J. Math. Anal. Appl. 434 (2016) 464–500

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

The structure of optimal parameters for image restoration problems

J.C. De Los Reyes [a], C.-B. Schönlieb [b], T. Valkonen [a,b,*]

[a] *Research Center on Mathematical Modelling (MODEMAT), Escuela Politécnica Nacional, Ladrón de Guevara E11-253, Quito, Ecuador*
[b] *Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, CB3 0WA, Cambridge, United Kingdom*

ARTICLE INFO

Article history:
Received 8 May 2015
Available online 16 September 2015
Submitted by H. Frankowska

Keywords:
Total variation
Total generalised variation
Bi-level optimisation
Optimality
Parameter choice

ABSTRACT

We study the qualitative properties of optimal regularisation parameters in variational models for image restoration. The parameters are solutions of bilevel optimisation problems with the image restoration problem as constraint. A general type of regulariser is considered, which encompasses total variation (TV), total generalised variation (TGV) and infimal-convolution total variation (ICTV). We prove that under certain conditions on the given data optimal parameters derived by bilevel optimisation problems exist. A crucial point in the existence proof turns out to be the boundedness of the optimal parameters away from 0 which we prove in this paper. The analysis is done on the original – in image restoration typically non-smooth variational problem – as well as on a smoothed approximation set in Hilbert space which is the one considered in numerical computations. For the smoothed bilevel problem we also prove that it Γ converges to the original problem as the smoothing vanishes. All analysis is done in function spaces rather than on the discretised learning problem.

1. Introduction

In this paper we consider the general variational image reconstruction problem that, given parameters $\alpha = (\alpha_1, \dots, \alpha_N)$, $N \geq 1$, aims to compute an image

$$u_\alpha \in \operatorname*{arg\,min}_{u \in X} J(u; \alpha).$$

The image depends on α and belongs in our setting to a generic function space X. Here J is a generic energy modelling our prior knowledge on the image u_α. The quality of the solution u_α of variational imaging

* Corresponding author.
E-mail addresses: juan.delosreyes@epn.edu.ec (J.C. De Los Reyes), cbs31@cam.ac.uk (C.-B. Schönlieb), tuomo.valkonen@iki.fi (T. Valkonen).

http://dx.doi.org/10.1016/j.jmaa.2015.09.023

approaches like this one crucially relies on a good choice of the parameters α. We are particularly interested in the case

$$J(u;\alpha) = \Phi(Ku) + \sum_{j=1}^{N} \alpha_j \|A_j u\|_j,$$

with K a generic bounded forward operator, Φ a fidelity function, and A_j linear operators acting on u. The values $A_j u$ are penalised in the total variation or Radon norm $\|\mu\|_j = \|\mu\|_{\mathcal{M}(\Omega;\mathbb{R}^{m_j})}$, and combined constitute the image regulariser. In this context, α represents the regularisation parameter that balances the strength of regularisation against the fitness Φ of the solution to the idealised forward model K. The size of this parameter depends on the level of random noise and the properties of the forward operator. Choosing it too large results in over-regularisation of the solution and in turn may cause the loss of potentially important details in the image; choosing it too small under-regularises the solution and may result in a noisy and unstable output. In this work we will discuss and thoroughly analyse a bilevel optimisation approach that is able to determine the optimal choice of α in $J(;\alpha)$.

Recently, bilevel approaches for variational models have gained increasing attention in image processing and inverse problems in general. Based on prior knowledge of the problem in terms of a training set of image data and corresponding model solutions or knowledge of other model determinants such as the noise level, optimal reconstruction models are conceived by minimising a cost functional – called F in the sequel – constrained to the variational model in question. We will explain this approach in more detail in the next section. Before, let us give an account of the state of the art of bilevel optimisation for model learning. In machine learning bilevel optimisation is well established. It is a supervised learning method that optimally adapts itself to a given dataset of measurements and desirable solutions. In [41,42,28,29,18,19], for instance the authors consider bilevel optimisation for finite dimensional Markov random field (MRF) models. In inverse problems the optimal inversion and experimental acquisition setup is discussed in the context of optimal model design in works by Haber, Horesh and Tenorio [34,32,33], as well as Ghattas et al. [13,8]. Recently parameter learning in the context of functional variational regularisation models also entered the image processing community with works by the authors [23,14], Kunisch, Pock and co-workers [37,17] and Chung et al. [20]. A very interesting contribution can be found in a preprint by Fehrenbach et al. [31] where the authors determine an optimal regularisation procedure introducing particular knowledge of the noise distribution into the learning approach.

Apart from the work of the authors [23,14], all approaches for bilevel learning in image processing so far are formulated and optimised in the discrete setting. Our subsequent modelling, analysis and optimisation will be carried out in function space rather than on a discretisation of the variational model. In this context, a careful analysis of the bilevel problem is of great relevance for its application in image processing. The structure of optimal regularisers is important, among others, for the development of solution algorithms. In particular, if the parameters are bounded and lie in the interior of a closed connected set, then efficient optimisation methods can be used for solving the problem. Previous results on optimal parameters for inverse problems with partial differential equations have been obtained in, e.g., [16].

In this paper we study the qualitative structure of regularisation parameters arising as solutions from the bilevel optimisation of variational models. In our framework the variational models are typically convex but non-smooth and posed in Banach spaces. The total variation and total generalised variation regularisation models are particular instances. Alongside the optimisation of the non-smooth variational model, we also consider a smoothed approximation in Hilbert space which is typically the one considered in numerical computation. Under suitable conditions, we prove that – for both the original non-smooth optimisation problem as well as the regularised Hilbert space problem – the optimal regularisers are bounded and lie in the interior of the positive orthant. The conditions necessary to prove this turn out to be very natural

conditions on the given data in the case of an L^2 cost functional F. Indeed, for the total variation regularisers with L^2-squared cost and fidelity, we will merely require

$$\mathrm{TV}(f) > \mathrm{TV}(f_0),$$

with f_0 the ground-truth and f the noisy image. That is, the noisy image should oscillate more in terms of the total variation functional, than the ground-truth. For second-order total generalised variation [11], we obtain an analogous condition. Apart from the standard L^2 costs, we also discuss costs that constitute a smoothed L^1 norm of the gradient of the original data – we will call this the Huberised total variation cost in the sequel – typically resulting in optimal solutions superior to the ones minimising an L^2 cost. For this case, however, the interior property of optimal parameters could be verified for a finite dimensional version of the cost only. Eventually, we also show that as the numerical smoothing vanishes the optimal parameters for the smoothed models tend to optimal parameters of the original model.

The results derived in this paper are motivated by problems in image processing. However, their applicability goes well beyond that and can be generally applied to parameter estimation problems of variational inequalities of the second kind, for instance the parameter estimation problem in Bingham flow [22]. Previous analysis in this context either required box constraints on the parameters in order to prove existence of solutions or the addition of a parameter penalty to the cost functional [3,6,7,23]. In this paper, we require neither but rather prove that under certain conditions on the variational model and for reasonable data and cost functional, optimal parameters are indeed positive and guaranteed to be bounded away from 0 and ∞. As we will see later this is enough for proving existence of solutions and continuity of the solution map. The next step from our work in this here is deriving numerically useful characterisations of solutions to the ensuing bi-level programs. For the most basic problems considered herein this has been done in [23] under numerical H^1 regularisation. We will consider in the follow-up work [24] the optimality conditions for higher-order regularisers and the new cost functionals introduced in this work. For an extension characterisation of optimality systems for bi-level optimisation in finite dimensions, we point the reader to [27] as a starting point.

Outline of the paper In Section 2 we introduce the general bilevel learning problem, stating assumptions on the setup of the cost functional F and the lower level problem given by a variational regularisation approach. The bilevel problem is discussed in its original non-smooth form (P) as well as in a smoothed form in a Hilbert space setting ($\mathrm{P}^{\gamma,\epsilon}$) in Section 2.2, that will be the one used in the numerical computations. The bilevel problem is put in context with parameter learning for non-smooth variational regularisation models, typical in image processing, by proving the validity of the assumptions for examples such as TV, TGV and ICTV regularisation. The main results of the paper – existence of positive optimal parameters for L^2, Huberised TV and L^1 type costs and the convergence of the smoothed numerical problem to the original non-smooth problem – are stated in Section 3. Auxiliary results, such as coercivity, lower semicontinuity and compactness results for the involved functionals, is the topic of Section 4. Proofs for existence and convergence of optimal parameters are contained in Section 5. The paper finishes with a brief numerical discussion in Section 6.

