_{\lambda, N}) : \overline{\Lambda} \to [0, \infty]^N.$$

Based on these definitions, we introduce $\overline{\mathcal{I}} : \overline{\Lambda} \to [0, \infty]$ as the extension of the upper level functional $\mathcal{I}$ given by

$$\overline{\mathcal{I}}(\lambda) := \min_{w \in \overline{K}_\lambda} \|w - u^c\|_{L^2(\Omega; \mathbb{R}^N)}^2, \tag{2.6}$$

where $\overline{K}_{\lambda, j} := \operatorname{argmin}_{u \in L^2(\Omega)} \overline{\mathcal{J}}_{\lambda, j}(u)$ and $\overline{K}_\lambda := \overline{K}_{\lambda, 1} \times \overline{K}_{\lambda, 2} \times \cdots \times \overline{K}_{\lambda, N}$ for $\lambda \in \overline{\Lambda}$. Observe that $\overline{K}_{\lambda, j}$ is L^2-weakly closed because the functional $\overline{\mathcal{J}}_{\lambda, j}$, as a $\Gamma(w\text{-}L^2)$-limit by (2.5), is L^2-weakly lower semicontinuous. Hence, the minimum in the definition of $\overline{\mathcal{I}}$ is actually attained. Notice that taking constant sequences in the parameter space in (2.5) and using the weak lower semicontinuity of the regularizers $\mathcal{R}_\lambda$ in (H), we conclude that $\overline{\mathcal{J}}_\lambda$ coincides with $\mathcal{J}_\lambda$ whenever $\lambda \in \Lambda$. In that sense, we can think of $\{\overline{\mathcal{J}}_\lambda\}_{\lambda \in \overline{\Lambda}}$ as the extension of the family $\{\mathcal{J}_\lambda\}_{\lambda \in \Lambda}$ to the closure of Λ.

All together, this leads to the extended bi-level problem

(Upper-level)   Minimize $\overline{\mathcal{I}}(\lambda) := \min_{w \in \overline{K}_\lambda} \|w - u^c\|_{L^2(\Omega; \mathbb{R}^N)}^2$ over $\lambda \in \overline{\Lambda}$,

(Lower-level) $\overline{K}_\lambda := \left\{ w \in L^2(\Omega; \mathbb{R}^N) : w_j \in \underset{u \in L^2(\Omega)}{\arg\min} \overline{\mathcal{J}}_{\lambda, j}(u) \text{ for all } j \in \{1, \ldots, N\} \right\}.$

$(\overline{\mathcal{T}})$

The theorem below compares the extended upper level functional $\overline{\mathcal{I}}$ with the relaxation of $\mathcal{I}$ (after trivial extension to $\overline{\Lambda}$ by ∞), that is, with its lower semicontinuous envelope $\mathcal{I}^{\text{rlx}} : \overline{\Lambda} \to [0, \infty]$ given by

$$\mathcal{I}^{\text{rlx}}(\lambda) := \inf \left\{ \liminf_{k \to \infty} \mathcal{I}(\lambda_k) : (\lambda_k)_k \subset \Lambda, \lambda_k \to \lambda \text{ in } \overline{\Lambda} \right\}. \tag{2.7}$$

As we will see, the key assumption to obtain the equality between $\overline{\mathcal{I}}$ and $\mathcal{I}^{\text{rlx}}$ is the Mosco-convergence of the family of regularizers in (2.9), which is stronger than the Γ-convergence of the reconstruction functionals in (2.5). It even implies the Mosco-convergence

$$\overline{\mathcal{J}}_{\lambda, j} = \text{Mosc}(L^2)\text{-} \lim_{\lambda' \to \lambda} \mathcal{J}_{\lambda', j}$$

and, in this case, the limit passage can be performed additively in the fidelity and regularizing term; thus, for all $j \in \{1, \ldots, N\}$, we have

$$\overline{\mathcal{J}}_{\lambda, j}(u) = \|u - u_j^\eta\|_{L^2(\Omega)} + \overline{\mathcal{R}}_\lambda(u) \qquad \text{for } u \in L^2(\Omega). \tag{2.8}$$

Theorem 2.5 *Consider the bi-level optimization problems ($\mathcal{T}$) and ($\overline{\mathcal{T}}$), assume (2.5), and recall the definitions in (2.6) and (2.7). Suppose in addition that*

(i) *the Mosco-limits*

$$\overline{\mathcal{R}}_\lambda := \text{Mosc}(L^2)\text{-} \lim_{\lambda' \to \lambda} \mathcal{R}_{\lambda'} \tag{2.9}$$

exist for each $\lambda \in \overline{\Lambda}$, with λ' taking values on sequences in Λ, and

(ii) *$\overline{K}_\lambda$ is a singleton for every $\lambda \in \overline{\Lambda} \setminus \Lambda$.*

Then, the extension $\overline{\mathcal{I}}$ of $\mathcal{I}$ to the closure $\overline{\Lambda}$ coincides with the relaxation of $\mathcal{I}$, i.e., $\overline{\mathcal{I}} = \mathcal{I}^{\text{rlx}}$ on $\overline{\Lambda}$.

Proof To show that $\overline{\mathcal{I}} \leq \mathcal{I}^{\text{rlx}}$, we take $\lambda \in \overline{\Lambda}$ and let $(\lambda_k)_k \subset \Lambda$ with $\lambda_k \to \lambda$ in $\overline{\Lambda}$ be an admissible sequence for $\mathcal{I}^{\text{rlx}}(\lambda)$ in (2.7). We may even assume that $\infty > \liminf_{k \to \infty} \mathcal{I}(\lambda_k) = \lim_{k \to \infty} \mathcal{I}(\lambda_k)$. Then, recalling (2.2) and fixing $\delta > 0$, we can find $w_k \in K_{\lambda_k}$ such that

$$\lim_{k \to \infty} \mathcal{I}(\lambda_k) \geq \liminf_{k \to \infty} \|w_k - u^c\|_{L^2(\Omega; \mathbb{R}^N)}^2 - \delta.$$

In particular, $(w_k)_k$ is uniformly bounded in $L^2(\Omega; \mathbb{R}^N)$, which allows us to extract an L^2-weakly converging subsequence (not relabeled) with limit $\bar{w} \in L^2(\Omega; \mathbb{R}^N)$. By the properties of Γ-convergence on cluster points of minimizing sequences recalled above (see also [25, Corollary 7.20]), we infer from (2.5) that $\bar{w}_j \in \operatorname{argmin}_{u \in L^2(\Omega)} \overline{\mathcal{J}}_{\lambda, j}(u)$ for all $j \in \{1, \ldots, N\}$; in other words, $\bar{w} \in \overline{K}_\lambda$. Thus,

$$\lim_{k \to \infty} \mathcal{I}(\lambda_k) \geq \|\bar{w} - u^c\|_{L^2(\Omega; \mathbb{R}^N)}^2 - \delta \geq \overline{\mathcal{I}}(\lambda) - \delta.$$

By letting $\delta \to 0$ first, and then taking the infimum over all admissible sequences for $\mathcal{I}^{\mathrm{rlx}}(\lambda)$ in (2.7), it follows that $\overline{\mathcal{I}}(\lambda) \leq \mathcal{I}^{\mathrm{rlx}}(\lambda)$.

To prove the reverse inequality, we start by recalling that for $\lambda \in \Lambda$, $\mathcal{J}_\lambda$ is weakly L^2-lower semicontinuous by Assumption (H); thus, (2.5) yields $\overline{\mathcal{J}}_\lambda = \mathcal{J}_\lambda$ for $\lambda \in \Lambda$. Consequently, $\overline{\mathcal{I}}(\lambda) = \mathcal{I}(\lambda) \geq \mathcal{I}^{\mathrm{rlx}}(\lambda)$ for $\lambda \in \Lambda$. We are then left to consider $\lambda \in \overline{\Lambda} \backslash \Lambda$ and find a sequence $(\lambda_k)_k \subset \Lambda$ converging to λ in $\overline{\Lambda}$ and satisfying $\liminf_{k \to \infty} \mathcal{I}(\lambda_k) \leq \overline{\mathcal{I}}(\lambda)$. To that end, take any $(\lambda_k)_k \subset \Lambda$ with $\lambda_k \to \lambda$ in $\overline{\Lambda}$, and let $w_k \in K_{\lambda_k}$ for $k \in \mathbb{N}$. Recalling (ii), denote by $w_\lambda = (w_{\lambda, 1}, \ldots, w_{\lambda, N})$ the unique element in $\overline{K}_\lambda$. Then, using (2.5) and the equi-coercivity of $(\mathcal{J}_\lambda)_{\lambda \in \Lambda}$, we obtain by the theory of Γ-convergence (see [25, Corollary 7.24]) that $(w_k)_k$ converges weakly in $L^2(\Omega; \mathbb{R}^N)$ to w_λ; moreover, it holds for all $j \in \{1, \ldots, N\}$ that

$$\mathcal{J}_{\lambda_k, j}(w_{k, j}) \to \overline{\mathcal{J}}_{\lambda, j}(w_{\lambda, j}) \qquad \text{as } k \to \infty. \tag{2.10}$$

The following shows that $(w_k)_k$ converges even strongly in $L^2(\Omega; \mathbb{R}^N)$. Indeed, fixing $j \in \{1, \ldots, N\}$, we infer from (2.10) along with the Mosco-convergence of the regularizers in (i) and (2.8) that

$$\begin{aligned}
\|w_{\lambda, j} - u_j^\eta\|_{L^2(\Omega)}^2 + \overline{\mathcal{R}}_\lambda(w_{\lambda, j}) = \overline{\mathcal{J}}_{\lambda, j}(w_{\lambda, j}) &= \lim_{k \to \infty} \mathcal{J}_{\lambda_k, j}(w_{k, j}) \\
&= \lim_{k \to \infty} \left[\|w_{k, j} - u_j^\eta\|_{L^2(\Omega)}^2 + \mathcal{R}_{\lambda_k}(w_{k, j}) \right] \\
&\geq \limsup_{k \to \infty} \|w_{k, j} - u_j^\eta\|_{L^2(\Omega)}^2 + \overline{\mathcal{R}}_\lambda(w_{\lambda, j}).
\end{aligned}$$

Hence, $\|w_{\lambda, j} - u_j^\eta\|_{L^2(\Omega)}^2 \geq \limsup_{k \to \infty} \|w_{k, j} - u_j^\eta\|_{L^2(\Omega)}^2$, which together with the weak lower semicontinuity of the L^2-norm yields

$$\lim_{k \to \infty} \|w_{k, j} - u_j^\eta\|_{L^2(\Omega)}^2 = \|w_{\lambda, j} - u_j^\eta\|_{L^2(\Omega)}^2;$$

thus, $w_k \to w_\lambda$ strongly in $L^2(\Omega; \mathbb{R}^N)$ using the combination of weak convergence and convergence of norms by the Radon–Riesz property. With this, we finally conclude that

$$\begin{aligned}
\liminf_{k \to \infty} \mathcal{I}(\lambda_k) \leq \liminf_{k \to \infty} \|w_k - u^c\|_{L^2(\Omega; \mathbb{R}^N)}^2 &= \|w_\lambda - u^c\|_{L^2(\Omega; \mathbb{R}^N)}^2 \\
&= \min_{w \in \overline{K}_\lambda} \|w - u^c\|_{L^2(\Omega; \mathbb{R}^N)}^2 = \overline{\mathcal{I}}(\lambda),
\end{aligned}$$

finishing the proof.   □

Remark 2.6 By inspecting the proof, it becomes clear that the estimate $\overline{\mathcal{I}} \leq \mathcal{I}^{\text{rlx}}$ holds without the additional assumptions (i) and (ii) from the previous theorem; in other words, $\overline{\mathcal{I}}$ always provides a lower bound for the relaxation of $\mathcal{I}$.

The identity $\overline{\mathcal{I}} = \mathcal{I}^{\text{rlx}}$ mail fail if either of the assumptions (i) or (ii) in Theorem 2.5 is dropped as the following example shows.

Example 2.7 a) To see why (i) is necessary, consider $\Lambda = (0, 1]$, a single data point (u^c, u^η) with $u^c = u^\eta = 0$, and

$$\mathcal{R}_\lambda = \frac{1}{\lambda} \| \cdot - v_\lambda \|^2_{L^2(\Omega)} \qquad \text{with } v_\lambda = v(\cdot/\lambda) \in L^2(\Omega)$$

for a given $v \in L^\infty(\mathbb{R}^n)$ with the properties that v is $(0, 1)^n$-periodic, $v \in \{-1, 1\}$ almost everywhere, and $\int_{(0,1)^n} v \, dx = 0$. Under these specifications, the Γ-limits $\overline{\mathcal{J}}_\lambda = \Gamma(w\text{-}L^2)\text{-} \lim_{\lambda' \to \lambda} \mathcal{J}_{\lambda'}$ (cf. (2.5) and (2.1)) exist and are given by

$$\overline{\mathcal{J}}_\lambda(u) = \begin{cases} \|u\|^2_{L^2(\Omega)} + \frac{1}{\lambda}\|u - v_\lambda\|^2_{L^2(\Omega)} & \text{for } \lambda \in (0, 1], \\ |\Omega| + \chi_{\{0\}}(u) & \text{for } \lambda = 0, \end{cases} \tag{2.11}$$

where χ_E denotes the indicator function of a set $E \subset L^2(\Omega)$, i.e.,

$$\chi_E(u) = \begin{cases} 0 & \text{if } u \in E, \\ \infty & \text{if } u \notin E, \end{cases} \qquad \text{for } u \in L^2(\Omega).$$

The non-trivial case is when $\lambda = 0$. In this case, we observe that we can take $(v_{\lambda'})_{\lambda'}$ as a recovery sequence for $u = 0$ because it converges weakly in $L^2(\Omega)$ as $\lambda' \to 0$ to $\int_{(0,1)^n} v \, dx = 0$ by the Riemann–Lebesgue lemma for periodically oscillating sequences. For the liminf inequality, let $u_{\lambda'} \rightharpoonup u$ as $\lambda' \to 0$ and suppose without loss of generality that $\sup_{\lambda'} \mathcal{R}_{\lambda'}(u_{\lambda'}) < \infty$. Then, $u_{\lambda'} = v_{\lambda'} + r_{\lambda'}$ with $r_{\lambda'} \to 0$ in $L^2(\Omega)$ as $\lambda' \to 0$, which implies $u = 0$ and, recalling that $v \in \{-1, 1\}$ almost everywhere,

$$\liminf_{\lambda' \to 0} \mathcal{J}_{\lambda'}(u_{\lambda'}) \geq \lim_{\lambda' \to 0} \|v_{\lambda'} + r_{\lambda'}\|^2_{L^2(\Omega)} = \lim_{\lambda' \to 0} \|v_{\lambda'}\|^2_{L^2(\Omega)} = |\Omega| = \overline{\mathcal{J}}_0(0),$$

which completes the proof of (2.11) when $\lambda = 0$.

In view of (2.11), one can now read off that $K_\lambda = \overline{K}_\lambda = \{v_\lambda/(1+\lambda)\}$ for $\lambda \in (0, 1]$ and $\overline{K}_0 = \{0\}$. In particular, condition (ii) on the uniqueness of minimizers of the extended lower-level problem is fulfilled here. Hence,

$$\mathcal{I}(\lambda) = \left(\frac{1}{1+\lambda}\right)^2 |\Omega| \tag{2.12}$$

for $\lambda \in (0, 1]$, and

$$\overline{\mathcal{I}}(\lambda) = \begin{cases} \left(\dfrac{1}{1+\lambda}\right)^2 |\Omega| & \text{if } \lambda \in (0, 1], \\ 0 & \text{if } \lambda = 0 \end{cases}$$

for $\lambda \in [0, 1]$. It is immediate to see from (2.12) that

$$\overline{\mathcal{I}}(0) = 0 < |\Omega| = \mathcal{I}^{\text{rlx}}(0).$$

Notice that this example hinges on the fact that the minimizers $v_\lambda/(1+\lambda)$ only converge weakly as $\lambda \to 0$, which, in view of the proof of Theorem 2.5, implies that the family of regularizers $\{\mathcal{R}_\lambda\}_{\lambda \in \Lambda}$ does not Mosco-converge in $L^2(\Omega)$ in the sense of (2.9), thus failing to satisfy (i).

   b) For the necessity of (ii), consider $\Lambda = (0, 1]$, a single data point (u^c, u^η) with $u^c = 0$ and $\|u^\eta\|^2_{L^2(\Omega)} = 1$, and

$$\mathcal{R}_\lambda(u) = \begin{cases} \lambda & \text{if } u = 0, \\ 1 & \text{if } u \neq 0. \end{cases}$$

While it is straightforward to check that condition (i) in Theorem 2.5 regarding the Mosco-limits of $\{\mathcal{R}_\lambda\}_{\lambda \in \Lambda}$ is satisfied with

$$\overline{\mathcal{R}}_\lambda(u) = \begin{cases} \lambda & \text{if } u = 0, \\ 1 & \text{if } u \neq 0 \end{cases}$$

for $\lambda \in [0, 1]$, which clearly coincides with $\mathcal{R}_\lambda$ for $\lambda \in \Lambda = (0, 1]$, condition (ii) fails. Indeed, it follows from (2.8) that $\overline{\mathcal{J}}_\lambda(u^\eta) = \mathcal{R}_\lambda(u^\eta) = 1$ and $\overline{\mathcal{J}}_\lambda(0) = \|u^\eta\|^2_{L^2(\Omega)} + \lambda = 1 + \lambda$ for all $\lambda \in [0, 1]$. Consequently, for $\lambda \in (0, 1]$, we have $\overline{\mathcal{J}}_\lambda = \mathcal{J}_\lambda$ and u^η is its unique minimizer; in contrast, for $\lambda = 0$, $\overline{\mathcal{J}}_0$ has two minimizers, namely $\overline{K}_0 = \{u^\eta, 0\} = \{u^\eta, u^c\}$. Finally, we observe that the conclusion of Theorem 2.5 fails here because

$$\overline{\mathcal{I}}(0) = 0 \quad \text{and} \quad \mathcal{I}(\lambda) = \|u^c - u^\eta\|^2_{L^2(\Omega)} = 1 \text{ for all } \lambda \in (0, 1],$$

which yields $\overline{\mathcal{I}}(0) = 0 < 1 = \mathcal{I}^{\text{rlx}}(0)$.

   The following result is a direct consequence of Theorem 2.5 and standard properties of relaxation.

Corollary 2.8 *Under the assumptions of Theorem 2.5 and if $\overline{\Lambda}$ is compact, it holds that:*

(i) *The extension $\overline{\mathcal{I}}$ has at least one minimizer and*

$$\min_{\overline{\Lambda}} \overline{\mathcal{I}} = \inf_\Lambda \mathcal{I}.$$

(ii) *Any minimizing sequence $(\lambda_k)_k \subset \Lambda$ of $\mathcal{I}$ converges up to subsequence to a minimizer $\lambda \in \overline{\Lambda}$ of $\overline{\mathcal{I}}$.*

(iii) *If $\lambda \in \Lambda$ minimizes $\overline{\mathcal{I}}$, then λ is also a minimizer of $\mathcal{I}$.*

We conclude this section on the theoretical framework with a brief comparison with related works on optimal control problems. By setting $K = \{(w, \lambda) \in L^2(\Omega) \times \Lambda : w \in K_\lambda\}$, the bi-level optimization problem ($\mathcal{I}$) can be equivalently rephrased into minimizing

$$\widehat{\mathcal{I}}(u, \lambda) = \|u - u^c\|_{L^2(\Omega)}^2 + \chi_K(u, \lambda), \qquad (u, \lambda) \in L^2(\Omega) \times \Lambda,$$

as a functional of two variables; observe that

$$\mathcal{I}(\lambda) = \inf_{w \in L^2(\Omega)} \widehat{\mathcal{I}}(w, \lambda).$$

Similar functionals and their relaxations have been studied in the literature, including [9, 16, 17]. Especially the paper [9] by Belloni, Buttazzo, & Freddi, where the authors propose to extend the control space to its closure and find a description of the relaxed optimal control problem, shares many parallels with our results. Apart from some differences in the assumptions and abstract set-up, the main reason why their results are not applicable here is the continuity condition of the cost functional with respect to the state variable [9, Eq. (2.11)]. In our setting, this would translate into weak continuity of the L^2-norm, which is clearly false. The argument in the proof of Theorem 2.5 exploiting the Mosco-convergence of the regularizers (see (2.9)) is precisely what circumvents this issue.

3 Learning the Optimal Weight of the Regularization Term

In this section, we study the optimization of a weight factor, often called tuning parameter, in front of a fixed regularization term. Such tuning parameters are typically employed in practical implementations of variational denoising models to adjust the best level of regularization. This setting constitutes a simple, yet non-trivial, application of our general theory and therefore helps to exemplify the abstract results from the previous section.

As above, $\Omega \subset \mathbb{R}^n$ is a bounded open set and $u^c, u^\eta \in L^2(\Omega; \mathbb{R}^N)$ are the given data representing pairs of clean and noisy images. We take $\Lambda = (0, \infty)$ describing the range of a weight factor and, to distinguish the various parameters throughout this paper, denote by α an arbitrary point in $\overline{\Lambda} = [0, \infty]$. For a fixed map $\mathcal{R} : L^2(\Omega) \to [0, \infty]$ with the properties that

(H1$_\alpha$) $\mathcal{R}$ is convex, vanishes exactly on constant functions, and Dom $\mathcal{R}$ is dense in $L^2(\Omega)$,

(H2$_\alpha$) $\mathcal{R}$ is lower semicontinuous on $L^2(\Omega)$,

we define the weighted regularizers

$$\mathcal{R}_\alpha = \alpha \mathcal{R} \qquad \text{for } \alpha \in (0, \infty). \tag{2.1}$$

Note that (H1$_\alpha$) and (H2$_\alpha$) imply that the family $\{\mathcal{R}_\alpha\}_{\alpha \in (0,\infty)}$ satisfies (H) because convexity and lower semicontinuity yield weak lower semicontinuity, making this setting match with the framework of Sect. 2.

Following the definition of the training scheme ($\mathcal{T}$), we introduce here for $\alpha \in (0, \infty)$ and $j \in \{1, \ldots, N\}$ the reconstruction functionals

$$\mathcal{J}_{\alpha,j}(u) = \|u - u_j^\eta\|_{L^2(\Omega)}^2 + \mathcal{R}_\alpha(u) \quad \text{for } u \in L^2(\Omega),$$

cf. (2.1), and consider accordingly the upper level functional $\mathcal{I} : (0, \infty) \to [0, \infty)$ given by

$$\mathcal{I}(\alpha) = \inf_{w \in K_\alpha} \|w - u^c\|_{L^2(\Omega;\mathbb{R}^N)}^2 \quad \text{for } \alpha \in (0, \infty), \tag{2.2}$$

with $K_\alpha = K_{\alpha,1} \times \cdots \times K_{\alpha,N}$ and $K_{\alpha,j} = \arg\min_{u \in L^2(\Omega)} \mathcal{J}_{\alpha,j}(u)$, cf. (2.2). Further, the following set of hypotheses on the training data will play a crucial role for our main result in this section (Theorem 3.2):

(H3$_\alpha$) It holds that

$$\sum_{j=1}^N \mathcal{R}(u_j^c) < \sum_{j=1}^N \mathcal{R}(u_j^\eta);$$

(H4$_\alpha$) the data u^η and u^c satisfy

$$\|u^\eta - u^c\|_{L^2(\Omega;\mathbb{R}^N)}^2 < \left\| \fint_\Omega u^\eta \, \mathrm{d}x - u^c \right\|_{L^2(\Omega;\mathbb{R}^N)}^2.$$

Remark 3.1 (*Discussion of the hypotheses* (H1$_\alpha$)–(H4$_\alpha$)) a) Note that (H1$_\alpha$) implies that the set of minimizers for the reconstruction functionals, K_α, has cardinality one, owing to the convexity of $\mathcal{R}$ and the strict convexity of the fidelity term, considering also that $\mathcal{J}_{\alpha,j} \not\equiv \infty$. In the following, we write $w^{(\alpha)} = (w_1^{(\alpha)}, \ldots, w_N^{(\alpha)}) \in L^2(\Omega; \mathbb{R}^N)$ for the single element of K_α, i.e., $K_\alpha = \{w^{(\alpha)}\}$.

b) An example of a nonlocal regularizer satisfying (H1$_\alpha$) and (H2$_\alpha$) is

$$\mathcal{R}(u) := \int_\Omega \int_\Omega a(x, y) \, g(u(x) - u(y)) \, \mathrm{d}x \, \mathrm{d}y \quad \text{for } u \in L^2(\Omega),$$

where $g : \mathbb{R} \to [0, \infty)$ is a convex function such that $g^{-1}(0) = \{0\}$ and $a : \Omega \times \Omega \to [0, \infty]$ is a suitable kernel ensuring that $C_c^\infty(\Omega) \subset \mathrm{Dom}\,\mathcal{R}$. As an explicit choice, one can take $g(t) = t^p$ for $t \in \mathbb{R}$ and $a(x, y) = |y - x|^{-n-sp}$ for $x, y \in \Omega$ with some $s \in (0, 1)$ and $p \geq 1$, which corresponds to a fractional Sobolev regularization.

c) Assumption (H3$_\alpha$) asserts that the regularizer penalizes the noisy images more than the clean ones on average. This is a natural condition because any good regularizer should reflect the prior knowledge on the training data, favoring the clean images.

d) The second condition on the data, (H4$_\alpha$), means that the noisy image lies closer to the clean image than its mean value, which can be considered a reasonable assumption in the case of moderate noise and a non-trivial ground truth. Indeed, suppose the noise is bounded by $\|u_j^\eta - u_j^c\|_{L^2(\Omega)} \leq \delta$ for all $j \in \{1, \dots, N\}$ and some $\delta > 0$; then, (H4$_\alpha$) is satisfied if

$$\left\| \fint_\Omega u_j^c \, dx - u_j^c \right\|_{L^2(\Omega)} > \delta\left(1 + |\Omega|^{-\frac{1}{2}}\right) \quad \text{for all } j \in \{1, \dots, N\}$$

because

$$\left\| \fint_\Omega u_j^\eta \, dx - u_j^c \right\|_{L^2(\Omega)} \geq \left\| \fint_\Omega u_j^c \, dx - u_j^c \right\|_{L^2(\Omega)} - \left\| \fint_\Omega (u_j^\eta - u_j^c) \, dx \right\|_{L^2(\Omega)}$$
$$> \delta\left(1 + |\Omega|^{-\frac{1}{2}}\right) - |\Omega|^{-\frac{1}{2}} \|u_j^\eta - u_j^c\|_{L^2(\Omega)}$$
$$\geq \delta \geq \|u_j^\eta - u_j^c\|_{L^2(\Omega)},$$

where the second inequality is due to Jensen's inequality.

Next, we prove that the assumptions (H1$_\alpha$)–(H4$_\alpha$) on the regularization term and on the training set give rise to optimal weight parameters that stay away from the extremal regimes, $\alpha = 0$ and $\alpha = \infty$. Thus, in this case, the bi-level parameter optimization procedure preserves the structure of the original denoising model.

Theorem 3.2 (Structure preservation) *Suppose that* (H1$_\alpha$)–(H4$_\alpha$) *hold. Then, the learning scheme corresponding to the minimization of* $\mathcal{I}$ *in* (2.2) *admits a solution* $\bar\alpha \in (0, \infty)$.

A related statement in the same spirit can be found in [30, Theorem 1], although some of the details of the proof were not entirely clear to us. Our proof of Theorem 3.2 is based on a different approach and hinges on the following two lemmas, the first of which determines the Mosco-limits of the regularizers, and thereby provides an explicit formula of the extension $\overline{\mathcal{I}}$ of $\mathcal{I}$ as introduced in (2.6).

Proposition 3.3 (Mosco-convergence of the regularizer) *Let* $\mathcal{R} : L^2(\Omega) \to [0, \infty]$ *satisfy* (H1$_\alpha$) *and* (H2$_\alpha$), *and let* $\{\mathcal{R}_\alpha\}_{\alpha \in (0,\infty)}$ *be as in* (2.1). *Then,*

$$\overline{\mathcal{R}}_\alpha := \text{Mosc}(L^2)\text{-}\lim_{\alpha' \to \alpha} \mathcal{R}_{\alpha'} = \begin{cases} \mathcal{R}_\alpha & \text{for } \alpha \in (0, \infty), \\ 0 & \text{for } \alpha = 0, \\ \chi_C & \text{for } \alpha = \infty, \end{cases} \quad (2.3)$$

for $\alpha \in [0, \infty]$, *where* χ_C *is the indicator function of* $C := \{u \in L^2(\Omega) : u$ *is constant*\}.

Proof Using standard arguments, we show that the Mosco-limit of $(\mathcal{R}_{\alpha_k})_k$ exists for every sequence $(\alpha_k)_k$ of positive real numbers with $\alpha_k \to \alpha \in [0, \infty]$, and corresponds to the right hand side of (2.3).

Case 1: $\alpha \in (0, \infty)$. Using (H2$_\alpha$) for the liminf inequality and a constant recovery sequence for the upper bound, we conclude that the Mosco-limit of $(\mathcal{R}_{\alpha_k})_k$ coincides with $\mathcal{R}_\alpha$.

Case 2: $\alpha = 0$. The liminf inequality is trivial. For the recovery sequence, take $u \in L^2(\Omega)$ and let $(u_k)_k \subset \mathrm{Dom}\,\mathcal{R}$ converge strongly to u in $L^2(\Omega)$, which is feasible due to (H1$_\alpha$). By possibly repeating certain entries of the sequence $(u_k)_k$ (not relabeled), one can slowdown the speed at which $\mathcal{R}(u_k)$ potentially blows up and assume that $\alpha_k \mathcal{R}(u_k) \to 0$ as $k \to \infty$. Thus,

$$\lim_{k \to \infty} \mathcal{R}_{\alpha_k}(u_k) = \lim_{k \to \infty} \alpha_k \mathcal{R}(u_k) = 0.$$

Case 3: $\alpha = \infty$. The limsup inequality follows by choosing constant recovery sequences. For the proof of the lower bound, consider $u_k \rightharpoonup u$ in $L^2(\Omega)$ with $r := \sup_{k \in \mathbb{N}} \alpha_k \mathcal{R}(u_k) = \sup_{k \in \mathbb{N}} \mathcal{R}_{\alpha_k}(u_k) < \infty$. Then, along with the weak lower semicontinuity of $\mathcal{R}$ (see Remark 3.1 a)),

$$\mathcal{R}(u) \leq \liminf_{k \to \infty} \mathcal{R}(u_k) \leq \lim_{k \to \infty} \frac{r}{\alpha_k} = 0.$$

This shows that $\mathcal{R}(u) = 0$, which implies by the assumption on the zero level set of $\mathcal{R}$ in (H1$_\alpha$) that u is constant, i.e., $u \in C$.   □

As a consequence of the previous proposition, we deduce that the extension $\overline{\mathcal{I}} : \overline{\Lambda} \to [0, \infty]$ of $\mathcal{I}$ in the sense of (2.6) can be explicitly determined as

$$\overline{\mathcal{I}}(\alpha) = \begin{cases} \mathcal{I}(\alpha) & \text{for } \alpha \in (0, \infty), \\ \|u^\eta - u^c\|^2_{L^2(\Omega; \mathbb{R}^N)} & \text{for } \alpha = 0, \\ \left\| \fint_\Omega u^\eta \, \mathrm{d}x - u^c \right\|^2_{L^2(\Omega; \mathbb{R}^N)} & \text{for } \alpha = \infty. \end{cases} \tag{2.4}$$

Indeed, a straight-forward calculation of the unique componentwise minimizer of the extended reconstruction functionals $\overline{\mathcal{J}}_\alpha$ at the boundary points $\alpha = 0$ and $\alpha = \infty$ leads to

$$\overline{K}_0 = \{u^\eta\} \quad \text{and} \quad \overline{K}_\infty = \left\{ \fint_\Omega u^\eta \, \mathrm{d}x \right\}.$$

Since the assumptions (i) and (ii) of Theorem 2.5 are satisfied, $\overline{\mathcal{I}}$ coincides with the relaxation $\mathcal{I}^{\mathrm{rlx}}$. By Corollary 2.8 (i), $\overline{\mathcal{I}}$ attains its minimum at some $\bar{\alpha} \in [0, \infty]$. The degenerate cases $\bar{\alpha} \in \{0, \infty\}$ cannot be excluded a priori, but the next lemma shows that the minimum is attained in the interior $(0, \infty)$ under suitable assumptions on the training data.

Lemma 3.4 *Suppose that* (H1$_\alpha$) *and* (H2$_\alpha$) *hold, and let* $K_\alpha = \{w^{(\alpha)}\}$ *with* $w^{(\alpha)} = (w_1^{(\alpha)}, \ldots, w_N^\alpha) \in L^2(\Omega; \mathbb{R}^N)$ *for* $\alpha \in (0, \infty)$, *cf. Remark* 3.1a).