2. The general problem setup

Let $\Omega \subset \mathbb{R}^m$ be an open bounded domain with Lipschitz boundary. This will be our image domain. Usually $\Omega = (0, w) \times (0, h)$ for w and h the width and height of a two-dimensional image, although no such assumptions are made in this work. Our noisy or corrupted data f is assumed to lie in a Banach space Y, which is the dual of Y_*, while our ground-truth f_0 lies in a general Banach space Z. Usually, in our model, we choose $Z = Y = L^2(\Omega; \mathbb{R}^d)$. This holds for denoising or deblurring, for example, where the

data is just a corrupted version of the original image. Further $d = 1$ for grayscale images, and $d = 3$ for colour images in typical colour spaces. For sub-sampled reconstruction from Fourier samples, we might use a finite-dimensional space $Y = \mathbb{C}^n$ – there are however some subtleties with that, and we refer the interested reader to [1].

As our parameter space for regularisation functional weights we take

$$\mathcal{P}_\alpha^\infty := [0, \infty]^N,$$

where N is the dimension of the parameter space, that is $\alpha = (\alpha_1, \dots, \alpha_N)$, $N \geq 1$. Observe that we allow infinite and zero values for α. The reason for the former is that in case of TGV^2, it is not reasonable to expect that *both* α_1 and α_2 are bounded; such conditions would imply that TGV^2 performs better than both TV^2 and TV. But we want our learning algorithm to find out whether that is the case! Regarding zero values, one of our main tasks is proving that for reasonable data, optimal parameters in fact lie in the interior

$$\operatorname{int} \mathcal{P}_\alpha^\infty = (0, \infty]^N.$$

This is required for the existence of solutions and the continuity of the solution map parametrised by additional regularisation parameters needed for the numerical realisation of the model. We also set

$$\mathcal{P}_\alpha := [0, \infty)^N$$

for some occasions when we need a bounded parameter.

Remark 1. In much of our treatment, we could allow for spatially dependent parameters α. However, the parameters would need to lie in a finite-dimensional subspace of $C_0(\Omega; \mathbb{R}^N)$ in our theory. Minding our general definition of the functional $J^{\gamma,0}(\,\cdot\,; \alpha)$ below, no generality is lost by taking α to be vectors in $\mathbb{R}^N$. We could simply replace the sum in the functional as a larger sum modelling integration over parameters with values in a finite-dimensional subspace of $C_0(\Omega; \mathbb{R}^N)$.

In our general learning problem, we look for $\alpha = (\alpha_1, \dots, \alpha_N)$ solving for some convex, proper, weak* lower semicontinuous cost functional $F : X \to \mathbb{R}$ the problem

$$\min_{\alpha \in \mathcal{P}_\alpha^\infty} F(u_\alpha) \tag{P}$$

subject to

$$u_\alpha \in \operatorname*{arg\,min}_{u \in X} J(u; \alpha), \tag{D_α}$$

with

$$J(u; \alpha) := \Phi(Ku) + \sum_{j=1}^{N} \alpha_j \|A_j u\|_j.$$

Here we denote for short the total variation norm

$$\|\mu\|_j := \|\mu\|_{\mathcal{M}(\Omega; \mathbb{R}^{m_j})}, \quad (\mu \in \mathcal{M}(\Omega; \mathbb{R}^{m_j})).$$

The following covers our assumptions with regard to A, K, and Φ. We discuss various specific examples in Section 2.1 immediately after stating the assumptions.

Assumption A-KA *(Operators A and K).* We assume that Y is a Banach space, and X a normed linear space, both themselves duals of Y_* and X_*, respectively. We then assume that the linear operators

$$A_j : X \to \mathcal{M}(\Omega; \mathbb{R}^{m_j}), \quad (j = 1, \dots, N),$$

and

$$K : X \to Y,$$

are bounded. Regarding K, we also assume the existence of a bounded a right-inverse $K^\dagger : \mathcal{R}(K) \to X$, where $\mathcal{R}(K)$ denotes the range of the operator K. That is, $KK^\dagger = I$ on $\mathcal{R}(K)$. We further assume that

$$\|u\|'_X := \sum_{j=1}^{N} \|A_j u\|_j + \|Ku\|_Y \tag{1}$$

is a norm on X, equivalent to the standard norm. In particular, by the Banach–Alaoglu theorem, any sequence $\{u^i\}_{i=1}^\infty \subset X$ with $\sup_i \|u^i\|'_X < \infty$ possesses a weakly* convergent subsequence.

Assumption A-Φ *(The fidelity Φ).* We suppose $\Phi : Y \to (-\infty, \infty]$ is convex, proper, weakly* lower semi-continuous, and coercive in the sense that

$$\Phi(v) \to +\infty \quad \text{as} \quad \|v\|_Y \to +\infty. \tag{2}$$

We assume that $0 \in \operatorname{dom} \Phi$, and the existence of $f \in \arg\min_{v \in Y} \Phi(v)$ such that $f = K\bar{f}$ for some $\bar{f} \in X$. Finally, we require that either K is compact, or Φ is continuous and strongly convex.

Remark 2. When $\bar{f}$ exists, we can choose $\bar{f} = K^\dagger f$.

Remark 3. Instead of $0 \in \operatorname{dom} \Phi$, it would suffice to assume, more generally, that $\operatorname{dom} \Phi \cap \bigcap_{j=1}^N \ker A_j \neq \emptyset$.

We also require the following technical assumption on the relationship of the regularisation terms and the fidelity Φ. Roughly speaking, in most interesting cases, it says that for each ℓ, we can closely approximate the noisy data f with functions $f_{\delta,\ell}$ of "order ℓ". The order here is not directly in terms of derivatives, but as defined by the operators $\{A_j\}$ through the conditions (3): $f_{\delta,j} = 0$ for $j \neq \ell$, so there is no information at other orders, while $A_\ell f_{\delta,\ell}$ is bounded, so the information at order ℓ is bounded.

Assumption A-δ *(Order reduction).* We assume that for every $\ell \in \{1, \dots, N\}$ and $\delta > 0$, there exists $\bar{f}_{\delta,\ell} \in X$ such that

$$\Phi(K\bar{f}_{\delta,\ell}) < \delta + \inf \Phi, \tag{3a}$$

$$\|A_\ell \bar{f}_{\delta,\ell}\|_\ell < \infty, \quad \text{and,} \tag{3b}$$

$$\sum_{j \neq \ell} \|A_j \bar{f}_{\delta,\ell}\|_j = 0. \tag{3c}$$

Example 1 *(Orders related to* TGV^2 *and* ICTV*).* For ICTV, as formulated in Example 6, "order 2" means that $f_{\delta,2} \in \mathrm{BV}^2(\Omega)$, while "order 1" means $f_{\delta,1} \in \mathrm{BV}(\Omega)$. For the definition of $\mathrm{BV}^2(\Omega)$, we refer to [26]. For TGV^2, as formulated in Example 5, "order 1" is likewise $\mathrm{BV}(\Omega)$, but "order 2" is the space of functions $f_{\delta,2} \in W^{1,1}(\Omega)$, such that $\nabla f_{\delta,2} \in \mathrm{BD}(\Omega)$.

2.1. Specific examples

We now discuss a few examples to motivate the abstract framework above.

Example 2 *(Squared L^2 fidelity).* With $Y = L^2(\Omega)$, $f \in Y$, and

$$\Phi(v) = \frac{1}{2}\|v - f\|_Y^2,$$

we recover the standard L^2-squared fidelity, modelling Gaussian noise. On a bounded domain Ω, Assumption A-Φ follows immediately.

With fidelity functionals as in Example 2, the operator K has a double-role in (D_α). It models any forward operator T, such as a blurring kernel or a sub-sampled Fourier transform, but also acts to extract the interesting components of u in the correct space. In its latter role, the operator is also required within F. We therefore introduce the operator K_0 for that purpose. Then for denoising we want to take $K = K_0$, but for other image processing tasks, we would combine this with the forward operator T, taking $K = TK_0$.