(i) *Under the additional assumption* (H3$_\alpha$)*, there exists* $\alpha \in (0, \infty)$ *such that*

$$\|w^{(\alpha)} - u^c\|^2_{L^2(\Omega;\mathbb{R}^N)} < \|u^\eta - u^c\|^2_{L^2(\Omega;\mathbb{R}^N)}.$$

(ii) *Under the additional assumption* (H4$_\alpha$)*, there exists* $\alpha_0 \in (0, \infty)$ *such that, for all* $\alpha \in (0, \alpha_0)$,

$$\|w^{(\alpha)} - u^c\|^2_{L^2(\Omega;\mathbb{R}^N)} < \left\|\fint_\Omega u^\eta \, dx - u^c\right\|^2_{L^2(\Omega;\mathbb{R}^N)}. \tag{2.5}$$

Proof We start by providing two useful auxiliary results about the asymptotic behavior of the reconstruction vector $w^{(\alpha)}$ as α tends to zero; precisely,

$$\lim_{\alpha \to 0} \|w^{(\alpha)} - u^\eta\|_{L^2(\Omega;\mathbb{R}^N)} = 0 \quad \text{and} \quad \lim_{\alpha \to 0} \mathcal{R}(w_j^{(\alpha)}) = \mathcal{R}(u_j^\eta) \text{ for every } j \in \{1, \ldots, N\}. \tag{2.6}$$

Fix $j \in \{1, \ldots, N\}$ and let $(\alpha_k)_k \subset (0, \infty)$ be such that $\alpha_k \to 0$ as $k \to \infty$. Take $u \in \text{Dom}\, \mathcal{R}$ with $\|u - u_j^\eta\|^2_{L^2(\Omega)} \le \varepsilon$ for some $\varepsilon > 0$, which is possible by (H1$_\alpha$). Then, the minimality of $w_j^{(\alpha_k)}$ for $\mathcal{J}_{\alpha_k, j}$ yields

$$\|w_j^{(\alpha_k)} - u_j^\eta\|^2_{L^2(\Omega)} \le \mathcal{J}_{\alpha_k, j}(w_j^{(\alpha_k)}) \le \mathcal{J}_{\alpha_k, j}(u) = \|u - u_j^\eta\|^2_{L^2(\Omega)} + \alpha_k \mathcal{R}(u) \le \varepsilon + \alpha_k \mathcal{R}(u).$$

Since $\mathcal{R}(u) < \infty$, we find

$$\limsup_{k \to \infty} \|w_j^{(\alpha_k)} - u_j^\eta\|^2_{L^2(\Omega)} \le \varepsilon,$$

which proves the first part of (2.6) due to the arbitrariness of ε. Exploiting the minimality of $w_j^{(\alpha)}$ for $\mathcal{J}_{\alpha, j}$ again with $\alpha \in (0, \infty)$ entails

$$\alpha \mathcal{R}(w_j^{(\alpha)}) = \mathcal{R}_\alpha(w_j^{(\alpha)}) \le \mathcal{J}_{\alpha, j}(w_j^{(\alpha)}) \le \mathcal{J}_{\alpha, j}(u_j^\eta) = \mathcal{R}_\alpha(u_j^\eta) = \alpha \mathcal{R}(u_j^\eta);$$

hence, $\mathcal{R}(w_j^{(\alpha)}) \le \mathcal{R}(u_j^\eta)$ and, together with the first part of (2.6) and the lower semicontinuity of $\mathcal{R}$ by (H2$_\alpha$), it follows then that

$$\mathcal{R}(u_j^\eta) \ge \limsup_{k \to \infty} \mathcal{R}(w_j^{(\alpha_k)}) \ge \liminf_{k \to \infty} \mathcal{R}(w_j^{(\alpha_k)}) \ge \mathcal{R}(u_j^\eta).$$

Thus, $\lim_{k \to \infty} \mathcal{R}(w_j^{(\alpha_k)}) = \mathcal{R}(u_j^\eta)$, showing the second part of (2.6).

Regarding (i), we observe that the minimality of $w_j^{(\alpha)}$ for $\mathcal{J}_{\alpha, j}$ for any $\alpha \in (0, \infty)$ and $j \in \{1, \ldots, N\}$ imposes the necessary condition $0 \in \partial \mathcal{J}_{\alpha, j}(w_j^{(\alpha)})$ or, equivalently,

$$2(u_j^\eta - w_j^{(\alpha)}) \in \partial \mathcal{R}_\alpha(w_j^{(\alpha)}) = \alpha \partial \mathcal{R}(w_j^{(\alpha)}),$$

where $\partial \mathcal{C}(u) \in L^2(\Omega)' \cong L^2(\Omega)$ is the subdifferential of a convex function $\mathcal{C} : L^2(\Omega) \to [0, \infty]$ at $u \in L^2(\Omega)$. Then,

$$\|u_j^\eta - u_j^c\|_{L^2(\Omega)}^2 - \|w_j^{(\alpha)} - u_j^c\|_{L^2(\Omega)}^2 = 2\langle u_j^\eta - w_j^{(\alpha)}, w_j^{(\alpha)} - u_j^c \rangle_{L^2(\Omega)} + \|w_j^{(\alpha)} - u_j^\eta\|_{L^2(\Omega)}^2$$
$$\geq \mathcal{R}_\alpha(w_j^{(\alpha)}) - \mathcal{R}_\alpha(u_j^c) = \alpha\big(\mathcal{R}(w_j^{(\alpha)}) - \mathcal{R}(u_j^c)\big),$$

where $\langle \cdot, \cdot \rangle_{L^2(\Omega)}$ denotes the standard $L^2(\Omega)$-inner product. Summing both sides over $j \in \{1, \dots, N\}$ results in

$$\|u^\eta - u^c\|_{L^2(\Omega;\mathbb{R}^N)}^2 - \|w^{(\alpha)} - u^c\|_{L^2(\Omega;\mathbb{R}^N)}^2 \geq \alpha \sum_{j=1}^N \big(\mathcal{R}(w_j^{(\alpha)}) - \mathcal{R}(u_j^c)\big).$$

By (H3$_\alpha$) in combination with the second part of (2.6), there exists $\alpha_0 > 0$ such that

$$\sum_{j=1}^N \mathcal{R}(w_j^{(\alpha)}) > \sum_{j=1}^N \mathcal{R}(u_j^c)$$

for all $\alpha \in (0, \alpha_0)$, so that choosing $\bar{\alpha} \in (0, \alpha_0)$ concludes the proof of (i).

To show (ii), we exploit the first limit in (2.6). Due to (H4$_\alpha$), it follows then for any $(\alpha_k)_k$ of positive real numbers with $\alpha_k \to 0$ as $k \to \infty$ that

$$\limsup_{k \to \infty} \|w^{(\alpha_k)} - u^c\|_{L^2(\Omega;\mathbb{R}^N)} \leq \limsup_{k \to \infty} \|w^{(\alpha_k)} - u^\eta\|_{L^2(\Omega;\mathbb{R}^N)} + \|u^\eta - u^c\|_{L^2(\Omega;\mathbb{R}^N)}$$
$$< \left\| \fint_\Omega u^\eta \, dx - u^c \right\|_{L^2(\Omega;\mathbb{R}^N)},$$

which gives rise to (2.5) for all k sufficiently large. $\square$

Proof of Theorem 3.2 Since $\overline{\mathcal{I}}$ in (2.4) attains its infimum at a point $\bar{\alpha} \in (0, \infty)$ by Lemma 3.4, we conclude from Corollary 2.8 (iii) that $\bar{\alpha}$ is also a minimizer of $\mathcal{I}$. $\square$

Let us finally remark that the assumptions (H3$_\alpha$) and (H4$_\alpha$) on the training data are necessary to obtain structure preservation in the sense of Theorem 3.2.

Remark 3.5 To see that (H3$_\alpha$) and (H4$_\alpha$) can generally not be dropped, consider, for example, a regularizer $\mathcal{R} : L^2(\Omega) \to [0, \infty]$ that satisfies (H1$_\alpha$) and (H2$_\alpha$) and is 2-homogeneous, i.e., $\mathcal{R}(\mu u) = \mu^2 \mathcal{R}(u)$ for all $u \in L^2(\Omega)$ and $\mu \in \mathbb{R}$. With a single, non-constant noisy image $u^\eta \in L^2(\Omega)$, so that $\mathcal{R}(u^\eta) \neq 0$, one has for any $\alpha \in (0, \infty)$ that the quadratic polynomial

$$\mu \mapsto \mathcal{J}_\alpha(\mu u^\eta) = (1 - \mu)^2 \|u^\eta\|_{L^2(\Omega)}^2 + \mu^2 \alpha \mathcal{R}(u^\eta),$$

is not minimized at $\mu = 0$ or $\mu = 1$ because the derivative with respect to μ does not vanish there. Hence,

$$\mathcal{J}_\alpha(w^{(\alpha)}) < \mathcal{J}_\alpha(0) \quad \text{and} \quad \mathcal{J}_\alpha(w^{(\alpha)}) < \mathcal{J}_\alpha(u_\eta).$$

As a result, it follows that

$$w^{(\alpha)} \notin \{0, u^\eta\}.$$

If we now take $u^c = 0$ and suppose additionally that u^η has zero mean value, then $\mathcal{I}(\alpha) > 0$ for all $\alpha \in (0, \infty)$, while clearly $\overline{\mathcal{I}}(\infty) = 0$, that is, the minimum of $\overline{\mathcal{I}}$ is only attained at the boundary point $\alpha = \infty$. Similarly, for $u^c = u^\eta$, the unique minimizer of $\overline{\mathcal{I}}$ is $\alpha = 0$.

4 Optimal Integrability Exponents

Here, we study the optimization of an integrability parameter, p, for a fixed nonlocal regularizer. Our motivation comes from the appearance of different L^p-norms in image processing, such as in quadratic, TV, and Lipschitz regularization [50, Sect. 4]. We focus on the parameter range $\Lambda = [1, \infty)$ with closure $\overline{\Lambda} = [1, \infty]$, paying particular attention to the structural change occurring at $p = \infty$.

Let $\Omega \subset \mathbb{R}^n$ be a bounded Lipschitz domain and consider a function $f : \Omega \times \Omega \times \mathbb{R} \times \mathbb{R} \to [0, \infty)$ that is Carathéodory, i.e., measurable in the first two and continuous with respect to the last two variables, and that satisfies the following bounds and convexity condition:

(H1$_p$) There exist $M, \delta > 0$ and $\beta \in [0, 1]$ such that for all $\xi, \zeta \in \mathbb{R}$, we have

$$f(x, y, \xi, \zeta) \leq M \left(\frac{|\xi - \zeta|}{|x - y|^\beta} + |\xi| + |\zeta| + 1 \right) \quad \text{for a.e. } x, y \in \Omega,$$

and

$$M^{-1} \frac{|\xi - \zeta|}{|x - y|^\beta} - M \leq f(x, y, \xi, \zeta) \quad \text{for a.e. } x, y \in \Omega \text{ with } |x - y| < \delta.$$

(H2$_p$) f is separately convex in the second two variables, i.e., $f(x, y, \cdot, \zeta)$ and $f(x, y, \xi, \cdot)$ are convex for a.e. $x, y \in \Omega$ and every $\xi, \zeta \in \mathbb{R}^n$.

In this setting, we take $p \in [1, \infty)$ and consider the regularization term $\mathcal{R}_p : L^2(\Omega) \to [0, \infty]$ defined by

$$\mathcal{R}_p(u) := \left(\frac{1}{|\Omega \times \Omega|} \int_\Omega \int_\Omega f^p(x, y, u(x), u(y)) \, dx \, dy \right)^{1/p}. \tag{2.1}$$

Remark 4.1 a) Since the regularizer $\mathcal{R}_p$ is invariant under symmetrization, one can assume without loss of generality that f is symmetric in the sense that $f(x, y, \xi, \zeta) = f(y, x, \zeta, \xi)$ for all $x, y \in \Omega$ and $\xi, \zeta \in \mathbb{R}$.

b) Let $p, q \in [1, \infty)$ with $p > q$. Hölder's inequality then yields for every $u \in \text{Dom} \, \mathcal{R}_p = \{u \in L^2(\Omega) : \mathcal{R}_p(u) < \infty\}$ that

$$\left(\int_\Omega \int_\Omega f^p(x, y, u(x), u(y)) \, dx \, dy \right)^{1/p} \geq |\Omega \times \Omega|^{\frac{q-p}{pq}} \left(\int_\Omega \int_\Omega f^q(x, y, u(x), u(y)) \, dx \, dy \right)^{1/q},$$

which translates into $\mathcal{R}_p(u) \geq \mathcal{R}_q(u)$; in particular, Dom $\mathcal{R}_p \subset$ Dom $\mathcal{R}_q$.

A basic example of a symmetric Carathéodory function f satisfying (H1$_p$) with $\beta = 0$ and (H2$_p$) is

$$f(x, y, \xi, \zeta) = a(x - y)|\xi - \zeta| \quad \text{for } x, y \in \Omega \text{ and } \xi, \zeta \in \mathbb{R},$$

where $a \in L^\infty(\mathbb{R}^n)$ is an even function such that ess $\inf_{\mathbb{R}^n} a > 0$. Another example of such a function f with $\beta = 1$ in (H1$_p$) is

$$f(x, y, \xi, \zeta) = b\frac{|\xi - \zeta|}{|x - y|} \quad \text{for } x, y \in \Omega \text{ and } \xi, \zeta \in \mathbb{R},$$

with $b > 0$; note that for the $p > n$ case, the corresponding regularizer $\mathcal{R}_p$ is, up to a multiplicative constant, the Gagliardo semi-norm of the fractional Sobolev space $W^{1-\frac{n}{p}, p}(\Omega)$.

Before showing how the framework of Sect. 2 can be applied here, let us first collect and discuss a few properties of the regularizers $\mathcal{R}_p$ with $p \in [1, \infty)$. We introduce the notation

$$[u]_{p,\beta} := \left(\int_\Omega \int_\Omega \frac{|u(x) - u(y)|^p}{|x - y|^{\beta p}} \, \mathrm{d}x \, \mathrm{d}y \right)^{1/p}$$

to indicate a suitable (p, β)-nonlocal seminorm. Our first lemma shows that the boundedness of the regularizer $\mathcal{R}_p$ is equivalent to the simultaneous boundedness of the L^p-norm and of the (p, β)-seminorm.

Lemma 4.2 *There exists a constant $C > 0$, depending on n, p, Ω, M, δ, and β, such that*

$$\|u\|_{L^p(\Omega)} \leq C\big(\mathcal{R}_p(u) + \|u\|_{L^2(\Omega)} + 1\big), \tag{2.2}$$

$$[u]_{p,\beta} \leq C\big(\mathcal{R}_p(u) + \|u\|_{L^p(\Omega)} + 1\big), \tag{2.3}$$

and

$$\mathcal{R}_p(u) \leq C\big([u]_{p,\beta} + \|u\|_{L^p(\Omega)} + 1\big) \tag{2.4}$$

for all $u \in L^2(\Omega)$, and for all $p \in [1, \infty)$.

Proof Properties (2.2) and (2.3) are direct consequences of the coercivity bound on the double-integrand f in (H1$_p$). In fact, for (2.2), we use the nonlocal Poincaré inequality in [7, Proposition 4.2], which also holds for $u \in L^2(\Omega)$ via a truncation argument. From the upper bound in (H1$_p$), we infer (2.4). □

The next result provides a characterization of the domain of $\mathcal{R}_p$.

Lemma 4.3 *For any $p \in [1, \infty)$ there holds*

$$Dom \, \mathcal{R}_p = \left\{ u \in L^p(\Omega) \cap L^2(\Omega) : [u]_{p,\beta} < \infty \right\}. \tag{2.5}$$

If, additionally, $\beta p < n$, then

$$Dom \, \mathcal{R}_p = L^p(\Omega) \cap L^2(\Omega).$$

If, instead, $\beta p > n$, then

$$Dom \, \mathcal{R}_p = W^{\beta - \frac{n}{p}, p}(\Omega) \cap L^2(\Omega). \tag{2.6}$$

Proof By combining (2.2) and (2.3) with (2.4), we deduce (2.5). In the case $\beta p < n$, a direct computation shows that $[u]_{p,\beta} < \infty$ for all $u \in L^p(\Omega)$, hence we infer the statement. Property (2.6) follows by observing that for $\beta p > n$, the quantity $[u]_{p,\beta}$ corresponds to the Gagliardo semi-norm of the fractional Sobolev space $W^{\beta - \frac{n}{p}, p}(\Omega)$ (cf. e.g. [34]). □

As a consequence of Lemma 4.3, we deduce, in particular, that $C_c^\infty(\mathbb{R}^n) \subset Dom \, \mathcal{R}_p$, where the functions in $C_c^\infty(\mathbb{R}^n)$ are implicitly restricted to Ω.

The next lemma shows that any element of the domain of $\mathcal{R}_p$ can be extended to a function having compact support and finite (p, β)-seminorm.

Lemma 4.4 *Let $p \in [1, \infty)$. For any $u \in Dom \, \mathcal{R}_p$, there is a $\bar{u} \in L^p(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ with compact support inside some bounded open set Ω' with $\Omega \subset \Omega' \subset \mathbb{R}^n$ satisfying $\bar{u} = u$ on Ω and*

$$\int_{\Omega'} \int_{\Omega'} \frac{|\bar{u}(x) - \bar{u}(y)|^p}{|x - y|^{\beta p}} \, dx \, dy < \infty. \tag{2.7}$$

Proof If $\beta > \frac{n}{p}$, this follows directly from well-established extension results for fractional Sobolev spaces on Ω to those on $\mathbb{R}^n$ (cf. [34, Theorem 5.4]), considering (2.6). If $1 \le \beta p \le n$, the map $x \mapsto |x - y|^{-\beta p}$ is no longer integrable at infinity. Property (2.7) follows by minor modifications to the arguments in [34, Sect. 5]. □

Elements of the domain of $\mathcal{R}_p$ can be approximated by sequences of smooth maps with compact support.