Example 3 *(Cost functionals).* For the cost functional F, given noise-free data $f_0 \in Z = L^2(\Omega)$, we consider in particular the L^2 cost

$$F_{L_2^2}(u) := \frac{1}{2}\|f_0 - K_0 u\|_{L^2(\Omega)}^2,$$

as well as the Huberised total variation cost

$$F_{L_\eta^1 \nabla}(u) := \|D(f_0 - K_0 u)\|_\gamma$$

with noise-free data $f_0 \in Y := \mathrm{BV}(\Omega)$. For the definition of the Huberised total variation, we refer to Section 2.2 on the numerics of the bi-level framework (P). In the numerical counterpart [24] to the present paper, we have observed that the Huberised total variation cost produces improved results compared to the L^2 cost when measured with the SSIM [46], while having the advantage of being convex unlike the SSIM.

Example 4 *(Total variation denoising).* We take K_0 as the embedding of $X = \mathrm{BV}(\Omega) \cap L^2(\Omega)$ into $Z = L^2(\Omega)$, and $A_1 = D$. We equip X with the norm

$$\|u\|_X := \|u\|_{L^2(\Omega)} + \|Du\|_{\mathcal{M}(\Omega;\mathbb{R}^n)}.$$

This makes K_0 a bounded linear operator. If the domain Ω has Lipschitz boundary, and the dimension satisfies $n \in \{1, 2\}$, the space $\mathrm{BV}(\Omega)$ continuously embeds into $L^2(\Omega)$ [2, Corollary 3.49]. Therefore, we may identify X with $\mathrm{BV}(\Omega)$ as a normed space. Otherwise, if $n \geq 3$, without going into the details of constructing X as a dual space,[1] we *define* weak* convergence in X as combined weak* convergence in $\mathrm{BV}(\Omega)$ and $L^2(\Omega)$. Any bounded sequence in X will then have a weak* convergent subsequence. This is the only property we would use from X being a dual space.

Now, combined with Example 2 and the choice $K = K_0$, $Y = Z$, we get total variation denoising for the sub-problem (D_α). Assumption A-KA holding is immediate from the previous discussion. Assumption A-δ is also easily satisfied, as with $f \in \mathrm{BV}(\Omega) \cap L^2(\Omega)$, we may simply pick $\bar{f}_{\delta,1} = f$ for the only possible case $\ell = 1$. Observe however that K is not compact unless $n = 1$, see [2, Corollary 3.49], so the strong

[1] This can be achieved by allowing $\phi_0 \in L^2(\Omega)$ instead of the $C_0(\Omega)$ in the construction of [2, Remark 3.12].

convexity of Φ is crucial here. If the data $f \in L^\infty(\Omega)$, then it is well-known that solutions $\hat{u}$ to (D_α) satisfy $\|\hat{u}\|_{L^\infty(\Omega)} \leq \|f\|_{L^\infty(\Omega)}$. We could therefore construct a compact embedding by adding some artificial constraints to the data f. This changes in the next two examples, as boundedness of solutions for higher-order regularisers is unknown; see also [43].

Example 5 *(Second order total generalised variation denoising).* We take

$$X = (\mathrm{BV}(\Omega) \cap L^2(\Omega)) \times \mathrm{BD}(\Omega),$$

the first part with the same topology as in Example 4. We also take $Z = L^2(\Omega)$, denote $u = (v, w)$, and set

$$K_0(v, w) = v, \quad A_1 u = Dv - w, \quad \text{and} \quad A_2 u = Ew$$

for E the symmetrised differential. With $K = K_0$ and $Y = Z$, this yields second-order total generalised variation (TGV^2) denoising [11] for the sub-problem (D_α). Assuming for simplicity that $\alpha_1, \alpha_2 > 0$ are constants, to show Assumption A-KA, we recall from [12] the existence of a constant $c = c(\Omega)$ such that

$$c^{-1}\|v\|_{\mathrm{BV}(\Omega)} \leq \|v\|_{L^1(\Omega)} + \|Dv\|_{\mathcal{M}(\Omega;\mathbb{R}^n)} \leq c\|v\|_{\mathrm{BV}(\Omega)},$$

where the norm

$$\|v\|_{\mathrm{BV}(\Omega)} := \mathrm{TGV}^2_{(1,1)}(v) + \|v\|_{L^1(\Omega)}.$$

We may thus approximate

$$\begin{aligned}
\|w\|_{L^1(\Omega;\mathbb{R}^n)} &\leq \|Dv - w\|_{\mathcal{M}(\Omega;\mathbb{R}^n)} + \|Dv\|_{\mathcal{M}(\Omega;\mathbb{R}^n)} \\
&\leq \|Dv - w\|_{\mathcal{M}(\Omega;\mathbb{R}^n)} + c\big(\mathrm{TGV}^2_{(1,1)}(v) + \|u\|_{L^1(\Omega)}\big) \\
&\leq (1+c)\big(\|Dv - w\|_{\mathcal{M}(\Omega;\mathbb{R}^n)} + \|Ew\|_{\mathcal{M}(\Omega;\mathbb{R}^{n\times n})}\big) + c\|v\|_{L^1(\Omega)}.
\end{aligned}$$

For some $C > 0$, it follows

$$\begin{aligned}
\|u\|_X &= \big(\|v\|_{L^2(\Omega)} + \|Dv\|_{\mathcal{M}(\Omega;\mathbb{R}^n)}\big) + \big(\|w\|_{L^1(\Omega;\mathbb{R}^n)} + \|Ew\|_{\mathcal{M}(\Omega;\mathbb{R}^{n\times n})}\big) \\
&\leq C\left(\|Dv - w\|_{\mathcal{M}(\Omega;\mathbb{R}^n)} + \|Ew\|_{\mathcal{M}(\Omega;\mathbb{R}^{n\times n})} + \|v\|_{L^2(\Omega)}\right) \\
&= C\left(\sum_{i=1}^N \|A_j u\|_j + \|Ku\|_{L^2(\Omega)}\right).
\end{aligned}$$

This shows $\|u\|_X \leq C\|u\|'_X$. The inequality $\|u\|_X \geq \|u\|'_X$ follows easily from the triangle inequality, namely

$$\|u\|_X \geq \|v\|_{L^2(\Omega)} + \|Dv - w\|_{\mathcal{M}(\Omega;\mathbb{R}^n)} + \|Ew\|_{\mathcal{M}(\Omega;\mathbb{R}^{n\times n})}.$$

Thus $\|\cdot\|'_X$ is equivalent to $\|\cdot\|_X$.

Next we observe that clearly $Dv^k - w^k \overset{*}{\rightharpoonup} Dv - w$ in $\mathcal{M}(\Omega;\mathbb{R}^n)$ and $Ew^k \overset{*}{\rightharpoonup} Ew$ in $\mathcal{M}(\Omega;\mathbb{R}^{n\times n})$ if $v^k \overset{*}{\rightharpoonup} v$ in $\mathrm{BV}(\Omega)$ and $w^k \overset{*}{\rightharpoonup} w$ weakly* in $\mathrm{BD}(\Omega)$. Thus A_1 and A_2 are weak* continuous. Assumption A-KA follows.

Consider then the satisfaction of Assumption A-δ. If $\ell = 1$, we may then pick $\bar{f}_{\delta,1} = (f, 0)$, in which case $A_1 f_{\delta,1} = Df$, and $A_2 f_{\delta,1} = 0$. If $\ell = 2$, which is the only other case, we may pick a smooth approximate $f_{\delta,2}$ to f such that

$$\frac{1}{2}\|f - f_{\delta,2}\|^2_{L^2(\Omega)} < \delta.$$

Then we set $\bar{f}_{\delta,2} := (f_{\delta,2}, \nabla f_{\delta,2})$, yielding $A_1 \bar{f}_{\delta,2} = 0$ and $A_2 \bar{f}_{\delta,2} = \nabla^2 f_{\delta,2}$. By $f_{\delta,2}$ being smooth we see that $\|A_2 \bar{f}_{\delta,2}\|_{\mathcal{M}} < \infty$. Thus Assumption A-$\delta$ is satisfied for the squared L^2 fidelity $\Phi(v) = \|f - v\|^2_{L^2(\Omega)}$.

Example 6 *(Infimal convolution TV denoising).* Let us take $Z = L^2(\Omega)$, and $X = (\mathrm{BV}(\Omega) \cap L^2(\Omega)) \times X_2$, where

$$X_2 := \{v \in W^{1,1}(\Omega) \mid \nabla v \in \mathrm{BV}(\Omega; \mathbb{R}^n)\},$$

and $\mathrm{BV}(\Omega) \cap L^2(\Omega)$ again has the topology of Example 4. Setting $u = (v, w)$, as well as

$$K_0(v, w) = v + w, \quad A_1 u = Dv, \quad \text{and} \quad A_2 u = D\nabla w,$$

we obtain for (D_α) with $K = K_0$ and $Y = Z$ the infimal convolution total variation denoising model of [15]. Assumptions A-KA, A-δ, and A-H are verified analogously to TGV^2 in Example 5.

Example 7 *(Deblurring and sub-sampled Fourier transforms for MRI).* Let K_0 and the A_js be given by one of the regularisers of Example 4 to Example 6. Also take the cost $F = F_{L^2_2}$ or $F_{L^1_\eta \nabla}$ as in Example 3, and Φ as the squared L^2 fidelity of Example 2. However, let us now take $K = TK_0$ for some bounded linear operator $T : Z \to Y$. The operator T could be, for example, a blurring kernel or a (sub-sampled) Fourier transform, in which case we obtain a model for learning the parameters for deblurring or reconstruction from Fourier samples. The latter would be important, for example for magnetic resonance imaging (MRI) [5,44,45]. Unfortunately, our theory does no extend to many of these cases because we will require, roughly, $K_0^* K_0 \leq C K^* K$ for some constant $C > 0$. In other words, *the cost functional cannot see what the fidelity does not.*

Example 8 *(Parameter estimation in Bingham flow).* Bingham fluids are materials that behave as solids if the magnitude of the stress tensor stays below a plasticity threshold, and as liquids if that quantity surpasses the threshold. In a cross sectional pipe, of section Ω, the behaviour is modelled by the energy minimisation functional

$$\min_{u \in H^1_0(\Omega)} \frac{\mu}{2} \|u\|^2_{H^1_0(\Omega)} - (f|u)_{H^{-1}(\Omega), H^1_0(\Omega)} + \alpha \int_\Omega |\nabla u| \, dx, \tag{4}$$

where $\mu > 0$ stands for the viscosity coefficient, $\alpha > 0$ for the plasticity threshold and $f \in H^{-1}(\Omega)$. In many practical situations, the plasticity threshold is not known in advance and has to be estimated from experimental measurements. One then aims at minimising a least squares term

$$F_{L^2_2} = \frac{1}{2} \|u - f_0\|^2_{L^2(\Omega)}$$

subject to (4).

The bilevel optimisation problem can then be formulated as problem (P)–(D_α), with the choices $X = Y = H^1_0(\Omega)$, K the identity, $A_1 = D$ and

$$\phi(v) = \frac{\mu}{2} \|v\|^2_{H^1_0(\Omega)} - (f|u)_{H^{-1}(\Omega), H^1_0(\Omega)}.$$

Concentrating in the rest of this paper primarily on image processing applications, we will however briefly return to Bingham flow in Example 13.

2.2. Considerations for numerical implementation

For the numerical solution of the denoising sub-problem, we will in a follow-up work [24] expand upon the infeasible semi-smooth quasi-Newton approach taken in [35] for L^2-TV image restoration problems. This depends on Huber-regularisation of the total variation measures, as well as enforcing smoothness through Hilbert spaces. This is usually done by a squared penalty on the gradient, i.e., H^1 regularisation, but we formalise this more abstractly in order to simplify our notation and arguments later on. Therefore, we take a convex, proper, and weak* lower-semicontinuous smoothing functional $H : X \to [0, \infty]$, and generally expect it to satisfy the following.

Assumption A-*H* *(Smoothing).* We assume that $0 \in \operatorname{dom} H$ and for every $\delta \geq 0$, every $\alpha \in \mathcal{P}_\alpha^\infty$, and every $u \in X$, the existence of $u^\delta \in X$ satisfying

$$H(u^\delta) < \infty \quad \text{and} \quad J(u^\delta; \alpha) \leq J(u; \alpha) + \delta. \tag{5}$$

Example 9 *(H^1 smoothing in* $\mathrm{BV}(\Omega)$*).* Usually, with $H^1(\Omega; \mathbb{R}^m) \cap X \neq \emptyset$, we take

$$H(u) := \begin{cases} \frac{1}{2}\|\nabla u\|^2_{L^2(\Omega;\mathbb{R}^{m\times n})}, & u \in H^1(\Omega;\mathbb{R}^m) \cap X, \\ \infty, & \text{otherwise.} \end{cases}$$

This is in particular the case with Example 4 (TV), where $X = \mathrm{BV}(\Omega) \cap L^2(\Omega) \supset H^1(\Omega)$, and Example 5 ($\mathrm{TGV}^2$), where $X = (\mathrm{BV}(\Omega) \cap L^2(\Omega)) \times \mathrm{BD}(\Omega) \supset H^1(\Omega) \times H^1(\Omega; \mathbb{R}^n)$ on a bounded domain Ω. In both of these cases, weak* lower semicontinuity is apparent; for completeness we record this in Lemma 1 in Section 4.2. In case of Example 4, (5) is immediate from approximating u strictly by functions in $C^\infty(\Omega)$ using standard strict approximation results in $\mathrm{BV}(\Omega)$ [2]. In case of Example 5, this also follows by a simple generalisation of the same argument to TGV-strict approximation, as presented in [43,10].

For parameters $\epsilon \geq 0$ and $\gamma \in (0, \infty]$, we then consider the problem

$$\min_{\alpha \in \mathcal{P}_\alpha^\infty} F(u_{\alpha,\gamma,\epsilon}) \tag{$\mathrm{P}^{\gamma,\epsilon}$}$$

where $u_{\alpha,\gamma,\epsilon} \in X \cap \operatorname{dom} \epsilon H$ solves

$$\min_{u \in X} J^{\gamma,\epsilon}(u; \alpha) \tag{$\mathrm{D}^{\gamma,\epsilon}$}$$

for

$$J^{\gamma,\epsilon}(u; \alpha) := \epsilon H(u) + \Phi(Ku) + \sum_{j=1}^{N} \alpha_j \|A_j u\|_{\gamma,j}.$$

Here we denote for short the Huber-regularised total variation norm

$$\|\mu\|_{\gamma,j} := \|\mu\|_{\gamma,\mathcal{M}(\Omega;\mathbb{R}^{m_j})}, \quad (\mu \in \mathcal{M}(\Omega; \mathbb{R}^{m_j})),$$

as given by the following definition. There we interpret $\gamma = \infty$ to give back the standard unregularised total variation measure. Clearly $J^{\infty,0} = J$, and (D_α) corresponds to ($\mathrm{D}^{\gamma,\epsilon}$) and (P) to ($\mathrm{P}^{\gamma,\epsilon}$) with $(\gamma, \epsilon) = (\infty, 0)$.

Definition 1. Given $\gamma \in (0, \infty]$, we define for the norm $\|\cdot\|_2$ on $\mathbb{R}^m$, the Huber regularisation

$$|g|_\gamma = \begin{cases} \|g\|_2 - \frac{1}{2\gamma}, & \|g\|_2 \geq 1/\gamma, \\ \frac{\gamma}{2}\|g\|_2^2, & \|g\|_2 < 1/\gamma. \end{cases}$$

We observe that this can equivalently be written using convex conjugates as

$$\alpha|g|_\gamma = \sup\Big\{ \langle q, g\rangle - \frac{1}{2\gamma\alpha}\|q\|_2^2 \;\Big|\; \|q\|_2 \leq \alpha \Big\}. \tag{6}$$

Then if $\mu = f\mathcal{L}^n + \mu^s$ is the Lebesgue decomposition of $\mu \in \mathcal{M}(\Omega; \mathbb{R}^m)$ into the absolutely continuous part $f\mathcal{L}^n$ and the singular part μ^s, we set

$$|\mu|_\gamma(V) := \int_V |f(x)|_\gamma \, dx + |\mu^s|(V), \quad (V \subset \Omega \text{ Borel-measureable}).$$

The measures $|\mu|_\gamma$ is the Huber-regularisation of the total variation measures $|\mu|$, and we define its Radon norm as the Huber regularisation of the Radon norm of μ, that is

$$\|\mu\|_{\gamma, \mathcal{M}(\Omega;\mathbb{R}^m)} := \||\mu|_\gamma\|_{\mathcal{M}(\Omega;\mathbb{R}^m)}.$$

Remark 4. The parameter γ varies in the literature. In this paper, we use the convention of [23], where large γ means small Huber regularisation. In [35], small γ means small Huber regularisation. That is, their regularisation parameter is $1/\gamma$ in our notation.