Lemma 4.5 *Let $p \in [1, \infty)$. For every $u \in Dom \, \mathcal{R}_p$, there exists a sequence $(u_l)_l \subset C_c^\infty(\mathbb{R}^n)$ such that $u_l \to u$ in $L^p(\Omega)$ and $\lim_{l \to \infty} \mathcal{R}_p(u_l) = \mathcal{R}_p(u)$ as $l \to \infty$.*

Proof Let $\bar{u}$ be an extension of u as in Lemma 4.3. We define $u_l = \varphi_{1/l} * \bar{u} \in C_c^\infty(\mathbb{R}^n)$ for $l \in \mathbb{N}$ with $(\varphi_\varepsilon)_{\varepsilon > 0}$ a family of smooth standard mollifiers satisfying $0 \le \varphi_\varepsilon \le 1$ and $\int_{\mathbb{R}^n} \varphi_\varepsilon \, dx = 1$, and whose support lies in the ball centered at the origin and with radius $\varepsilon > 0$, supp $\varphi_\varepsilon \subset B_\varepsilon(0) \subset \mathbb{R}^n$. Then, $u_l \to u$ in $L^p(\Omega)$ and $u_l \to u$ pointwise a.e. in Ω as $l \to \infty$. To show that Lebesgue's dominated convergence theorem can

be applied, we use the upper bound in (H1$_p$) to derive the following estimate for any $l \in \mathbb{N}$:

$$f^p(x, y, u_l(x), u_l(y)) \le 4^{p-1} M^p \left(\frac{|u_l(x) - u_l(y)|^p}{|x - y|^{\beta p}} + |u_l(x)|^p + |u_l(y)|^p + 1 \right)$$
(2.8)

for a.e. $(x, y) \in \Omega \times \Omega$. By Jensen's inequality and Fubini's theorem,

$$
\begin{aligned}
[u_l]_{p,\beta}^p &\le \int_{B_{1/l}(0)} \varphi_{1/l}(z) \int_\Omega \int_\Omega \frac{|\bar{u}(x - z) - \bar{u}(y - z)|^p}{|x - y|^{\beta p}} \, dx \, dy \, dz \\
&\le \int_{\Omega_{1/l}} \int_{\Omega_{1/l}} \frac{|\bar{u}(x) - \bar{u}(y)|^p}{|x - y|^{\beta p}} \, dx \, dy < \infty,
\end{aligned}
$$

with $\Omega_{1/l} = \{x \in \mathbb{R}^n : d(x, \Omega) < 1/l\}$; thus, $\limsup_{l \to \infty} [u_l]_{p,\beta}^p \le [u]_{p,\beta}^p$. Conversely, the a.e. pointwise convergence of the mollified sequence gives $\liminf_{l \to \infty} [u_l]_{p,\beta}^p \ge [u]_{p,\beta}^p$ by Fatou's lemma. Along with the L^p-convergence of $(u_l)_l$, the upper bound in (2.8) is thus a converging sequence in $L^1(\Omega \times \Omega)$. This concludes the proof of the lemma. □

Finally, we characterize the weak lower-semicontinuity of the regularizers. We refer to [8, 36, 48] for a discussion on sufficient (and necessary) conditions for the weak lower semicontinuity of inhomogeneous double-integral functionals.

Lemma 4.6 *For every $p \in [1, \infty)$, the regularizer $\mathcal{R}_p$ is L^2-weak lower semicontinuous.*

Proof The statement is an immediate consequence of the nonnegativity of f and (H2$_p$), see e.g. [49, Theorem 2.5] or [45]. □

Remark 4.7 Observe that Lemmas 4.3 and 4.6 imply in particular that the hypothesis (H) from Sect. 2 is fulfilled.

Given a collection of noisy images $u^\eta \in L^2(\Omega; \mathbb{R}^N)$ and $p \in [1, \infty)$, we set, for each $j \in \{1, \ldots, N\}$,

$$\mathcal{J}_{p,j}(u) := \|u - u_j^\eta\|_{L^2(\Omega)}^2 + \mathcal{R}_p(u) \quad \text{for } u \in L^2(\Omega),$$

with $K_{p,j} := \arg \min \mathcal{J}_{p,j} \neq \emptyset$ since (H) is satisfied. As in ($\mathcal{I}$), we define $\mathcal{I} : [1, \infty) \to [0, \infty)$ by

$$\mathcal{I}(p) = \inf_{w \in K_p} \|w - u^c\|_{L^2(\Omega; \mathbb{R}^N)}^2 \quad \text{for } p \in [1, \infty),$$

where $K_p = K_{p,1} \times K_{p,2} \times \cdots \times K_{p,N}$. Next, we prove the Mosco-convergence result that will provide us with an extension of $\mathcal{I}$ to $\overline{\Lambda} = [1, \infty]$. It is an L^p-approximation statement in the present nonlocal setting, which can be obtained from a modification

of the arguments by Champion, De Pascale, & Prinari [20] in the local case, and those by Kreisbeck, Ritorto, & Zappale [42, Theorem 1.3], where the case of homogeneous double-integrands is studied.

Proposition 4.8 (Mosco-convergence of the regularizers) *Let* $\Lambda = [1, \infty)$, $\mathcal{R}_p$ *for* $p \in [1, \infty)$ *as in* (2.1), *and suppose that* (H1$_p$) *and* (H2$_p$) *are satisfied. Then, for* $p \in \overline{\Lambda} = [1, \infty]$,

$$\overline{\mathcal{R}}_p := \text{Mosc}(L^2)\text{-} \lim_{p' \to p} \mathcal{R}_{p'} = \begin{cases} \mathcal{R}_p & \text{if } p \in [1, \infty), \\ \mathcal{R}_\infty & \text{if } p = \infty, \end{cases} \tag{2.9}$$

with $\mathcal{R}_\infty : L^2(\Omega) \to [0, \infty]$ *given by*

$$\mathcal{R}_\infty(u) := \text{ess sup}_{(x,y) \in \Omega \times \Omega} f(x, y, u(x), u(y)).$$

Proof To show (2.9), it suffices to show that for every sequence $(p_k)_k \subset [1, \infty)$ converging to $p \in [1, \infty]$, (2.9) holds with p' replaced by p_k. We divide the proof into two cases.

Case 1: $p \in [1, \infty)$. For the recovery sequence, consider $u \in \text{Dom }\mathcal{R}_p$ and take $(u_l)_l \subset C_c^\infty(\mathbb{R})$ as in Lemma 4.5, satisfying $u_l \to u$ in $L^p(\Omega)$ and $\mathcal{R}_p(u_l) \to \mathcal{R}_p(u)$ as $l \to \infty$. In view of Lemma 4.3, we know that $(u_l)_l$ is contained in $\text{Dom }\mathcal{R}_p$ and $\text{Dom }\mathcal{R}_{p_k}$ for all $k \in \mathbb{N}$, and we conclude via Lebesgue's dominated convergence theorem that

$$\lim_{k \to \infty} \mathcal{R}_{p_k}(u_l) = \mathcal{R}_p(u_l)$$

for every $l \in \mathbb{N}$. Hence,

$$\lim_{l \to \infty} \lim_{k \to \infty} \mathcal{R}_{p_k}(u_l) = \lim_{l \to \infty} \mathcal{R}_p(u_l) = \mathcal{R}_p(u),$$

so that one can find a recovery sequence by extracting an appropriate diagonal sequence.

To prove the lower bound, let $u_k \rightharpoonup u$ in $L^2(\Omega)$ be such that $\lim_{k \to \infty} \mathcal{R}_{p_k}(u_k) = \liminf_{k \to \infty} \mathcal{R}_{p_k}(u_k) < \infty$, and fix $s \in (1, p)$ (or $s = 1$ if $p = 1$). Observe that $p_k \geq s$ for all k sufficiently large because $p_k \to p$ for $k \to \infty$. Then, Remark 4.1 b) and the weak lower semicontinuity of $\mathcal{R}_s$ according to Lemma 4.6 imply that

$$\lim_{k \to \infty} \mathcal{R}_{p_k}(u_k) \geq \liminf_{k \to \infty} \mathcal{R}_s(u_k) \geq \mathcal{R}_s(u).$$

If $s = p = 1$ the argument is complete, whereas in the case $p > 1$, an additional application of Fatou's lemma shows $\liminf_{s \nearrow p} \mathcal{R}_s(u) \geq \mathcal{R}_p(u)$, giving rise to the desired liminf inequality.

Case 2: $p = \infty$. That constant sequences serve as recovery sequences results from the observation that $\mathcal{R}_{p_k}(u) \to \mathcal{R}_\infty(u)$ as $k \to \infty$ for all $u \in \text{Dom }\mathcal{R}_\infty$. The latter is an immediate consequence of classical L^p-approximation, i.e., the well-known fact

that $\lim_{p \to \infty} \|v\|_{L^p(V)} = \|v\|_{L^\infty(V)} = \text{ess sup}_{x \in V} |v(x)|$ for all $v \in L^\infty(V)$ with $V \subset \mathbb{R}^m$ open and bounded.

To prove the lower bound, we argue via Young measure theory (see, e.g., [37, 48] for a general introduction). Let $u_k \rightharpoonup u$ in $L^2(\Omega)$, and denote by $\nu = \{\nu_x\}_{x \in \Omega}$ the Young measure generated by a (non-relabeled) subsequence of $(u_k)_k$. The barycenter of $[\nu_x] := \int_\mathbb{R} \xi \, d\nu_x(\xi)$ then coincides with $u(x)$ for a.e. $x \in \Omega$. Without loss of generality, one can suppose that $\infty > \liminf_{k \to \infty} \mathcal{R}_{p_k}(u_k) = \lim_{k \to \infty} \mathcal{R}_{p_k}(u_k)$. Recalling Remark 4.1 b), we have that

$$\lim_{k \to \infty} \mathcal{R}_{p_k}(u_k) \geq \liminf_{q \to \infty} \liminf_{k \to \infty} \mathcal{R}_q(u_k). \tag{2.10}$$

On the other hand, with the nonlocal field v_u associated with some $u : \Omega \to \mathbb{R}$ defined by

$$v_u(x, y) := (u(x), u(y)) \qquad \text{for } (x, y) \in \Omega \times \Omega,$$

the statement of [48, Proposition 2.3] allows us to extract a subsequence $(v_{u_k})_k$ that generates the Young measure $\{\nu_x \otimes \nu_y\}_{(x,y) \in \Omega \times \Omega}$. Hence, a standard result on Young measure lower semicontinuity (see e.g. [37, Sect. 8.1]) yields

$$\liminf_{k \to \infty} \mathcal{R}_q(u_k) \geq \left(\frac{1}{|\Omega \times \Omega|} \int_\Omega \int_\Omega \int_\mathbb{R} \int_\mathbb{R} f^q(x, y, \xi, \zeta) \, d\nu_x(\xi) \, d\nu_y(\zeta) \, dx \, dy \right)^{1/q}.$$

Letting $q \to \infty$, we use classical L^p-approximation results and the Jensen's type inequality for separately convex functions in [43, Lemma 3.5 (iv)] to conclude that

$$\liminf_{q \to \infty} \liminf_{k \to \infty} \mathcal{R}_q(u_k) \geq \text{ess sup}_{(x,y) \in \Omega \times \Omega} \left[(\nu_x \otimes \nu_y)\text{-} \text{ess sup}_{(\xi,\zeta) \in \mathbb{R} \times \mathbb{R}} f(x, y, \xi, \zeta) \right]$$

$$\geq \text{ess sup}_{(x,y) \in \Omega \times \Omega} f(x, y, [\nu_x], [\nu_y])$$

$$= \text{ess sup}_{(x,y) \in \Omega \times \Omega} f(x, y, u(x), u(y)) = \mathcal{R}_\infty(u);$$

note that $(\nu_x \otimes \nu_y)\text{-ess sup}_{(\xi,\zeta) \in \mathbb{R} \times \mathbb{R}} f(x, y, \xi, \zeta) = \inf\{c \in \mathbb{R} : f(x, y, \cdot, \cdot) \leq c \ (\nu_x \otimes \nu_y)\text{-a.e. in } \mathbb{R} \times \mathbb{R}\}$. Finally, the lower bound follows from the previous estimate and (2.10). $\qquad \square$

The above result implies that the reconstruction functional for $p = \infty$ and $j \in \{1, \ldots, N\}$ is given by

$$\overline{\mathcal{J}}_{\infty,j}(u) := \|u - u_j^\eta\|_{L^2(\Omega)}^2 + \mathcal{R}_\infty(u) \qquad \text{for } u \in L^2(\Omega).$$

Under the additional convexity condition on the given function $f : \Omega \times \Omega \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ that

(H3$_p$) f is (jointly) level convex in its last two variables,

where level convexity means convexity of the sub-level sets of the function, the supremal functional $\mathcal{R}_\infty$ also becomes level convex. In combination with the strict convexity of the fidelity term, the reconstruction functional $\overline{\mathcal{J}}_{\infty,j}$ then admits a unique minimizer. Since level convexity is weaker than convexity, we do not necessarily have that $\mathcal{J}_{p,j}$ for $p \in [1, \infty)$ is (level) convex, and it may have multiple minimizers.

If we suppose that f fulfills (H1$_p$)–(H3$_p$), then Theorem 2.5 and Proposition 4.8 imply that the extension $\overline{\mathcal{I}} : [1, \infty] \to [0, \infty]$ is given by

$$\overline{\mathcal{I}}(p) = \begin{cases} \mathcal{I}(p) & \text{for } p \in [1, \infty), \\ \|w^{(\infty)} - u^c\|^2_{L^2(\Omega;\mathbb{R}^N)} & \text{for } p = \infty, \end{cases}$$

for $p \in [1, \infty]$, where $w^{(\infty)}$ denotes the unique componentwise minimizer of $\overline{\mathcal{J}}_\infty$. In particular, the hypothesis (ii) of Theorem 2.5 is satisfied, which shows that $\overline{\mathcal{I}}$ is the relaxation of $\mathcal{I}$ and, thus, admits a minimizer $\bar{p} \in \overline{\Lambda} = [1, \infty]$.

We conclude this section with a discussion of examples when optimal values of the integrability exponents are obtained in the interior of the original interval Λ or at its boundary, respectively. In one case, the presence of noise causes $\mathcal{R}_\infty$ to penalize u^c more than u^η, while $\mathcal{R}_q$ for some $q \in [1, \infty)$ prefers the clean image. This entails that the optimal parameter is attained in $\Lambda = [1, \infty)$. In the second case instead, the reconstruction functional for $p = \infty$ gives back the exact clean image and outperforms the reconstruction functionals for other parameter values.

Example 4.9 a) Let $f = \alpha \widehat{f} : \Omega \times \Omega \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, for some $\alpha > 0$ to be specified later, be a double-integrand satisfying (H1$_p$), (jointly) convex in the last two variables, and vanishing exactly on $\{(x, y, \xi, \xi) : x, y \in \Omega, \ \xi \in \mathbb{R}\}$. Following (2.1), we set

$$\mathcal{R}_p(u) = \alpha \left(\frac{1}{|\Omega \times \Omega|} \int_\Omega \int_\Omega \widehat{f}^p(x, y, u(x), u(y)) \, dx \, dy \right)^{1/p} =: \alpha \widehat{\mathcal{R}}_p(u)$$

for $u \in L^2(\Omega)$ and $p \in [1, \infty)$.

We further introduce the following two conditions on the given data $u^\eta, u^c \in L^2(\Omega;\mathbb{R}^N)$:

(H4$_p$) $\sum_{j=1}^N \mathcal{R}_q(u^c_j) < \sum_{j=1}^N \mathcal{R}_q(u^\eta_j)$ for some $q \in [1, \infty)$;

(H5$_p$) $\sum_{j=1}^N \mathcal{R}_\infty(2u^\eta_j - u^c_j) < \sum_{j=1}^N \mathcal{R}_\infty(u^\eta_j)$.

By applying Lemma 3.4 (i) from the previous section with $\mathcal{R} = \widehat{\mathcal{R}}_q$ — the conditions (H1$_\alpha$), (H2$_\alpha$), and (H3$_\alpha$) are immediate to verify in view of Lemma 4.3, Lemma 4.6, and (H4$_p$) — we can then deduce for small enough α that $\overline{\mathcal{I}}(q) < \|u^\eta - u^c\|^2_{L^2(\Omega;\mathbb{R}^N)}$. On the other hand, due to (H5$_p$), the same lemma can be applied to $\mathcal{R} = \widehat{\mathcal{R}}_\infty$ with $\widehat{\mathcal{R}}_\infty(u) = \text{ess sup}_{(x,y)\in\Omega\times\Omega} \widehat{f}(x, y, u(x), u(y))$ for $u \in L^2(\Omega)$ to find

$$\|w^{(\infty)} - (2u^\eta - u^c)\|^2_{L^2(\Omega;\mathbb{R}^N)} < \|u^\eta - u^c\|^2_{L^2(\Omega;\mathbb{R}^N)}, \tag{2.11}$$

provided α is sufficiently small. The reverse triangle inequality then yields

$$\overline{\mathcal{I}}(\infty) \geq \left(\|w^{(\infty)} - (2u^\eta - u^c)\|^2_{L^2(\Omega;\mathbb{R}^N)} - 2\|u^\eta - u^c\|^2_{L^2(\Omega;\mathbb{R}^N)} \right)^2$$
$$> \|w^{(\infty)} - (2u^\eta - u^c)\|^2_{L^2(\Omega;\mathbb{R}^N)} > \|u^\eta - u^c\|^2_{L^2(\Omega;\mathbb{R}^N)} > \overline{\mathcal{I}}(q),$$

where in the second and third inequality we have used (2.11). This proves that the optimal parameter is attained inside $[1, \infty)$, and, therefore, is also a minimizer of $\mathcal{I}$.

b) We illustrate a) with a specific example. Consider $\Omega = (0, 1)$ and let $\widehat{f}(x, y, \xi, \zeta) = |\xi - \zeta|/|x - y|$ for $x, y \in \Omega$ and $\xi, \zeta \in \mathbb{R}^n$. This leads then to the difference quotient regularizers

$$\mathcal{R}_p(u) = \alpha \left(\int_0^1 \int_0^1 \frac{|u(x) - u(y)|^p}{|x - y|^p} \, dx \, dy \right)^{1/p} =: \alpha \widehat{\mathcal{R}}_p(u) \tag{2.12}$$

and

$$\mathcal{R}_\infty(u) = \alpha \operatorname{ess\,sup}_{(x,y)\in(0,1)^2} \frac{|u(x) - u(y)|}{|x - y|} = \alpha \operatorname{Lip}(u), \tag{2.13}$$

with $\operatorname{Lip}(u)$ denoting the Lipschitz constant of (a representative of) u, which could be infinite.