2.3. Shorthand notation

Writing for notational lightness

$$Au := (A_1 u, \ldots, A_N u),$$

and

$$R^\gamma(\mu_1, \ldots, \mu_N; \alpha) := \sum_{j=1}^N \alpha_j \|\mu_j\|_{\gamma, j}, \quad R(\cdot\,; \alpha) := R^\infty(\cdot\,; \alpha),$$

our problem ($\mathrm{P}^{\gamma,\epsilon}$) becomes

$$\min_{\alpha \in \mathcal{P}_\alpha^\infty} F(u_\alpha) \quad \text{subject to} \quad u_\alpha \in \arg\min_{u \in X} J^{\gamma,\epsilon}(u; \alpha)$$

for

$$J^{\gamma,\epsilon}(u; \alpha) := \epsilon H(u) + \Phi(Ku) + R^\gamma(Au; \alpha).$$

Further, given $\alpha \in \mathcal{P}_\alpha$, we define the "marginalised" regularisation functional

$$T^\gamma(v; \alpha) := \inf_{\bar{v} \in K^{-1}v} R^\gamma(A\bar{v}; \alpha), \quad T(\cdot\,; \bar{\alpha}) := T^\infty(\cdot\,; \bar{\alpha}). \tag{7}$$

Here K^{-1} stands for the preimage, so the constraint is $\bar{v} \in X$ with $v = K\bar{v}$. Then in the case $(\gamma, \epsilon) = (\infty, 0)$ and $F = F_0 \circ K$, our problem may also be written as

$$\min_{\alpha \in \mathcal{P}_\alpha^\infty} F_0(v_\alpha)$$

subject to

$$v_\alpha \in \arg\min_{v \in Y} \Phi(v) + T(v; \alpha).$$

This gives the problem a much more conventional flair, as the following examples demonstrate.

Example 10 *(*TV *as a marginal).* Consider the total variation regularisation of Example 4. Then $A_1 = D$, $N = 1$, and

$$T(v; \alpha) = R(Av; \alpha) = \alpha \mathrm{TV}(v).$$

Example 11 *(*TGV^2 *as a marginal).* In case of the TGV^2 regularisation of Example 5, we have $K(v, w) = v$ and $K^\dagger v = (v, 0)$. Thus $KK^\dagger f = f$, etc., so

$$T(v; \alpha) = \inf_{w \in \mathrm{BD}(\Omega)} R(Dv - w, Ew; \alpha) = \mathrm{TGV}_\alpha^2(v).$$

3. Main results

Our task now is to study the characteristics of optimal solutions, and their existence. Our results, based on natural assumptions on the data and the original problem (P), derive properties of the solutions to this problem and all numerically regularised problems ($\mathrm{P}^{\gamma,\epsilon}$) sufficiently close to the original problem: large $\gamma > 0$ and small $\epsilon > 0$. We denote by $u_{\alpha,\gamma,\epsilon}$ any solution to ($\mathrm{D}^{\gamma,\epsilon}$) for any given $\alpha \in \mathcal{P}_\alpha$, and by $\alpha_{\gamma,\epsilon}$ any solution to ($\mathrm{P}^{\gamma,\epsilon}$). Solutions to ($\mathrm{D}_\alpha$) and (P) we denote, respectively, by $u_\alpha = u_{\alpha,\infty,0}$, and $\hat{\alpha} = \alpha_{\infty,0}$.

3.1. L^2-squared cost and L^2-squared fidelity

Our main existence result regarding L^2-squared costs and L^2-squared fidelities is the following.

Theorem 1. *Let Y and Z be Hilbert spaces, $F(u) = \frac{1}{2}\|K_0 u - f_0\|_Z^2$, and $\Phi(v) = \frac{1}{2}\|f - v\|_Y^2$ for some $f \in \mathcal{R}(K)$, $f_0 \in Z$, and a bounded linear operator $K_0 : X \to Z$ satisfying*

$$\|K_0 u\|_Z \leq C_0 \|Ku\|_Y, \quad \text{for all } u \in X \quad \text{for some constant } C_0 > 0. \tag{8}$$

Suppose Assumptions A-KA and A-δ hold. If for some $\bar{\alpha} \in \operatorname{int} \mathcal{P}_\alpha$ and $t \in (0, 1/C_0]$ holds

$$T(f; \bar{\alpha}) > T(f - t(K_0 K^\dagger)^*(K_0 K^\dagger f - f_0); \bar{\alpha}), \tag{9}$$

then problem (P) admits a solution $\hat{\alpha} \in \operatorname{int} \mathcal{P}_\alpha^\infty$.

If, moreover, Assumption A-H holds, then there exist $\bar{\gamma} \in (0, \infty)$ and $\bar{\epsilon} \in (0, \infty)$ such that the problem ($\mathrm{P}^{\gamma,\epsilon}$) with $(\gamma, \epsilon) \in [\bar{\gamma}, \infty] \times [0, \bar{\epsilon}]$ admits a solution $\alpha_{\gamma,\epsilon} \in \operatorname{int} \mathcal{P}_\alpha^\infty$, and the solution map

$$\mathcal{S}(\gamma, \epsilon) := \arg\min_{\alpha \in \mathcal{P}_\alpha^\infty} F(u_{\alpha,\gamma,\epsilon})$$

is outer semicontinuous within $[\bar{\gamma}, \infty] \times [0, \bar{\epsilon}]$.

We prove this result in Section 5. Outer semicontinuity of a set-valued map $S : \mathbb{R}^k \rightrightarrows \mathbb{R}^m$ means [39] that for any convergent sequence $x^k \to x$ and $S(x^k) \ni y^k \to y$, we have $y \in S(x)$. In particular, the outer semicontinuity of $\mathcal{S}$ means that as the numerical regularisation vanishes, the optimal parameters for the regularised models ($P^{\gamma,\epsilon}$) tend to optimal parameters of the original model (P).

Remark 5. Let $Z = Y$ and $K_0 = K$ in Theorem 1. Then (9) reduces to

$$T(f;\bar{\alpha}) > T(f_0;\bar{\alpha}). \tag{10}$$

Also observe that our result requires, $\Phi \circ K$ to measure all the data that F measures, in the more precise sense given by (8). If (8) did not hold, an oscillating solution u_α for $\alpha \in \partial\mathcal{P}_\alpha^\infty$, could largely pass through the nullspace of K, hence have low value for the objective J of the inner problem, yet have a large cost given by F.

Corollary 1 *(Total variation Gaussian denoising). Suppose $f, f_0 \in \mathrm{BV}(\Omega) \cap L^2(\Omega)$, and*

$$\mathrm{TV}(f) > \mathrm{TV}(f_0). \tag{11}$$

Then there exist $\bar{\epsilon}, \bar{\gamma} > 0$ such that any optimal solution $\alpha_{\gamma,\epsilon}$ to the problem

$$\alpha_{\gamma,\epsilon} \in \operatorname*{arg\,min}_{\alpha \geq 0} \frac{1}{2}\|f_0 - u_{\alpha,\gamma,\epsilon}\|^2_{L^2(\Omega)}$$

with

$$u_{\alpha,\gamma,\epsilon} \in \operatorname*{arg\,min}_{u \in \mathrm{BV}(\Omega)} \Big(\frac{1}{2}\|f - u\|^2_{L^2(\Omega)} + \alpha\|Du\|_{\gamma,\mathcal{M}} + \frac{\epsilon}{2}\|\nabla v\|^2_{L^2(\Omega;\mathbb{R}^n)} \Big)$$

satisfies $\alpha_{\gamma,\epsilon} > 0$ whenever $\epsilon \in [0, \bar{\epsilon}]$, $\gamma \in [\bar{\gamma}, \infty]$.