With the sawtooth function $v : [0, 1] \to \mathbb{R}$ defined by

$$v(x) = \begin{cases} x & \text{for } 0 \leq x \leq 1/4, \\ -x + 1/2 & \text{for } 1/4 < x \leq 3/4, \\ x - 1 & \text{for } 3/4 < x \leq 1, \end{cases}$$

we take a single clean and noisy image given by

$$u^c(x) = \begin{cases} 0 & \text{for } 0 < x \leq 1/3, \\ 10v(3x - 1) & \text{for } 1/3 < x \leq 2/3 \\ 0 & \text{for } 2/3 < x < 1. \end{cases}$$

$$\text{and} \quad u^\eta(x) = \begin{cases} v(3x) & \text{for } 0 < x \leq 1/3, \\ (10 - \varepsilon)v(3x - 1) & \text{for } 1/3 < x \leq 2/3, \\ v(3x - 2) & \text{for } 2/3 < x < 1, \end{cases}$$

respectively, where $\varepsilon > 0$ is small; see Fig. 1.

We observe that u^c is constant near the boundaries and only slightly steeper than u^η in the middle of the domain. Numerical calculations show that for small ε, such as $\varepsilon = 0.1$, the estimate $\mathcal{R}_2(u^c) < \mathcal{R}_2(u^\eta)$, and hence (H4$_p$) with $q = 2$, holds; moreover, (H5$_p$) holds since the clean image has a higher Lipschitz constant than the noisy image in the sense that

$$\operatorname{Lip}(2u^\eta - u^c) = 30 - 6\varepsilon < 30 - 3\varepsilon = \operatorname{Lip}(u^\eta).$$
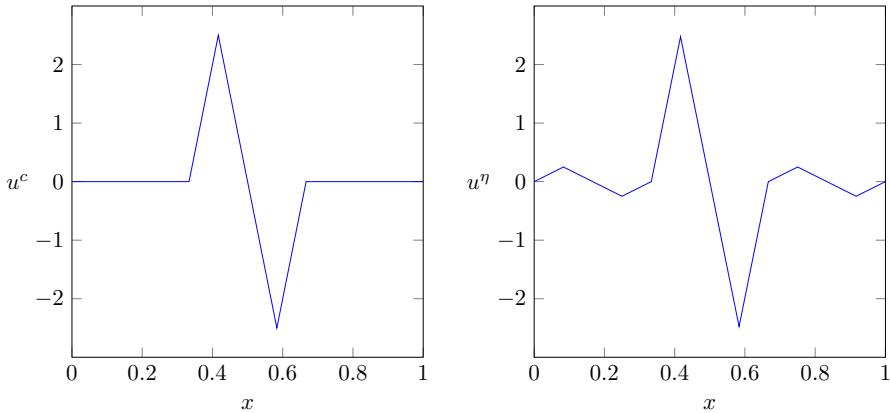
Fig. 1 The graphs of the functions u^c and u^η from Example 4.9 a) with $\varepsilon = 0.1$

Therefore, we find that for $\alpha > 0$ small enough, the optimal parameter lies inside $\Lambda = [1, \infty)$.

c) If we work with the same regularizers as in b), there are reasonable images for which the Lipschitz regularizer in (2.13) performs better than the other regularizers in (2.12). Let us consider with $\alpha > 0$ chosen as in b), the images

$$u^c(x) = x - 1/2 \quad \text{and} \quad u^\eta = (1 + 6\alpha)u^c.$$

Since u^η is affine, we can show that the reconstruction with the Lipschitz regularizer is also an affine function. Indeed, for every other function, one can find an affine function with at most the same Lipschitz constant without increasing the distance to u^η anywhere. This, in combination with the fact that the images are odd functions with respect to $x = 1/2$, shows that $w^{(\infty)}$ is of the form $w^{(\infty)}(x) = \gamma(x - 1/2) = \gamma u^c$ with $\gamma \geq 0$. Due to the optimality of $w^{(\infty)}$, the constant γ has to minimize the quantity

$$\|\gamma u^c - u^\eta\|^2_{L^2((0,1))} + \alpha \mathrm{Lip}(\gamma u^c) = \frac{1}{12}(\gamma - (1 + 6\alpha))^2 + \alpha\gamma,$$

which yields $\gamma = 1$. Hence, $w^{(\infty)}$ coincides with the clean image and therefore $\overline{\mathcal{I}}(\infty) = 0$, which implies that $p = \infty$ is the optimal parameter in this case.

5 Varying the Amount of Nonlocality

Next, we study two classes of nonlocal regularizers, $\mathcal{R}_\delta$ with $\delta \in \Lambda := (0, \infty)$, considered by Brezis & Nguyen [15] and Aubert & Kornprobst [5], respectively, in the context of image processing. In both cases, we aim at optimizing the parameter δ that encodes the amount of nonlocality in the problem. We mention further that both families of functionals recover the classical TV-reconstruction model in the limit $\delta \to 0$, cf. [5, 15].

To set the stage for our analysis, consider training data $(u^c, u^\eta) \in L^2(\Omega; \mathbb{R}^N) \times L^2(\Omega; \mathbb{R}^N)$ and the reconstruction functionals $\mathcal{J}_{\delta, j} : L^2(\Omega) \to [0, \infty]$ with $\delta \in \Lambda$ and $j \in \{1, 2, \ldots, N\}$ given by

$$\mathcal{J}_{\delta, j}(u) = \|u - u_j^\eta\|_{L^2(\Omega)}^2 + \mathcal{R}_\delta(u).$$

After showing that the sets

$$K_{\delta, j} = \underset{u \in L^2(\Omega)}{\arg \min} \, \mathcal{J}_{\delta, j}(u). \tag{2.1}$$

are non-empty for each of the two choices of the regularizers $\mathcal{R}_\delta$, the upper-level functional from $(\mathcal{T})$ in Sect. 2 becomes

$$\mathcal{I} : (0, \infty) \to [0, \infty), \quad \mathcal{I}(\delta) = \inf_{w \in K_\delta} \|w - u^c\|_{L^2(\Omega; \mathbb{R}^N)}^2 \tag{2.2}$$

with $K_\delta = K_{\delta, 1} \times K_{\delta, 2} \times \cdots \times K_{\delta, N}$. In order to find its extension $\overline{\mathcal{I}}$ defined on $\overline{\Lambda} = [0, \infty]$, we determine the Mosco-limits of the regularizers (cf. (2.6) and Theorem 2.5). This is the content of Propositions 5.3 and 5.5 below, which provide the main results of this section.

5.1 Brezis & Nguyen Setting

For every $\delta \in (0, \infty)$ and $u \in L^1(\Omega)$, we consider the regularizers

$$\mathcal{R}_\delta(u) := \delta \int_\Omega \int_\Omega \frac{\varphi(|u(x) - u(y)|/\delta)}{|x - y|^{n+1}} \, dx \, dy,$$

where, following [15], the function $\varphi : [0, \infty) \to [0, \infty)$ is assumed to satisfy the following hypotheses:

(H1$_\delta$) φ is lower semicontinuous in $[0, \infty)$ and continuous in $[0, \infty)$ except at a finite number of points, where it admits left- and right-side limits;

(H2$_\delta$) there exists a constant $a > 0$ such that $\varphi(t) \leq \min\{at^2, a\}$ for all $t \in [0, \infty)$;

(H3$_\delta$) φ is non-decreasing;

(H4$_\delta$) it holds that $\gamma_n \int_0^\infty \varphi(t) t^{-2} \, dt = 1$ with $\gamma_n := \int_{\mathbb{S}^{n-1}} |e \cdot \sigma| \, d\sigma$ for any $e \in \mathbb{S}^{n-1}$.

Note that the assumptions on φ imply that the functional $\mathcal{R}_\delta$ is never convex.

Example 5.1 Examples of functions φ with the properties (H1$_\delta$)–(H4$_\delta$) include suitable normalizations of

$$t \mapsto \begin{cases} 0 & \text{if } t \leq 1 \\ 1 & \text{if } t > 1 \end{cases}, \qquad t \mapsto \begin{cases} t^2 & \text{if } t \leq 1 \\ 1 & \text{if } t > 1 \end{cases}, \qquad t \mapsto 1 - e^{-t^2}$$

for $t \geq 0$, cf. [15].

To guarantee that the functionals $\mathcal{R}_\delta$ satisfy a suitable compactness property, see Theorem 5.2 b), we must additionally assume that

(H5$_\delta$) $\varphi(t) > 0$ for all $t > 0$.

Clearly, the last two functions from Example 5.1 satisfy the positivity condition, while the first one does not. In identifying the Mosco-limits $\overline{\mathcal{R}}_\delta$ in each of the three cases $\delta \in (0, \infty)$, $\delta = 0$, and $\delta = \infty$, we make repeated use of [15, Theorems 1, 2 and 3], which we recall here for the reader's convenience.

Theorem 5.2 (cf. [15, Theorems 1–3]) *Let $\Omega \subset \mathbb{R}^n$ be a bounded and smooth domain, and let φ satisfy* (H1$_\delta$)–(H4$_\delta$).
(a) *If $(\delta_k)_k \subset (0, \infty)$ is such that $\delta_k \to 0$, then the following statements hold:*

(i) *There exists a constant $K(\varphi) \in (0, 1]$, independent of Ω, such that $(\mathcal{R}_{\delta_k})_k$ Γ-converges as $k \to \infty$, with respect to the $L^1(\Omega)$-topology, to $\mathcal{R}_0 : L^1(\Omega) \to [0, \infty]$ defined for $u \in L^1(\Omega)$ by*

$$\mathcal{R}_0(u) := \begin{cases} K(\varphi)|Du|(\Omega) & \text{if } u \in BV(\Omega), \\ \infty & \text{if } u \in L^1(\Omega) \setminus BV(\Omega). \end{cases}$$

(ii) *If $(u_k)_k$ is a bounded sequence in $L^1(\Omega)$ with $\sup_k \mathcal{R}_{\delta_k}(u_k) < \infty$, then there exist a subsequence $(u_{k_l})_l$ of $(u_k)_k$ and a function $u \in L^1(\Omega)$ such that $\lim_{l\to\infty} \|u_{k_l} - u\|_{L^1(\Omega)} = 0$.*

(b) *Suppose that* (H5$_\delta$) *holds in addition to the above conditions, and let $(u_k)_k$ be a bounded sequence in $L^1(\Omega)$ with $\sup_k \mathcal{R}_\delta(u_k) < \infty$ for some $\delta > 0$. Then, there exists a subsequence $(u_{k_l})_l$ of $(u_k)_k$ and a function $u \in L^1(\Omega)$ such that $\lim_{l\to\infty} \|u_{k_l} - u\|_{L^1(\Omega)} = 0$.*

We point out that if φ fulfills (H1$_\delta$)–(H5$_\delta$), then (H) in Sect. 2 holds and the sets $K_{\delta,j}$ defined in (2.1) are non-empty (cf. [15, Corollary 7]). We are now in a position to characterize the asymptotic behavior of the regularizers $\mathcal{R}_{\delta'}$ as $\delta' \to \delta \in \overline{\Lambda} = [0, \infty]$.

Proposition 5.3 (Mosco-convergence of regularizers) *Let $\Lambda = (0, \infty)$ and $\Omega \subset \mathbb{R}^n$ be a bounded and smooth domain. Under the assumptions* (H1$_\delta$)–(H5$_\delta$) *on $\varphi : [0, \infty) \to [0, \infty)$, it holds that*

$$\overline{\mathcal{R}}_\delta := \text{Mosc}(L^2)\text{-}\lim_{\delta'\to\delta} \mathcal{R}_{\delta'} = \begin{cases} \mathcal{R}_\delta & \text{if } \delta \in (0, \infty), \\ \mathcal{R}_0 & \text{if } \delta = 0, \quad \text{for } \delta \in \overline{\Lambda} = [0, \infty]. \\ 0 & \text{if } \delta = \infty, \end{cases} \quad (2.3)$$

Proof Considering a sequence $(\delta_k)_k \subset (0, \infty)$ with limit $\delta \in [0, \infty]$, one needs to verify that the Mosco-limit of $(\mathcal{R}_{\delta_k})_k$ exist and is given by the right-hand side of (2.3). We split the proof into three cases.

Case 1: $\delta = 0$. Let $(u_k)_k \subset L^2(\Omega)$ and $u \in L^2(\Omega)$ be such that $u_k \rightharpoonup u$ in $L^2(\Omega)$. We aim to show that

$$\mathcal{R}_0(u) \le \liminf_{k\to\infty} \mathcal{R}_{\delta_k}(u). \quad (2.4)$$

One may thus assume without loss of generality that the limit inferior on the right-hand side of (2.4) is finite, and, after extracting a subsequence if necessary, also

$$\sup_k \mathcal{R}_{\delta_k}(u_k) < \infty.$$

Hence, by Theorem 5.2 a) (ii), it follows that $u_k \to u$ in $L^1(\Omega)$, which together with Theorem 5.2 a) (i) yields (2.4).

To complement this lower bound, we need to obtain for each $u \in L^2(\Omega) \cap BV(\Omega)$ a sequence $(u_k)_k \subset L^2(\Omega)$ such that $u_k \to u$ in $L^2(\Omega)$ and

$$\mathcal{R}_0(u) \geq \limsup_{k \to \infty} \mathcal{R}_{\delta_k}(u_k). \tag{2.5}$$

The idea is to suitably truncate a recovery sequence of the Γ-limit $\Gamma(L^1)$-$\lim_{k \to \infty} \mathcal{R}_{\delta_k}$ from Theorem 5.2 (i). For the details, fix $l \in \mathbb{N}$ and consider the truncation function, $T^l : \mathbb{R} \to \mathbb{R}$,

$$T^l(t) := \begin{cases} l & \text{if } t \geq l, \\ t & \text{if } -l \leq t \leq l, \\ -l & \text{if } t \leq -l. \end{cases}$$

By Theorem 5.2 (i), there exists a sequence $(v_k)_k \subset L^1(\Omega)$ such that $v_k \to u$ in $L^1(\Omega)$ and

$$\lim_{k \to \infty} \mathcal{R}_{\delta_k}(v_k) = K(\varphi)|Du|(\Omega) = \mathcal{R}_0(u). \tag{2.6}$$

Choosing a sequence $(l_k)_k \subset \mathbb{R}$ such that $l_k \to \infty$ and $l_k \|v_k - u\|_{L^1(\Omega)} \to 0$ as $k \to \infty$, we define

$$u_k := T^{l_k} \circ v_k \in L^\infty(\Omega) \quad \text{for all } k \in \mathbb{N}.$$

Then, an application of Hölder's inequality shows that

$$\begin{aligned} \|u_k - u\|_{L^2(\Omega)} &\leq \|u_k - T^{l_k} \circ u\|_{L^2(\Omega)} + \|T^{l_k} \circ u - u\|_{L^2(\Omega)} \\ &\leq \left(2l_k \|v_k - u\|_{L^1(\Omega)}\right)^{1/2} + \|T^{l_k} \circ u - u\|_{L^2(\Omega)} \to 0, \end{aligned}$$

as $k \to \infty$. Therefore, $u_k \to u$ in $L^2(\Omega)$ and, in view of the monotonicity of φ in (H3$_\delta$), we conclude that

$$\begin{aligned} \limsup_{k \to \infty} \mathcal{R}_{\delta_k}(u_k) &= \limsup_{k \to \infty} \delta_k \int_\Omega \int_\Omega \frac{\varphi(\delta_k^{-1}|T^{l_k}(v_k(x)) - T^{l_k}(v_k(y))|)}{|x - y|^{n+1}} \, dx \, dy \\ &\leq \lim_{k \to \infty} \delta_k \int_\Omega \int_\Omega \frac{\varphi(\delta_k^{-1}|v_k(x) - v_k(y)|)}{|x - y|^{n+1}} \, dx \, dy = \lim_{k \to \infty} \mathcal{R}_{\delta_k}(v_k), \end{aligned}$$

which implies (2.5) by (2.6).