That is, for the optimal parameter to be strictly positive, the noisy image f should, in terms of the total variation, oscillate more than the noise-free image f_0. This is a very natural condition: if the noise somehow had smoothed out features from f_0, then we should not smooth it anymore by TV regularisation!

Proof. Assumptions A-KA, A-δ, and A-H we have already verified in Example 4. We then observe that $K_0 = K$, so we are in the setting of Remark 5. Following the mapping of the TV problem to the general framework using the construction in Example 4, we have $K = I$ and $K^\dagger = I$ embeddings with $Y = L^2(\Omega)$. $K^\dagger$ is bounded on $\mathcal{R}(K) = L^2(\Omega) \cap \mathrm{BV}(\Omega)$. Moreover, by Example 10, $T(v;\bar{\alpha}) = \bar{\alpha}\mathrm{TV}(v)$. Thus (10) with the choice $t = 1$ reduces to (11). □

For TGV^2 we also have a very natural condition.

Corollary 2 *(Second-order total generalised variation Gaussian denoising). Suppose that the data $f, f_0 \in L^2(\Omega) \cap \mathrm{BV}(\Omega)$ satisfies for some $\alpha_2 > 0$ the condition*

$$\mathrm{TGV}^2_{(\alpha_2,1)}(f) > \mathrm{TGV}^2_{(\alpha_2,1)}(f_0). \tag{12}$$

Then there exist $\bar{\epsilon}, \bar{\gamma} > 0$ such that any optimal solution $\alpha_{\gamma,\epsilon} = ((\alpha_{\gamma,\epsilon})_1, (\alpha_{\gamma,\epsilon})_2)$ to the problem

$$\alpha_{\gamma,\epsilon} \in \operatorname*{arg\,min}_{\alpha \geq 0} \frac{1}{2}\|f_0 - v_{\alpha,\gamma,\epsilon}\|^2_{L^2(\Omega)}$$

with

$$(v_{\alpha,\gamma,\epsilon}, w_{\alpha,\gamma,\epsilon}) \in \underset{\substack{v\in \mathrm{BV}(\Omega)\\ w\in \mathrm{BD}(\Omega)}}{\arg\min} \Big(\frac{1}{2}\|f-v\|^2_{L^2(\Omega)} + \alpha_1\|Dv-w\|_{\gamma,\mathcal{M}} + \alpha_2\|Ew\|_{\gamma,\mathcal{M}} + \frac{\epsilon}{2}\|(\nabla v, \nabla w)\|^2_{L^2(\Omega;\mathbb{R}^n\times\mathbb{R}^{n\times n})}\Big)$$

satisfies $(\alpha_{\gamma,\epsilon})_1, (\alpha_{\gamma,\epsilon})_2 > 0$ *whenever* $\epsilon \in [0,\bar{\epsilon}]$, $\gamma \in [\bar{\gamma},\infty]$.

Proof. Assumptions A-KA, A-δ, and A-H we have already verified in Example 5. We then observe that $K_0 = K$, so we are in the setting of Remark 5. By Example 11, $T(v;\bar{\alpha}) = \mathrm{TGV}^2_{\bar{\alpha}}(v)$. Finally, similarly to (11), we get for (10) with $t = 1$ the condition (12). □

Example 12 *(Fourier reconstructions).* Let K_0 be given, for example as constructed in Example 4 or Example 5. If we take $K = \mathcal{F}K_0$ for $\mathcal{F}$ the Fourier transform – or any other unitary transform – then (8) is satisfied and $K^\dagger = K_0^\dagger\mathcal{F}^*$. Thus (9) becomes

$$T(f;\bar{\alpha}) > T(f - t(K_0K_0^\dagger\mathcal{F}^*)^*(K_0K_0^\dagger\mathcal{F}^*f - f_0);\bar{\alpha}).$$

With $\mathcal{F}^*f, f_0 \in \mathcal{R}(K_0)$ and $t = 1$ this just reduces to

$$T(f;\bar{\alpha}) > T(\mathcal{F}f_0;\bar{\alpha}).$$

Unfortunately, our results do not cover parameter learning for reconstruction from partial Fourier samples exactly because of (8). What we can do is to find the optimal parameters if we only know a part of the ground-truth, but have full noisy data.

3.2. Huberised total variation and other L^1-type costs with L^2-squared fidelity

We now consider the alternative "Huberised total variation" cost functional from Example 3. Unfortunately, we are unable to derive for $F_{L^1_\eta\nabla}$ easily interpretable conditions as for the $F_{L^2_2}$. The workaround is to "discretise" the norm, in the sense of testing only with a finite number of test functions. To be precise, recall that

$$F_{L^1_\eta\nabla}(u) = \|D(f_0 - K_0u)\|_\eta,$$

where

$$\|\mu\|_\eta = \sup_{\substack{\xi\in C_c^\infty(\Omega;\mathbb{R}^n),\\ \|\xi\|\le 1}} (\mu|\xi) - \frac{1}{2\eta}\|\xi\|_{L^2(\Omega;\mathbb{R}^n)}.$$

The idea is to replace the set of test functions ξ by a finite set

$$V = \{\xi_1,\dots,\xi_M\} \subset \{\xi \in C_c^\infty(\Omega;\mathbb{R}^n) \mid \|\xi\| \le 1\}.$$

That is, we approximate $F_{L^1_\eta\nabla}$ by

$$F^V_{L^1_\eta\nabla}(u) := \sup_{\xi\in V}(\mu|\xi) - \frac{1}{2\eta}\|\xi\|_{L^2(\Omega;\mathbb{R}^n)}, \quad \text{where} \quad \mu = D(f_0 - K_0u).$$

Choosing V suitably, this amounts to taking the finite-dimensional Huberised 1-norm of a discretised gradient of $f_0 - K_0 u$.

To study bi-level optimisation problems with the cost $F^V_{L^1_\eta \nabla}$, we rewrite the functional using a generic Huberised L^1-cost

$$F_{L^1_\eta}(z) := \sup_{\|\lambda\|_{Z^*} \le 1} (\lambda | z - \widetilde{f}_0) - \frac{1}{2\eta} \|\lambda\|^2_{Z^*}$$

on a reflexive Banach space Z, for some $\widetilde{f}_0 \in Z$. Indeed, defining

$$D^V v := \{(-\operatorname{div} \xi_1 | v), \dots, (-\operatorname{div} \xi_M | v)\}, \quad (v \in \mathrm{BV}(\Omega)),$$

we have

$$F^V_{L^1_\eta \nabla} := F_{L^1_\eta} \circ D^V \quad \text{where } \widetilde{f}_0 := D^V f_0 \text{ and } Z = \mathbb{R}^M \text{ with } \infty\text{-norm}.$$

Our results on $F^V_{L^1_\eta \nabla}$ are then based on the following general result.

Theorem 2. *Let Y be a Hilbert space, and Z a reflexive Banach space. Let $F = F_{L^1_\eta} \circ K_0$, and $\Phi(v) = \frac{1}{2}\|f - v\|^2_Y$ for some* compact *linear operator $K_0 : X \to Z$ satisfying (8) and $f \in \mathcal{R}(K)$. Suppose Assumptions A-KA and A-δ hold. If for some $\bar{\alpha} \in \operatorname{int} \mathcal{P}_\alpha$ and $t > 0$ holds*

$$T(f; \bar{\alpha}) > T(f - t(K_0 K^\dagger)^* \lambda; \bar{\alpha}), \quad \lambda \in \partial F_{L^1_\eta}(K_0 K^\dagger f), \tag{13}$$

then the problem (P) admits a solution $\hat{\alpha} \in \operatorname{int} \mathcal{P}^\infty_\alpha$.

If, moreover, Assumption A-H holds, then there exist $\bar{\gamma} \in (0, \infty)$ and $\bar{\epsilon} \in (0, \infty)$ such that the problem ($\mathrm{P}^{\gamma,\epsilon}$) with $(\gamma, \epsilon) \in [\bar{\gamma}, \infty] \times [0, \bar{\epsilon}]$ admits a solution $\alpha_{\gamma,\epsilon} \in \operatorname{int} \mathcal{P}^\infty_\alpha$, and the solution map $\mathcal{S}$ is outer semicontinuous within $[\bar{\gamma}, \infty] \times [0, \bar{\epsilon}]$.

We prove this result in Section 5.