Case 2: $\delta \in (0, \infty)$. Consider a sequence $(u_k)_k \subset L^2(\Omega)$ and $u \in L^2(\Omega)$ such that $u_k \rightharpoonup u$ in $L^2(\Omega)$ and

$$\sup_k \mathcal{R}_{\delta_k}(u_k) < \infty.$$

We start by observing that there exist $\bar{\delta} > 0$ and $K \in \mathbb{N}$ such that for all $k \geq K$, we have $\bar{\delta}/2 \leq \delta_k \leq \bar{\delta}$. Hence, the previous estimate and (H3$_\delta$) yield

$$\sup_{k \geq K} \mathcal{R}_{\bar{\delta}}(u_k) = \sup_{k \geq K} \left(\bar{\delta} \int_\Omega \int_\Omega \frac{\varphi(\bar{\delta}^{-1}|u_k(x) - u_k(y)|)}{|x - y|^{n+1}} \, dx \, dy \right) \leq 2 \sup_k \mathcal{R}_{\delta_k}(u_k) < \infty.$$

Consequently, in view of Theorem 5.2 b), we may further assume that

$$u_k \to u \text{ in } L^1(\Omega) \qquad \text{and} \qquad u_k(x) \to u(x) \text{ for a.e. } x \in \Omega. \tag{2.7}$$

Using Fatou's lemma first, and then (2.7) together with the lower semicontinuity of φ on $[0, \infty)$, we get

$$
\begin{aligned}
\liminf_{k \to \infty} \mathcal{R}_{\delta_k}(u_k) &= \liminf_{k \to \infty} \delta_k \int_\Omega \int_\Omega \frac{\varphi(\delta_k^{-1}|u_k(x) - u_k(y)|)}{|x - y|^{n+1}} \, dx \, dy \\
&\geq \delta \int_\Omega \int_\Omega \liminf_{k \to \infty} \frac{\varphi(\delta_k^{-1}|u_k(x) - u_k(y)|)}{|x - y|^{n+1}} \, dx \, dy \\
&\geq \delta \int_\Omega \int_\Omega \frac{\varphi(\delta^{-1}|u(x) - u(y)|)}{|x - y|^{n+1}} \, dx \, dy = \mathcal{R}_\delta(u),
\end{aligned}
$$

which proves the liminf inequality.

For the recovery sequence, fix $u \in L^2(\Omega)$ and take $u_k = \frac{\delta_k}{\delta} u$ for $k \in \mathbb{N}$. Then, $u_k \to u$ in $L^2(\Omega)$ as $k \to \infty$ and

$$\lim_{k \to \infty} \mathcal{R}_{\delta_k}(u_k) = \lim_{k \to \infty} \frac{\delta_k}{\delta} \mathcal{R}_\delta(u) = \mathcal{R}_\delta(u),$$

as desired.

Case 3: $\delta = \infty$. The lower bound follows immediately by the non-negativity of $\mathcal{R}_{\delta_k}$ for $k \in \mathbb{N}$. As a recovery sequence for $u \in L^2(\Omega)$, take a sequence $(u_k)_k \subset L^2(\Omega)$ such that $u_k \to u$ in $L^2(\Omega)$ and $\text{Lip}(u_k) \leq \delta_k^{1/4}$, which is possible since $\delta_k \to \infty$ as $k \to \infty$. Then, using (H2$_\delta$),

$$
\begin{aligned}
\mathcal{R}_{\delta_k}(u_k) &= \delta_k \int_\Omega \int_\Omega \frac{\varphi(\delta_k^{-1}|u_k(x) - u_k(y)|)}{|x - y|^{n+1}} \, dx \, dy \\
&\leq a \frac{\text{Lip}(u_k)^2}{\delta_k} \int_\Omega \int_\Omega \frac{1}{|x - y|^{n-1}} \, dx \, dy \leq a \delta_k^{-1/2} \int_\Omega \int_\Omega \frac{1}{|x - y|^{n-1}} \, dx \, dy.
\end{aligned}
$$

Hence, $\mathcal{R}_{\delta_k}(u_k) \to 0$ as $k \to \infty$, which concludes the proof.   □

5.2 Aubert & Kornprobst Setting

Let $\Omega \subset \mathbb{R}^n$ be a bounded Lipschitz domain. We fix a nonnegative function $\rho : [0, \infty) \to [0, \infty)$ satisfying

(H6$_\delta$) ρ is non-increasing and $\displaystyle\int_{\mathbb{R}^n} \rho(|x|) \, dx = 1$,

and consider the regularizers given for $\delta \in \Lambda = (0, \infty)$ and $u \in L^2(\Omega)$ by

$$\mathcal{R}_\delta(u) = \frac{1}{\delta^n} \int_\Omega \int_\Omega \frac{|u(x) - u(y)|}{|x - y|} \rho\left(\frac{|x - y|}{\delta}\right) dx \, dy. \qquad (2.8)$$

Remark 5.4 a) As ρ is non-increasing, we have for all $0 < \delta < \bar{\delta}$ and $x, y \in \Omega$ that $\rho(|x - y|/\delta) \le \rho(|x - y|/\bar{\delta})$; consequently,

$$\mathcal{R}_\delta(u) \le \frac{\bar{\delta}^n}{\delta^n} \mathcal{R}_{\bar{\delta}}(u)$$

for all $u \in L^2(\Omega)$.

b) Note that the assumption (H) from Sect. 2 is satisfied here; in particular, $\mathcal{R}_\delta$ is L^2-weakly lower semicontinuous. Indeed, as the dependence of the integrand on u is convex, it is enough to prove strong lower semicontinuity in $L^2(\Omega)$. This is in turn a simple consequence of Fatou's lemma.

c) In this set-up, the sets $K_{\delta, j}$ in (2.1) consist of a single element $w_j^{(\delta)} \in L^2(\Omega)$ in light of the strict convexity of the fidelity term and convexity of $\mathcal{R}_\delta$. The upper-level functional from (2.2) then becomes

$$\mathcal{I} : (0, \infty) \to [0, \infty), \quad \mathcal{I}(\delta) = \|w^{(\delta)} - u^c\|_{L^2(\Omega; \mathbb{R}^N)}^2.$$

The nonlocal functionals in (2.8) have been applied to problems in imaging in [5], providing a derivative-free alternative to popular local models. The localization behavior of these functionals as $\delta \to 0$ is well-studied, originally by Bourgain, Brezis, & Mironescu [12] and later extended to the BV-case in [26, 51]. Using these results, we show that, as $\delta \to 0$, the reconstruction functional in our bi-level scheme turns into the TV-reconstruction functional, see Proposition 5.5 below. Moreover, in order to get structural stability inside the domain Λ, we exploit the monotonicity properties of the functional $\mathcal{R}_\delta$, cf. Remark 5.4 a). Lastly, as $\delta \to \infty$, we observe that the regularization term vanishes.

Proposition 5.5 (Mosco-convergence of the regularizers) *Let $\Lambda = (0, \infty)$, $\Omega \subset \mathbb{R}^n$ be a bounded Lipschitz domain and assume that* (H6$_\delta$) *holds. Then,*

$$\overline{\mathcal{R}}_\delta := \text{Mosc}(L^2)\text{-} \lim_{\delta' \to \delta} \mathcal{R}_{\delta'} = \begin{cases} \mathcal{R}_\delta & \text{if } \delta \in (0, \infty), \\ \mathcal{R}_0 & \text{if } \delta = 0, \\ 0 & \text{if } \delta = \infty, \end{cases} \quad \text{for } \delta \in \overline{\Lambda} = [0, \infty], \ (2.9)$$

where

$$\mathcal{R}_0 : L^2(\Omega) \to [0, \infty], \quad \mathcal{R}_0(u) = \begin{cases} \kappa_n |Du|(\Omega) & \text{if } u \in BV(\Omega), \\ \infty & \text{if } u \in L^2(\Omega) \setminus BV(\Omega), \end{cases} \tag{2.10}$$

with $\kappa_n = \fint_{\mathbb{S}^{n-1}} |e \cdot \sigma| \, d\sigma$ *for any* $e \in \mathbb{S}^{n-1}$.

Proof Given $(\delta_k)_k \subset (0, \infty)$ with limit $\delta \in [0, \infty]$, the arguments below, subdivided into three different regimes, show that the Mosco-limit of $(\mathcal{R}_{\delta_k})_k$ exists and is equal to the right-hand side of (2.9).

Case 1: $\delta = 0$. For the lower bound, take a sequence $u_k \rightharpoonup u$ in $L^2(\Omega)$ and assume without loss of generality that

$$\sup_k \mathcal{R}_{\delta_k}(u_k) < \infty.$$

By [12, Theorem 4], $(u_k)_k$ is relatively compact in $L^1(\Omega)$, so that $u_k \to u$ in $L^1(\Omega)$. We now use the Γ-liminf result with respect to the $L^1(\Omega)$-convergence in [51, Corollary 8], to deduce that

$$\mathcal{R}_0(u) \leq \liminf_{k \to \infty} \mathcal{R}_{\delta_k}(u_k),$$

as desired. For the recovery sequence, we may suppose that $u \in L^2(\Omega) \cap BV(\Omega)$. Then, it follows from [51, Corollary 1] that

$$\lim_{k \to \infty} \frac{1}{\delta_k^n} \int_\Omega \int_\Omega \frac{|u(x) - u(y)|}{|x - y|} \rho \left(\frac{|x - y|}{\delta_k} \right) dx \, dy = \kappa_n |Du|(\Omega),$$

showing that the constant sequence $u_k = u$ for all $k \in \mathbb{N}$ provides a recovery sequence.

Case 2: $\delta \in (0, \infty)$. For the liminf inequality, take a sequence $(u_k)_k$ converging weakly to u in $L^2(\Omega)$. If $\bar{\delta} \in (0, \delta)$, then $\delta_k > \bar{\delta}$ for all $k \in \mathbb{N}$ large enough. Hence, it follows from Remark 5.4 a) that

$$\liminf_{k \to \infty} \mathcal{R}_{\delta_k}(u_k) \geq \liminf_{k \to \infty} \frac{\bar{\delta}^n}{\delta_k^n} \mathcal{R}_{\bar{\delta}}(u_k) \geq \frac{\bar{\delta}^n}{\delta^n} \mathcal{R}_{\bar{\delta}}(u),$$

where the last inequality uses the weak lower semicontinuity of $\mathcal{R}_{\bar{\delta}}$, cf. Remark 5.4 b). Letting $\bar{\delta} \nearrow \delta$ and using the monotone convergence theorem gives

$$\liminf_{k \to \infty} \mathcal{R}_{\delta_k}(u_k) \geq \mathcal{R}_\delta(u).$$

For the limsup inequality, consider $u \in L^2(\Omega)$ with $\mathcal{R}_\delta(u) < \infty$. Since ρ is non-increasing by (H6$_\delta$), we may extend u to a function $\bar{u} \in L^2(\mathbb{R}^n)$ by reflection across

the boundary of the Lipschitz domain Ω such that

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{|\bar{u}(x) - \bar{u}(y)|}{|x-y|} \rho\left(\frac{|x-y|}{\delta}\right) dx\, dy < \infty,$$

cf. [12, Proof of Theorem 4]. With $(\varphi_\varepsilon)_\varepsilon$ a family of smooth standard mollifiers, the sequence $u_l := \varphi_{1/l} * \bar{u}$ for $l \in \mathbb{N}$ converges to u in $L^2(\Omega)$ as $l \to \infty$, and we may argue similarly to the proof of Lemma 4.5 to conclude that

$$\lim_{l \to \infty} \mathcal{R}_\delta(u_l) = \mathcal{R}_\delta(u).$$

With $\rho_\delta := \delta^{-n} \rho(|\cdot|/\delta)$ and for a fixed $l \in \mathbb{N}$, we find that

$$|\mathcal{R}_\delta(u_l) - \mathcal{R}_{\delta_k}(u_l)| \leq \int_\Omega \int_\Omega \frac{|u_l(x) - u_l(y)|}{|x-y|} |\rho_\delta(x-y) - \rho_{\delta_k}(x-y)|\, dx\, dy$$
$$\leq \mathrm{Lip}(u_l) |\Omega| \|\rho_\delta - \rho_{\delta_k}\|_{L^1(\mathbb{R}^n)},$$

where $\mathrm{Lip}(u_l)$ is the Lipschitz constant of u_l. We have $\rho_{\delta_k} \to \rho_\delta$ in $L^1(\mathbb{R}^n)$ as $k \to \infty$ by a standard argument approximating ρ with smooth functions. Hence, we obtain

$$\lim_{k \to \infty} \mathcal{R}_{\delta_k}(u_l) = \mathcal{R}_\delta(u_l),$$

and, letting $l \to \infty$, results in

$$\lim_{l \to \infty} \lim_{k \to \infty} \mathcal{R}_{\delta_k}(u_l) = \mathcal{R}_\delta(u).$$

The limsup inequality now follows by extracting an appropriate diagonal sequence.

Case 3: $\delta = \infty$. The only nontrivial case is the limsup inequality, for which we take a sequence $(u_l)_l \subset C_c^\infty(\mathbb{R}^n)$ that converges to u in $L^2(\Omega)$. Then, with R larger than the diameter of Ω, one obtains for every $l \in \mathbb{N}$ that

$$\mathcal{R}_{\delta_k}(u_l) = \frac{1}{\delta_k^n} \int_\Omega \int_\Omega \frac{|u_l(x) - u_l(y)|}{|x-y|} \rho\left(\frac{|x-y|}{\delta_k}\right) dx\, dy$$
$$\leq \mathrm{Lip}(u_l) \int_\Omega \int_{\Omega/\delta_k} \rho\left(\left|z - \frac{y}{\delta_k}\right|\right) dz\, dy \leq \mathrm{Lip}(u_l) \int_\Omega \int_{B_{\frac{R}{\delta_k}}(0)} \rho(|w|)\, dw\, dy.$$

As $k \to \infty$, the last quantity goes to zero since $\rho(|\cdot|) \in L^1(\mathbb{R}^n)$. Therefore, we deduce that

$$\lim_{k \to \infty} \mathcal{R}_{\delta_k}(u_l) = 0,$$

and conclude again with a diagonal argument. □

5.3 Conclusions and Examples

In both the Brezis & Nguyen and the Aubert & Kornprobst settings, we now find that the extension $\overline{\mathcal{I}} : [0, \infty] \to [0, \infty]$ is given by

$$\overline{\mathcal{I}}(\delta) = \begin{cases} \mathcal{I}(\delta) & \text{if } \delta \in (0, \infty), \\ \|w^{(0)} - u^c\|^2_{L^2(\Omega;\mathbb{R}^N)} & \text{if } \delta = 0, \\ \|u^\eta - u^c\|^2_{L^2(\Omega;\mathbb{R}^N)} & \text{if } \delta = \infty, \end{cases}$$

where $w_j^{(0)}$ for $j \in \{1, \ldots, N\}$ is the unique minimizer of the TV-reconstruction functional $\overline{\mathcal{J}}_{0,j}$ (with different weight factors in the two cases). In particular, we deduce from Theorem 2.5 and Corollary 2.8 that $\overline{\mathcal{I}}$ is the relaxation of $\mathcal{I}$ and that these extended upper-level functionals $\overline{\mathcal{I}}$ admit minimizers $\overline{\delta} \in [0, \infty]$. To get an intuition about when this optimal parameter is attained at the boundary or in the interior of Λ, we present the following examples.

Example 5.6 a) For both settings analyzed in this section, it is clear that if the noisy and clean image coincide, $u^c \equiv u^\eta$, then the reconstruction model with parameter $\delta = \infty$ gives the exact clean image back. Hence, in this case the optimal parameter is attained at the boundary point $\delta = \infty$.

b) Next, we illustrate the case when the optimal parameter is attained at the boundary point $\delta = 0$. Consider the Aubert & Kornprobst setting in Sect. 5.2 and let $\Omega = (-1, 1)$, $N = 1$, $u^c = 0$, and $u^\eta(x) = \kappa_n x$ for $x \in (-1, 1)$. The reconstruction of u^η with the total variation regularizer $\mathcal{R}_0$ in (2.10) is of the form

$$w^{(0)} = \max\{\theta_1, \min\{\theta_2, u^\eta\}\} \quad \text{for some } \theta_1, \theta_2 \in \mathbb{R}.$$

To see this, we observe that $\overline{\mathcal{J}}_0(\tilde{u}) \le \overline{\mathcal{J}}_0(u)$ for any $u \in BV(-1, 1)$ with

$$\tilde{u} = \max\{u^-, \min\{u^+, u^\eta\}\},$$

where $u^- := \operatorname{ess\,inf}_{x \in (-1,1)} u(x)$ and $u^+ := \operatorname{ess\,sup}_{x \in (-1,1)} u(x)$. Indeed, the map $\tilde{u}$ has at most the same total variation as u and does not increase the distance to u^η anywhere. Next, since u^η is an odd function, the same should hold for the minimizer, meaning that $-\theta_1 = \theta_2 =: \theta \in [0, \kappa_n]$. We can now determine the value of θ by optimizing the quantity $\overline{\mathcal{J}}_0(w^{(0)})$ in θ. This boils down to minimizing

$$\frac{2}{3}\kappa_n^2 \left(1 - \frac{\theta}{\kappa_n}\right)^3 + 2\kappa_n\theta,$$

and yields $\theta = 0$. Hence, the reconstruction model for $\delta = 0$ yields the exact clean image, so that $\overline{\mathcal{I}}(0) = 0$. The same conclusions can be drawn for the Brezis & Nguyen setting by replacing κ_n in the example above with $K(\varphi)$.

c) Let us finally address the case when $\overline{\mathcal{I}}$ becomes minimal inside $\Lambda = (0, \infty)$. We work once again with the Aubert & Kornprobst model from Sect. 5.2, and assume

in addition to (H6$_\delta$) that the function ρ is equal to 1 in a neighborhood of zero. We consider the following conditions on the pair of data points $(u^c, u^\eta) \in L^2(\Omega; \mathbb{R}^N) \times L^2(\Omega; \mathbb{R}^N)$:

(H7$_\delta$) $\|u^\eta - u^c\|^2_{L^2(\Omega; \mathbb{R}^N)} < \|w^{(0)} - u^c\|^2_{L^2(\Omega; \mathbb{R}^N)}$;

(H8$_\delta$) $\sum_{j=1}^N \widetilde{\mathcal{R}}(u_j^c) < \sum_{j=1}^N \widetilde{\mathcal{R}}(u_j^\eta)$;

here, $w^{(0)}$ is the componentwise minimizer of the TV-reconstruction functional $\overline{\mathcal{J}}_0$ and we set

$$\widetilde{\mathcal{R}}(u) := \int_\Omega \int_\Omega \frac{|u(x) - u(y)|}{|x - y|} \, \mathrm{d}x \, \mathrm{d}y \quad \text{for } u \in L^2(\Omega). \tag{2.11}$$

The two hypotheses above can be realized, for example, by taking $u^\eta = (1 + \varepsilon)u^c$ for some small $\varepsilon > 0$ and $w^{(0)} \neq u^c$.

Notice that (H7$_\delta$) immediately rules out $\delta = 0$ as an optimal candidate, since the reconstruction at $\delta = \infty$ is better. On the other hand, ρ is supposed to be equal to 1 near the zero, so that we infer for large enough δ that

$$\mathcal{R}_\delta(u) = \frac{1}{\delta^n} \int_\Omega \int_\Omega \frac{|u(x) - u(y)|}{|x - y|} \, \mathrm{d}x \, \mathrm{d}y = \frac{1}{\delta^n} \widetilde{\mathcal{R}}(u) \tag{2.12}$$

for all $u \in L^2(\Omega)$. Since, for large δ, the dependence of the regularizer on δ is of the same type as the weight case from Sect. 3, we may apply Lemma 3.4 (i) in view of (H8$_\delta$). This yields, for all δ large enough, that

$$\|u^c - w^{(\delta)}\|^2_{L^2(\Omega; \mathbb{R}^N)} < \|u^c - u^\eta\|^2_{L^2(\Omega; \mathbb{R}^N)},$$

with $w^{(\delta)}$ the minimizer of $\mathcal{J}_\delta$. This shows that the optimal parameter is not attained at $\delta = \infty$ either and, as a result, needs to be attained inside $\Lambda = (0, \infty)$. Hence, the optimal regularizer lies within the class we started with.

The same conclusions can be drawn for the Brezis & Nguyen case described in Sect. 5.1 if we assume that $\varphi(t) = ct^r$ for small t with $c > 0$ and $r \geq 2$. One may take, for instance, the normalized version of the second function in Example 5.1. We then suppose that the pair of data points (u^c, u^η) satisfies (H7$_\delta$) and (H8$_\delta$), but now instead of (2.11), take

$$\widetilde{\mathcal{R}}(u) := c \int_\Omega \int_\Omega \frac{|u(x) - u(y)|^r}{|x - y|^{n+1}} \, \mathrm{d}x \, \mathrm{d}y \quad \text{for } u \in L^2(\Omega).$$

We observe with $l = \|u^\eta\|_{L^\infty(\Omega; \mathbb{R}^N)}$ (which we assume to be finite) and T^l the truncation as in the proof of Proposition 5.3 that

$$\mathcal{J}_\delta(T^l \circ u) \leq \mathcal{J}_\delta(u)$$

for all $u \in L^2(\Omega)$ and $\delta \in (0, \infty)$. Therefore, we may restrict our analysis to functions $u \in L^2(\Omega)$ with $|u(x) - u(y)| \leq 2l$ for all $x, y \in \Omega$. By additionally considering δ

large enough, we now find

$$\varphi\left(\frac{|u(x)-u(y)|}{\delta}\right) = c\,\frac{|u(x)-u(y)|^r}{\delta^r};$$

hence,

$$\mathcal{R}_\delta(u) = \frac{c}{\delta^{r-1}}\int_\Omega\int_\Omega \frac{|u(x)-u(y)|^r}{|x-y|^{n+1}}\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{\delta^{r-1}}\widetilde{\mathcal{R}}(u)$$

in analogy to (2.12).

6 Tuning the Fractional Parameter

This final section revolves around regularization via the L^2-norm of the spectral fractional Laplacian of order $s/2$, with s in the parameter range $\Lambda = (0,1)$. Our aim here is twofold. First, we determine the Mosco-limits of the regularizers, which allows us to conclude in view of the general theory in Sect. 2 that the extended bi-level problem recovers local models at the boundary points of $\overline{\Lambda} = [0,1]$. Second, we provide analytic conditions ensuring that the optimal parameter lies in the interior of $(0,1)$, and illustrate them with an explicit example.

The motivation behind the fractional Laplacian as a regularizer comes from [1], where the authors show that replacing the total variation in the classical ROF model [52] with a spectral fractional Laplacian can lead to comparable reconstruction results with a much smaller computational cost, if the order is chosen correctly. An abstract optimization of the fractional parameter for the spectral fractional Laplacian has already been undertaken in [6], although we remark that a convex penalization term is added there to the model to ensure that the optimal fractional parameter lies inside $(0,1)$.

We begin with the problem set-up. Let $\Omega \subset \mathbb{R}^n$ be a bounded Lipschitz domain and let $(\psi_m)_{m\in\mathbb{N}} \subset H_0^1(\Omega)$ be a sequence of eigenfunctions associated with the Laplace operator $(-\Delta)$ forming an orthonormal basis of $L^2(\Omega)$. With the corresponding eigenvalues $0 < \lambda_1 \le \lambda_2 \le \lambda_3 \le \cdots \nearrow \infty$, it holds for every $m \in \mathbb{N}$ that

$$\begin{cases} (-\Delta)\psi_m = \lambda_m\psi_m & \text{in } \Omega, \\ \psi_m = 0 & \text{on } \partial\Omega. \end{cases} \tag{2.1}$$

Denoting the projection of any $u \in L^2(\Omega)$ onto the mth eigenfunction ψ_m by

$$\hat{u}_m := \langle u, \psi_m\rangle_{L^2(\Omega)},$$

we have the representation $u = \sum_{m=1}^\infty \hat{u}_m\psi_m$.

With this at hand, one can define for $s \in (0, 1)$ the fractional Sobolev spaces

$$\mathbb{H}^s(\Omega) := \left\{ u = \sum_{m=1}^{\infty} \hat{u}_m \psi_m \in L^2(\Omega) : \sum_{m=1}^{\infty} \lambda_m^s \hat{u}_m^2 < \infty \right\},$$

endowed with the inner product

$$\langle u, v \rangle_{\mathbb{H}^s(\Omega)} := \sum_{m=1}^{\infty} \lambda_m^s \hat{u}_m \hat{v}_m.$$

It holds that $\mathbb{H}^s(\Omega)$ is a Hilbert space for every $s \in (0, 1)$; for more details on these spaces, we refer, e.g., to [18, 46]. In view of (2.1), the so-called spectral fractional Laplacian of order $s/2$ (with Dirichlet boundary conditions) on these spaces is defined as

$$(-\Delta_D)^{s/2} : \mathbb{H}^s(\Omega) \to L^2(\Omega), \quad (-\Delta_D)^{s/2} u = \sum_{m=1}^{\infty} \lambda_m^{s/2} \hat{u}_m \psi_m.$$

For $s \in (0, 1)$, we consider the regularizer

$$\mathcal{R}_s : L^2(\Omega) \to [0, \infty], \quad \mathcal{R}_s(u) = \begin{cases} \mu \| (-\Delta_D)^{s/2} u \|_{L^2(\Omega)}^2 & \text{for } u \in \mathbb{H}^s(\Omega), \\ \infty & \text{otherwise,} \end{cases} \quad (2.2)$$

with some $\mu > 0$. At the end of this section (see Remark 6.4), the weight parameter μ will be used to exhibit examples where structure preservation holds. The regularizers $\mathcal{R}_s$ coincide with $\mu \| \cdot \|_{\mathbb{H}^s(\Omega)}^2$ on $\mathbb{H}^s(\Omega)$, and are L^2-weakly lower semicontinuous because $u_k \rightharpoonup u$ in $L^2(\Omega)$ yields

$$\liminf_{k \to \infty} \mathcal{R}_s(u_k) = \liminf_{k \to \infty} \mu \sum_{m=1}^{\infty} \lambda_m^s \widehat{(u_k)}_m^2 \geq \mu \sum_{m=1}^{\infty} \lambda_m^s \widehat{u}_m^2 = \mathcal{R}_s(u)$$

by a discrete version of Fatou's lemma. Therefore, the hypotheses in (H) from Sect. 2 are satisfied.

Next, we determine the Mosco-limits of the regularizers, and thereby, provide the basis for extending the upper-level functional according to Sect. 2.

Proposition 6.1 (Mosco-convergence of the regularizers) *Let* $\Lambda := (0, 1)$ *and* $\mathcal{R}_s$ *for each* $s \in \Lambda$ *be given by* (2.2). *Then, for* $u \in L^2(\Omega)$ *and* $s \in \overline{\Lambda} = [0, 1]$,

$$\overline{\mathcal{R}}_s(u) = \text{Mosc}(L^2)\text{-} \lim_{s' \to s} \mathcal{R}_{s'}(u) = \begin{cases} \mathcal{R}_s(u) & \text{if } s \in (0, 1), \\ \mu \| u \|_{L^2(\Omega)}^2 & \text{if } s = 0, \quad (2.3) \\ \mu \| \nabla u \|_{L^2(\Omega)}^2 + \chi_{H_0^1(\Omega)}(u) & \text{if } s = 1. \end{cases}$$

Proof Let us observe up front that for all $u \in L^2(\Omega)$,

$$\|u\|_{L^2(\Omega)}^2 = \sum_{m=1}^{\infty} \widehat{u}_m^2 \quad \text{and} \quad \|\nabla u\|_{L^2(\Omega)}^2 + \chi_{H_0^1(\Omega)}(u) = \sum_{m=1}^{\infty} \lambda_m \widehat{u}_m^2; \qquad (2.4)$$

indeed, the first formula is simply Parseval's identity, while the second one is a consequence of $\nabla u = \sum_{m=1}^{\infty} \widehat{u}_m \nabla \psi_m$ for $u \in H_0^1(\Omega)$ and the orthogonality in $L^2(\Omega; \mathbb{R}^n)$ of the gradients $(\nabla \psi_m)_m$ with

$$\|\nabla \psi_m\|_{L^2(\Omega; \mathbb{R}^n)}^2 = -\int_{\Omega} \psi_m \, \Delta \psi_m \, dx = \int_{\Omega} \lambda_m \psi_m^2 \, dx = \lambda_m.$$

Fixing a sequence $(s_k)_k \subset (0, 1)$ with limit $s \in [0, 1]$, we want to prove now that the Mosco-limit of $(\mathcal{R}_{s_k})_k$ exists and is given by the right-hand side of (2.3).

Step 1: The liminf-inequality. Let $u_k \rightharpoonup u$ in $L^2(\Omega)$, and assume without loss of generality that $\liminf_{k \to \infty} \mathcal{R}_{s_k}(u_k) < \infty$. Then, since $\widehat{(u_k)}_m \to \widehat{u}_m$ for each $m \in \mathbb{N}$ as $k \to \infty$, it follows from a discrete version of Fatou's lemma that

$$\infty > \liminf_{k \to \infty} \mathcal{R}_{s_k}(u_k) = \liminf_{k \to \infty} \mu \sum_{m=1}^{\infty} \lambda_m^{s_k} \widehat{(u_k)}_m^2 \geq \mu \sum_{m=1}^{\infty} \lambda_m^s \widehat{u}_m^2.$$

In light of (2.4) for the cases $s \in \{0, 1\}$, the last quantity equals the regularizer on the right hand side of (2.3) in all the three regimes. This finishes the proof of the lower bound.

Step 2: Construction of a recovery sequence. We first consider the $u \in H_0^1(\Omega)$ case. By the regularity of u and Lebesgue's dominated converge theorem (applied to the counting measure) and by considering the constant recovery sequence $u_k = u$, we get

$$\lim_{k \to \infty} \mathcal{R}_{s_k}(u_k) = \lim_{k \to \infty} \mathcal{R}_{s_k}(u) = \lim_{k \to \infty} \mu \sum_{m=1}^{\infty} \lambda_m^{s_k} \widehat{u}_m^2 = \mu \sum_{m=1}^{\infty} \lambda_m^s \widehat{u}_m^2 = \mathcal{R}_s(u),$$

which concludes the proof for $u \in H_0^1(\Omega)$.

In the general case where $u \in \mathbb{H}^s(\Omega)$, we consider the sequence $(u_l)_l \subset H_0^1(\Omega)$ defined by $u_l := \sum_{m=1}^{l} \widehat{u}_m \psi_m$ for every $l \in \mathbb{N}$. Then, by construction, $u_l \to u$ strongly in $L^2(\Omega)$ and

$$\lim_{l \to \infty} \sum_{m=1}^{\infty} \lambda_m^s \widehat{(u_l)}_m^2 = \lim_{l \to \infty} \sum_{m=1}^{l} \lambda_m^s \widehat{u}_m^2 = \sum_{m=1}^{\infty} \lambda_m^s \widehat{u}_m^2.$$

The existence of a recovery sequence follows then by classical diagonalization arguments, using the previous case. $\qquad \square$

Given clean and noisy images, $u^c, u^\eta \in L^2(\Omega; \mathbb{R}^N)$, we work with the reconstruction functionals

$$\mathcal{J}_{s,j} : L^2(\mathbb{R}^n) \to [0, \infty], \quad \mathcal{J}_{s,j}(u) = \|u - u_j^\eta\|_{L^2(\Omega)}^2 + \mathcal{R}_s(u)$$

for $s \in (0, 1)$ and $j \in \{1, \ldots, N\}$. Recalling $(\mathcal{T})$ and $(\overline{\mathcal{T}})$, we obtain as a consequence of Proposition 6.1 that the extension of the upper-level functional $\mathcal{I}$ to $\overline{\Lambda}$ is given by

$$\overline{\mathcal{I}} : [0, 1] \to [0, \infty], \quad \overline{\mathcal{I}}(s) = \|w^{(s)} - u^c\|_{L^2(\Omega; \mathbb{R}^N)}^2;$$

here, $w^{(s)} = (w_1^{(s)}, \ldots, w_N^{(s)})$ with $w_j^{(s)}$ the unique minimizer of the strictly convex functional

$$\overline{\mathcal{J}}_{s,j}(u) = \|u - u_j^\eta\|_{L^2(\Omega)}^2 + \overline{\mathcal{R}}_s(u) = \sum_{m=1}^{\infty} (\hat{u}_m - \widehat{(u_j^\eta)}_m)^2 + \mu \lambda_m^s \hat{u}_m^2. \qquad (2.5)$$

By Theorem 2.5, $\overline{\mathcal{I}}$ is then the relaxation of $\mathcal{I}$ and has a minimizer in $\overline{\Lambda} = [0, 1]$.

We now continue by exhibiting conditions under which the minimum of $\overline{\mathcal{I}}$ is attained inside $(0, 1)$. This is based on a direct approach, observing that the components of $w^{(s)}$ can be determined explicitly by minimizing the entries of the sum in (2.5) individually. This gives the representation

$$w_j^{(s)} = \sum_{m=1}^{\infty} \frac{1}{1 + \mu \lambda_m^s} \widehat{(u_j^\eta)}_m \psi_m \qquad \text{for } j \in \{1, \ldots, N\}. \qquad (2.6)$$

The following lemma investigates how $w^{(s)}$ varies with s. In the $s > 0$ case, this lemma is essentially contained in [6, Theorem 2] (i.e., in a slightly different setting with periodic instead of Dirichlet boundary conditions). The proof below contains some additional details for the reader's convenience.