Remark 6. If $K_0 = K$, the condition (13) has the much more legible form

$$T(f; \bar{\alpha}) > T(f - t\lambda; \bar{\alpha}), \quad \lambda \in \partial F_{L^1_\eta}(f). \tag{14}$$

Also if K is compact, then the compactness of K_0 follows from (8). In the following applications, K is however not compact for typical domains $\Omega \subset \mathbb{R}^2$ or $\mathbb{R}^3$, so we have to make K_0 compact by making the range finite-dimensional.

Corollary 3 *(Total variation Gaussian denoising with discretised Huber-TV cost). Suppose that the data satisfies $f, f_0 \in \mathrm{BV}(\Omega) \cap L^2(\Omega)$ and for some $t > 0$ and $\xi \in V$ the condition*

$$\mathrm{TV}(f) > \mathrm{TV}(f + t \operatorname{div} \xi), \quad -\operatorname{div} \xi \in \partial F^V_{L^1_\eta \nabla}(f). \tag{15}$$

Then there exist $\bar{\epsilon}, \bar{\gamma} > 0$ such any optimal solution $\alpha_{\gamma,\epsilon}$ to the problem

$$\min_{\alpha \ge 0} F^V_{L^1_\eta \nabla}(f_0 - v_\alpha)$$

with

$$u_\alpha \in \operatorname*{arg\,min}_{u\in \mathrm{BV}(\Omega)} \Big(\frac{1}{2}\|f-u\|^2_{L^2(\Omega)} + \alpha \|Du\|_{\gamma,\mathcal{M}} + \frac{\epsilon}{2}\|\nabla v\|^2_{L^2(\Omega;\mathbb{R}^n)} \Big)$$

satisfies $\alpha_{\gamma,\epsilon} > 0$ *whenever* $\epsilon \in [0,\bar{\epsilon}]$, $\gamma \in [\bar{\gamma},\infty]$.

This says that for the optimal parameter to be strictly positive, the noisy image f should oscillate more than the image $f + t\,\mathrm{div}\,\xi$ in the direction of the *(discrete) total variation flow*. This is a very natural condition, and we observe that the non-discretised counterpart of (15) for $\gamma = \infty$ would be

$$\mathrm{TV}(f) > \mathrm{TV}(f + t\,\mathrm{div}\,\xi), \quad \xi \in \mathrm{Sgn}(Df_0 - Df),$$

where we define for a measure $\mu \in \mathcal{M}(\Omega;\mathbb{R}^m)$ the sign

$$\mathrm{Sgn}(\mu) := \{\xi \in L^1(\Omega;\mu) \mid \mu = \xi|\mu|\}.$$

That is, $-\mathrm{div}\,\xi$ is the total variation flow.

Proof. Analogous to Corollary 1 regarding the L^2 cost. Note that in the application of Theorem 2, the operator D^V is absorbed by K_0. Precisely, we take $K_0 = D^V$, while $K = I$. □

For TGV^2 we also have an analogous natural condition.

Corollary 4 *(*TGV^2 *Gaussian denoising with discretised Huber-TV cost). Suppose that the data* $f, f_0 \in L^2(\Omega)$ *satisfies for some* $t, \alpha_2 > 0$ *and* $\xi \in V$ *the condition*

$$\mathrm{TGV}^2_{(\alpha_2,1)}(f) > \mathrm{TGV}^2_{(\alpha_2,1)}(f + t\,\mathrm{div}\,\xi), \quad -\mathrm{div}\,\xi \in \partial F^V_{L^1_\eta \nabla}(f). \tag{16}$$

Then there exist $\bar{\epsilon}, \bar{\gamma} > 0$ *such any optimal solution* $\alpha_{\gamma,\epsilon} = ((\alpha_{\gamma,\epsilon})_1, (\alpha_{\gamma,\epsilon})_2)$ *to the problem*

$$\min_{\alpha \geq 0} F^V_{L^1_\eta \nabla}(f_0 - v_\alpha)$$

with

$$(v_\alpha, w_\alpha) \in \operatorname*{arg\,min}_{\substack{v\in \mathrm{BV}(\Omega)\\ w\in \mathrm{BD}(\Omega)}} \Big(\frac{1}{2}\|f-v\|^2_{L^2(\Omega)} + \alpha_1 \|Dv - w\|_{\gamma,\mathcal{M}} + \alpha_2 \|Ew\|_{\gamma,\mathcal{M}} + \frac{\epsilon}{2}\|(\nabla v, \nabla w)\|^2_{L^2(\Omega;\mathbb{R}^n\times\mathbb{R}^{n\times n})} \Big)$$

satisfies $(\alpha_{\gamma,\epsilon})_1, (\alpha_{\gamma,\epsilon})_2 > 0$ *whenever* $\epsilon \in [0,\bar{\epsilon}]$, $\gamma \in [\bar{\gamma},\infty]$.

Proof. Analogous to Corollary 2. Note that for the application of Theorem 2, where in the present setting $u = (v, w)$, we take $K_0 u = D^V v$, while $Ku = v$. □

4. A few auxiliary results

We record in this section some general results that will be useful in the proofs of the main results. These include the coercivity of the functional $J^{\gamma,0}(\,\cdot\,;\lambda,\alpha)$, recorded in Section 4.1. We then discuss some elementary lower semicontinuity facts in Section 4.2. We provide in Section 4.3 some new results for passing from strict convergence to strong convergence.

4.1. Coercivity

Observe that

$$R^\gamma(\mu_1,\dots,\mu_N;\bar\alpha) = \sup_{\substack{\psi_j\in C_c^\infty(\Omega;\mathbb{R}^{m_j}\times\mathbb{R}^n)\\ \sup_{x\in\Omega}\|\psi_j(x)\|_2^2\le 1}} \sum_{j=1}^N \bar\alpha_j\left(\mu_j(\psi_j) - \frac{1}{2\gamma}\|\psi_j\|^2_{L^2(\Omega;\mathbb{R}^n)}\right).$$

Thus

$$R(\mu;\alpha) \ge R^\gamma(\mu;\alpha) \ge R(\mu;\alpha) - \frac{C'}{2\gamma}\mathcal{L}^n(\Omega)$$

for some $C' = C'(\bar\alpha)$. Since Ω is bounded, it follows that given $\delta > 0$, for large enough $\gamma > 0$ and every $\epsilon \ge 0$ holds

$$J(u;\alpha) - \delta \le J^{\gamma,0}(u;\alpha) \le J^{\gamma,\epsilon}(u;\alpha) \le J^{0,\epsilon}(u;\alpha). \tag{17}$$

We will use these properties frequently. Based on the coercivity and norm equivalence properties in Assumptions A-KA and A-Φ, the following proposition states the important fact that $J^{\gamma,0}$ is coercive with respect to $\|\cdot\|'_X$ and thus also the standard norm of X.

Proposition 1. *Suppose Assumptions A-KA and A-Φ hold, and that $\alpha \in \operatorname{int}\mathcal{P}^\infty_\alpha$. Let $\epsilon \ge 0$ and $\gamma \in (0,\infty]$. Then*

$$J^{\gamma,\epsilon}(u;\alpha) \to +\infty \quad \text{as} \quad \|u\|_X \to +\infty. \tag{18}$$

Proof. Let $\{u^i\}_{i=1}^\infty \subset X$, and suppose $\sup_i J^{\gamma,\epsilon}(u^i;\alpha) < \infty$. Then in particular $\sup_i \Phi(Ku^i) < \infty$. By Assumption A-Φ then $\sup_i \|Ku^i\|_Y < \infty$. But Assumption A-KA says

$$\|u^i\|'_X = \sum_{j=1}^N \|A_j u^i\|_j + \|Ku^i\|_Y \le J^{\gamma,\epsilon}(u^i;\alpha) + \|Ku^i\|_Y.$$

This implies $\sup_i \|u^i\|'_X < \infty$. By the equivalence of norms in Assumption A-KA, we immediately obtain (18). □

4.2. Lower semicontinuity

We record the following elementary lower semicontinuity facts that we have already used to justify our examples.