Lemma 6.2 *Assume that $u^\eta \in \mathbb{H}^\varepsilon(\Omega; \mathbb{R}^N)$ for some $\varepsilon \in (0, 1)$. Then, the map $[0, 1] \mapsto L^2(\Omega; \mathbb{R}^N)$, $s \mapsto w^{(s)}$ is Fréchet-differentiable with derivative*

$$\partial_s w^{(s)} = -\sum_{m=1}^{\infty} \frac{\mu \log(\lambda_m) \lambda_m^s}{(1 + \mu \lambda_m^s)^2} \widehat{u^\eta}_m \psi_m. \qquad (2.7)$$

Proof For $j \in \{1, \ldots, N\}$, we set

$$v_j := -\sum_{m=1}^{\infty} \frac{\mu \log(\lambda_m) \lambda_m^s}{(1 + \mu \lambda_m^s)^2} \widehat{(u_j^\eta)}_m \psi_m,$$

which is a well-defined element of $L^2(\Omega)$ for all $s \in [0, 1]$ because $u_j^\eta \in \mathbb{H}^\varepsilon(\Omega)$. Since

$$\frac{w_j^{(\beta)} - w_j^{(s)}}{t - s} = \sum_{m=1}^\infty \frac{1}{t - s} \left(\frac{1}{1 + \mu\lambda_m^t} - \frac{1}{1 + \mu\lambda_m^s} \right) \widehat{(u_j^\eta)}_m \psi_m, \quad s, t \in [0, 1],$$

in view of (2.6), we can apply the mean value theorem to obtain, for each $m \in \mathbb{N}$, a value γ in between s and t such that

$$\left| \frac{1}{t - s} \left(\frac{1}{1 + \mu\lambda_m^t} - \frac{1}{1 + \mu\lambda_m^s} \right) \right| \le \left| \frac{\mu \log(\lambda_m)\lambda_m^\gamma}{(1 + \mu\lambda_m^\gamma)^2} \right| \le |\log(\lambda_m)|.$$

Exploiting once again that $u_j^\eta \in \mathbb{H}^\varepsilon(\Omega)$ gives

$$\left\| \frac{w_j^{(\beta)} - w_j^{(s)}}{t - s} - v_j \right\|_{L^2(\Omega)}^2 = \sum_{m=1}^\infty \left| \frac{1}{t - s} \left(\frac{1}{1 + \mu\lambda_m^t} - \frac{1}{1 + \mu\lambda_m^s} \right) + \frac{\mu \log(\lambda_m)\lambda_m^s}{(1 + \mu\lambda_m^s)^2} \right|^2 \widehat{(u_j^\eta)}_m^2$$

$$\le \sum_{m=1}^\infty |\log(\lambda_m)|^2 \widehat{(u_j^\eta)}_m^2 < \infty.$$

In particular, we may take the limit $t \to s$ on the left-hand side of the preceding estimate and interchange with the sum to show the claim. $\qquad\square$

It follows as a consequence of Lemma 6.2 that the upper level function $\overline{\mathcal{I}} : [0, 1] \to [0, \infty]$ is differentiable with derivative

$$\overline{\mathcal{I}}'(s) = 2\left\langle \partial_s w^{(s)}, w^{(s)} - u^c \right\rangle_{L^2(\Omega;\mathbb{R}^N)}$$

for $s \in [0, 1]$; at the boundary points $s = 0$ and $s = 1$, $\overline{\mathcal{I}}'(s)$ stands for the one-sided derivative. Plugging in the identities (2.7) and (2.6) in the inner product and using that the family $(\psi_m)_m$ is orthonormal yields

$$\overline{\mathcal{I}}'(s) = -2 \sum_{j=1}^N \sum_{m=1}^\infty \frac{\mu \log(\lambda_m)\lambda_m^s}{(1 + \mu\lambda_m^s)^2} \widehat{(u_j^\eta)}_m \left(\frac{1}{1 + \mu\lambda_m^s} \widehat{(u_j^\eta)}_m - \widehat{(u_j^c)}_m \right), \quad (2.8)$$

for $s \in [0, 1]$. Observe that the simple conditions

$$\overline{\mathcal{I}}'(0) < 0 \quad \text{and} \quad \overline{\mathcal{I}}'(1) > 0,$$

imply that $\overline{\mathcal{I}}$ does not attain its minimizer at $s = 0$ or at $s = 1$, respectively. After taking $s = 0$ and $s = 1$ in (2.8) and simplifying, these requirements can be written as follows:

(H1$_s$) $\displaystyle\sum_{j=1}^N \sum_{m=1}^\infty \log(\lambda_m) \widehat{(u_j^\eta)}_m \left(\widehat{(u_j^\eta)}_m - (1 + \mu) \widehat{(u_j^c)}_m \right) > 0;$

$(H2_s)$ $\displaystyle\sum_{j=1}^{N}\sum_{m=1}^{\infty}\frac{\log(\lambda_m)\lambda_m}{(1+\mu\lambda_m)^3}\widehat{(u_j^\eta)}_m\left(\widehat{(u_j^\eta)}_m-(1+\mu\lambda_m)\widehat{(u_j^c)}_m\right)<0.$

Since $(H1_s)$ guarantees that the minimizer of $\overline{\mathcal{I}}$ is not $s=0$ and $(H2_s)$ ensures the minimizer to be different from $s=1$, Corollary 2.8 (iii) yields the following result.

Corollary 6.3 *Suppose that $u^\eta\in\mathbb{H}^\varepsilon(\Omega;\mathbb{R}^N)$ for some $\varepsilon\in(0,1)$, and that assumptions $(H1_s)$ and $(H2_s)$ are satisfied. Then, $\mathcal{I}$ admits a minimizer $\bar{s}\in(0,1)$.*

We close this section with an interpretation of the conditions $(H1_s)$ and $(H2_s)$, and a specific example in which they are both satisfied.

Remark 6.4 a) Suppose that $N=1$. Decomposing the noisy image into the sum of the clean image and the noise, i.e., $u^\eta=u^c+\eta$, turns $(H1_s)$ and $(H2_s)$ into

$$\begin{cases}\displaystyle\sum_{m=1}^{\infty}\log(\lambda_m)\left(-\mu\widehat{u^c}^2_m+(1-\mu)\widehat{u^c}_m\widehat{\eta}_m+\widehat{\eta}^2_m\right)>0,\\[2mm]\displaystyle\sum_{m=1}^{\infty}\frac{\log(\lambda_m)\lambda_m}{(1+\mu\lambda_m)^3}\left(-\mu\lambda_m\widehat{u^c}^2_m+(1-\mu\lambda_m)\widehat{u^c}_m\widehat{\eta}_m+\widehat{\eta}^2_m\right)<0.\end{cases}\tag{2.9}$$

If we assume that the noise has mostly high frequencies and that the clean image has mostly moderate frequencies, then the mixed terms in (2.9) will be small. The first condition is then close to

$$-\mu\sum_{m=1}^{\infty}\log(\lambda_m)\widehat{u^c}^2_m+\sum_{m=1}^{\infty}\log(\lambda_m)\widehat{\eta}^2_m>0,$$

which holds for sufficiently small μ. Similarly, for sufficiently large μ, the second condition is satisfied. As we analyse in b) below, there are instances where we can find a range for μ that implies both conditions.

b) In the case where $\Omega=(0,\pi)^2$, by indexing the eigenfunctions via $m=(m_1,m_2)\in\mathbb{N}^2$, we find

$$\psi_m(x)=\sin(m_1x_1)\sin(m_2x_2)$$

with corresponding eigenvalues $\lambda_m=m_1^2+m_2^2$. By choosing $u^c=\psi_{(1,1)}$ as the clean image and $\eta=\frac{1}{10}\psi_{(10,10)}$ as the noise, the condition (2.9) turns into

$$\begin{cases}-100\,\mu\log(2)+\log(200)>0,\\[2mm]-\mu\dfrac{4\log(2)}{(1+2\mu)^3}+\dfrac{2\log(200)}{(1+200\mu)^3}<0,\end{cases}$$

which is satisfied for

$$0.0236\approx\mu_-<\mu<\mu_+\approx0.0764.$$

On the other hand, when $\mu = 0.023$, then $s = 1$ is optimal, while the optimal solution for $\mu = 0.11$ is $s = 0$. This can be seen numerically as for these values of μ, the derivative $\overline{\mathcal{I}}'$ is either negative or positive on [0, 1], respectively.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

1. Antil, H., Bartels, S.: Spectral approximation of fractional PDEs in image processing and phase field modeling. Comput. Methods Appl. Math. **17**(4), 661–678 (2017)
2. Antil, H., Di, Z.W., Khatri, R.: Bilevel optimization, deep learning and fractional Laplacian regularization with applications in tomography. Inverse Problems **36**(6), 064001 (2020)
3. Antil, H., Díaz, H., Jing, T., Schikorra, A.: Nonlocal bounded variations with applications. arXiv:2208.11746 (2022)
4. Antil, H., Rautenberg, C.N.: Sobolev spaces with non-Muckenhoupt weights, fractional elliptic operators, and applications. SIAM J. Math. Anal. **51**(3), 2479–2503 (2019)
5. Aubert, G., Kornprobst, P.: Can the nonlocal characterization of Sobolev spaces by Bourgain et al. be useful for solving variational problems? SIAM J. Numer. Anal. **47**(2), 844–860 (2009)
6. Bartels, S., Weber, N.: Parameter learning and fractional differential operators: applications in regularized image denoising and decomposition problems. Math. Control Relat. Fields (2021)
7. Bellido, J.C., Mora-Corral, C.: Existence for nonlocal variational problems in peridynamics. SIAM J. Math. Anal. **46**(1), 890–916 (2014)
8. Bellido, J.C., Mora-Corral, C.: Lower semicontinuity and relaxation via Young measures for nonlocal variational problems and applications to peridynamics. SIAM J. Math. Anal. **50**(1), 779–809 (2018)
9. Belloni, M., Buttazzo, G., Freddi, L.: Completion by gamma-convergence for optimal control problems. Ann. Fac. Sci. Toulouse Math. **2**(2), 149–162 (1993)
10. Benning, M., Burger, M.: Modern regularization methods for inverse problems. Acta Numer. **27**, 1–111 (2018)
11. Boulanger, J., Elbau, P., Pontow, C., Scherzer, O.: Non-local functionals for imaging. In: Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer Optim. Appl., vol. 49, pp. 131–154. Springer, New York (2011)
12. Bourgain, J., Brezis, H., Mironescu, P.: Another look at Sobolev spaces. In: Optimal Control and Partial Differential Equations, pp. 439–455. IOS, Amsterdam (2001)
13. Braides, A.: Γ-convergence for beginners. Oxford Lecture Series in Mathematics and Its Applications, vol. 22. Oxford University Press, Oxford (2002)
14. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. SIAM J. Imaging Sci. **3**(3), 492–526 (2010)

15. Brezis, H., Nguyen, H.-M.: Non-local functionals related to the total variation and connections with image processing. Ann. PDE **4**(1), 9 (2018)
16. Buttazzo, G.: Some relaxation problems in optimal control theory. J. Math. Anal. Appl. **125**(1), 272–287 (1987)
17. Buttazzo, G., Dal Maso, G.: Γ-convergence and optimal control problems. J. Optim. Theory Appl. **38**(3), 385–407 (1982)
18. Caffarelli, L.A., Stinga, P.R.: Fractional elliptic equations, caccioppoli estimates and regularity. Ann. l'Inst. Henri Poinc. C **33**, 767–807 (2016)
19. Chambolle, A., Lions, P.-L.: Image recovery via total variation minimization and related problems. Numer. Math. **76**(2), 167–188 (1997)
20. Champion, T., De Pascale, L., Prinari, F.: Γ-convergence and absolute minimizers for supremal functionals. ESAIM Control Optim. Calc. Var. **10**(1), 14–27 (2004)
21. Chen, Y., Pock, T., Ranftl, R., Bischof, H.: Revisiting loss-specific training of filter-based mrfs for image restoration. In: Weickert, J., Hein, M., Schiele, B. (eds.) Pattern Recognition, pp. 271–281. Springer, Berlin Heidelberg (2013)
22. Chen, Y., Ranftl, R., Pock, T.: Insights into analysis operator learning: from patch-based sparse models to higher order MRFs. IEEE Trans. Image Process. **23**(3), 1060–1072 (2014)
23. Chung, C.V., de los Reyes, J.C., Schönlieb, C.B.: Learning optimal spatially-dependent regularization parameters in total variation image denoising. Inverse Problems **33**(7), 074005 (2017)
24. Dacorogna, B.: Direct Methods in the Calculus of Variations. Applied Mathematical Sciences, vol. 78, second edn. Springer, New York (2008)
25. Dal Maso, G.: An Introduction to Γ-convergence. Progress in Nonlinear Differential Equations and their Applications, vol. 8. Birkhäuser Boston Inc, Boston, MA (1993)
26. Dávila, J.: On an open question about functions of bounded variation. Calc. Var. Partial Differ. Equ. **15**(4), 519–527 (2002)
27. Davoli, E., Fonseca, I., Liu, P.: Adaptive image processing: first order PDE constraint regularizers and a bilevel training scheme. J. Nonlinear Sci. **33**(3), 41 (2023). https://doi.org/10.1007/s00332-023-09902-4
28. Davoli, E., Liu, P.: One dimensional fractional order TGV: gamma-convergence and bilevel training scheme. Commun. Math. Sci. **16**(1), 213–237 (2018)
29. De los Reyes, J.C., Schönlieb, C.-B.: Image denoising: learning the noise model via nonsmooth PDE-constrained optimization. Inverse Probl. Imaging **7**(4), 1183–1214 (2013)
30. De Los Reyes, J.C., Schönlieb, C.-B., Valkonen, T.: The structure of optimal parameters for image restoration problems. J. Math. Anal. Appl. **434**(1), 464–500 (2016)
31. De Los Reyes, J.C., Schönlieb, C.-B., Valkonen, T.: Bilevel parameter learning for higher-order total variation regularisation models. J. Math. Imaging Vis. **57**(1), 1–25 (2017)
32. D'Elia, M., De Los Reyes, J.C., Miniguano-Trujillo, A.: Bilevel parameter learning for nonlocal image denoising models. J. Math. Imaging Vis. **63**(6), 753–775 (2021)
33. Dempe, S., Zemkoho, A. (eds.): Bilevel Optimization—Advances and Next Challenges. Springer Optimization and Its Applications, vol. 161. Springer, Cham (2020)
34. Di Nezza, E., Palatucci, G., Valdinoci, E.: Hitchhiker's guide to the fractional Sobolev spaces. Bull. Sci. Math. **136**(5), 521–573 (2012)
35. Domke, J.: Generic methods for optimization-based modeling. In: Lawrence, N.D., Girolami, M. (eds.) Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 22, pp. 318–326, La Palma, Canary Islands, 21–23 Apr 2012. PMLR (2012)
36. Elbau, P.: Sequential lower semi-continuity of non-local functionals (2011)
37. Fonseca, I., Leoni, G.: Modern Methods in the Calculus of Variations: L^p Spaces. Springer Monographs in Mathematics, Springer, New York (2007)
38. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. Multiscale Model. Simul. **7**(3), 1005–1028 (2008)
39. Hintermüller, M., Rautenberg, C.N.: Optimal selection of the regularization function in a weighted total variation model. Part I: Modelling and theory. J. Math. Imaging Vis. **59**(3), 498–514 (2017)
40. Holler, G., Kunisch, K.: Learning nonlocal regularization operators. Math. Control Relat. Fields **12**(1), 81–114 (2022)
41. Holler, G., Kunisch, K., Barnard, R.C.: A bilevel approach for parameter learning in inverse problems. Inverse Problems **34**(11), 115012 (2018)

42. Kreisbeck, C., Ritorto, A., Zappale, E.: Cartesian convexity as the key notion in the variational existence theory for nonlocal supremal functionals. Nonlinear Anal. **225**, 113111 (2022)
43. Kreisbeck, C., Zappale, E.: Lower semicontinuity and relaxation of nonlocal L^∞-functionals. Calc. Var. Part. Differ. Equ. **59**(4), 138 (2020)
44. Liu, P., Schönlieb, C.-B.: Learning optimal orders of the underlying Euclidean norm in total variation image denoising. arXiv:1903.11953 (2019)
45. Muñoz, J.: Characterisation of the weak lower semicontinuity for a type of nonlocal integral functional: the n-dimensional scalar case. J. Math. Anal. Appl. **360**(2), 495–502 (2009)
46. Nochetto, R.H., Otárola, E., Salgado, A.J.: A PDE approach to fractional diffusion in general domains: a priori error analysis. Found. Comput. Math. **15**(3), 733–791 (2015)
47. Pagliari, V., Papafitsoros, K., Raiţă, B., Vikelis, A.: Bilevel training schemes in imaging for total-variation-type functionals with convex integrands. SIAM J. Imaging Sci. **15**(4), 1690–1728 (2022). https://doi.org/10.1137/21M1467328
48. Pedregal, P.: Parametrized Measures and Variational Principles. Progress in Nonlinear Differential Equations and Their Applications, vol. 30. Birkhäuser Verlag, Basel (1997)
49. Pedregal, P.: Weak lower semicontinuity and relaxation for a class of non-local functionals. Rev. Mat. Complut. **29**(3), 485–495 (2016)
50. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: Global solutions of variational models with convex regularization. SIAM J. Imaging Sci. **3**(4), 1122–1145 (2010)
51. Ponce, A.C.: A new approach to Sobolev spaces and connections to Γ-convergence. Calc. Var. Partial Differ. Equ. **19**(3), 229–255 (2004)
52. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D **60**(1–4), 259–268 (1992) (Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991))
53. Tappen, M.F., Liu, C., Adelson, E.H., Freeman, W.T.: Learning gaussian conditional random fields for low-level vision. In 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)