Lemma 1. *The following functionals are lower semicontinuous.*

(i) $\mu \mapsto \|\nu-\mu\|_{\gamma,\mathcal{M}}$ *with respect to weak* convergence in* $\mathcal{M}(\Omega;\mathbb{R}^d)$.
(ii) $v \mapsto \|f-v\|^2_{L^2(\Omega;\mathbb{R}^d)}$ *with respect to weak convergence in* $L^p(\Omega;\mathbb{R}^d)$ *for any* $1 < p \le 2$ *on a bounded domain* Ω.
(iii) $v \mapsto \|\nabla(f-v)\|^2_{L^2(\Omega;\mathbb{R}^{d\times n})}$ *with respect to strong convergence in* $L^1(\Omega;\mathbb{R}^d)$.

Proof. In each case, let $\{v^i\}_{i=1}^\infty$ converge to v. Denoting by G the involved functional, we write it as a convex conjugate, $G(v) = \sup\{\langle v, \varphi\rangle - G^*(\varphi)\}$. Taking a supremising sequence $\{\varphi^j\}_{j=1}^\infty$ for this functional at any point v, we easily see lower semicontinuity by considering the sequences $\{\langle v^i, \varphi^j\rangle - G^*(\varphi^j)\}_{i=1}^\infty$ for each j. In case (ii) we use the fact that $\varphi^j \in L^2(\Omega;\mathbb{R}^d) \subset [L^p(\Omega;\mathbb{R}^d)]^*$ when Ω is bounded.

In case (i), how exactly we write $G(\mu) = \|\nu - \mu\|_{\gamma,\mathcal{M}}$ as a convex conjugate demands explanation. We first of all recall that for $g \in \mathbb{R}^n$, the Huber-regularised norm may be written in dual form as

$$|g|_\gamma = \sup\left\{ \langle q, g\rangle - \frac{\gamma}{2}\|q\|_2^2 \;\middle|\; \|q\|_2 \le 1 \right\}.$$

Therefore, we find that

$$G(\mu) = \sup\left\{ \mu(\varphi) - \int_\Omega \frac{\gamma}{2}\|\varphi(x)\|_2^2\,dx \;\middle|\; \varphi \in C_c^\infty(\Omega),\ \sup_{x\in\Omega}\|\varphi(x)\|_2 \le \alpha \right\}.$$

This has the required form. □

We also show here that the marginal regularisation functional $T^\gamma(\,\cdot\,;\bar\alpha)$ is weakly* lower semicontinuous on Y. Choosing K as in Example 4 and Example 5, this provides in particular a proof that TV and TGV^2 are lower semicontinuous with respect to weak convergence in $L^2(\Omega)$ when $n = 1, 2$.

Lemma 2. *Suppose $\bar\alpha \in \operatorname{int}\mathcal{P}_\alpha^\infty$, and Assumption A-KA holds. Then $T^\gamma(\,\cdot\,;\bar\alpha)$ is lower semicontinuous with respect to weak* convergence in Y, and continuous with respect to strong convergence in $\mathcal{R}(K)$.*

Proof. Let $v^k \overset{*}{\rightharpoonup} v$ weakly* in Y. By the Banach–Steinhaus theorem, the sequence is bounded in Y. From the definition

$$T^\gamma(v;\bar\alpha) := \inf_{\bar v \in K^{-1}v} R^\gamma(A\bar v;\bar\alpha).$$

Therefore, if we pick $\epsilon > 0$ and $\bar v^k \in K^{-1}v^k$ such that

$$R^\gamma(A\bar v^k;\bar\alpha) \le T^\gamma(v^k;\bar\alpha) + \epsilon,$$

then referral to Assumption A-KA, yields for some constant $c > 0$ the bound

$$c\|\bar v^k\|_X \le \|v^k\| + R^\gamma(A\bar v^k;\bar\alpha) \le \|v^k\| + T^\gamma(v^k;\bar\alpha) + \epsilon.$$

Without loss of generality, we may assume that

$$\liminf_{k\to\infty} T^\gamma(v^k;\bar\alpha) < \infty,$$

because otherwise there is nothing to prove. Then $\{\bar v^k\}_{k=1}^\infty$ is bounded in X, and therefore admits a weakly* convergent subsequence. Let $\bar v$ be the limit of this, unrelabelled, sequence. Since K is continuous, we find that $K\bar v = v$. But $R^\gamma(\,\cdot\,;\bar\alpha)$ is clearly weak* lower semicontinuous in X; see Lemma 1. Thus

$$T^\gamma(v;\bar\alpha) \le R^\gamma(A\bar v;\bar\alpha) \le \liminf_{k\to\infty} R^\gamma(A\bar v^k;\bar\alpha) \le \liminf_{k\to\infty} T^\gamma(v^k;\bar\alpha) + \epsilon.$$

Since $\epsilon > 0$ was arbitrary, this proves weak* lower semicontinuity.

To see continuity with respect to strong convergence in $\mathcal{R}(K)$, we observe that if $v = Ku \in \mathcal{R}(K)$, then by the boundedness of the operators $\{A_j\}_{j=1}^N$ we get

$$T^\gamma(v;\bar{\alpha}) \leq R^\gamma(u;\bar{\alpha}) \leq R(u;\bar{\alpha}) \leq C\|u\|,$$

for some constant $C > 0$. So we know that $T^\gamma(\cdot\,;\bar{\alpha})|\mathcal{R}(K)$ is finite-valued and convex. Therefore it is continuous [30, Lemma I.2.1]. □

4.3. From Φ-strict to strong convergence

In Proposition 3, forming part of the proof of our main theorems, we will need to pass from "Φ-strict convergence" of Ku^k to v to strong convergence, using the following lemmas. The former means that $\Phi(Ku^k) \to \Phi(v)$ and $Ku^k \rightharpoonup v$ weakly* in Y. By strong convexity in a Banach space Y, we mean the existence of $\gamma > 0$ such that for every $y \in Y$ and $z \in \partial\Phi(y) \subset Y^*$ holds

$$\Phi(y') - \Phi(y) \geq (z|y' - y) + \frac{\gamma}{2}\|y' - y\|_Y^2, \quad (y' \in Y),$$

where $(z|y)$ denotes the dual product, and the subdifferential $\partial\Phi(y)$ is defined by z satisfying the same expression with $\gamma = 0$. With regard to more advanced strict convergence results, we point the reader to [25,38,36].

Lemma 3. *Suppose Y is a Banach space, and $\Phi : Y \to (-\infty, \infty]$ strongly convex. If $v^k \rightharpoonup \hat{v} \in \operatorname{dom}\partial\Phi$ weakly* in Y and $\Phi(v^k) \to \Phi(\hat{v})$, then $v^k \to \hat{v}$ strongly in Y.*

Remark 7. By standard convex analysis [30], $v \in \operatorname{dom}\partial\Phi$ if Φ has a finite-valued point of continuity and $v \in \operatorname{int}\operatorname{dom}\Phi$.

Proof. We first of all note that $-\infty < \Phi(\hat{v}) < \infty$ because $v \in \operatorname{dom}\partial\Phi$ implies $v \in \operatorname{dom}\Phi$. Let us pick $z \in \partial\Phi(\hat{v})$. From the strong convexity of Φ, for some $\gamma > 0$ then

$$\Phi(v^k) - \Phi(\hat{v}) \geq (z|v^k - \hat{v}) + \frac{\gamma}{2}\|v^k - \hat{v}\|_Y^2.$$

Taking the limit infimum, we observe

$$0 = \Phi(\hat{v}) - \Phi(\hat{v})d \geq \liminf_{k\to\infty} \frac{\gamma}{2}\|v^k - \hat{v}\|_Y^2.$$

This proves strong convergence. □

We now use the lemma to show strong convergence of minimising sequences.

Lemma 4. *Suppose Φ is strongly convex, satisfies Assumption A-Φ, and that $C \subset Y$ is non-empty, closed, and convex with $\operatorname{int} C \cap \operatorname{dom}\Phi \neq \emptyset$. Let*

$$\hat{v} := \arg\min_{v\in C} \Phi(v).$$

If $\{v^k\}_{k=1}^\infty \subset Y$ with $\lim_{k\to\infty}\Phi(v^k) = \Phi(\hat{v})$, then $v^k \to \hat{v}$ strongly in Y.

Proof. By the strict convexity of Φ, implied by strong convexity, and the assumptions on C, $\hat{v}$ is unique and well-defined. Moreover $\hat{v} \in \operatorname{dom} \partial\Phi$. Indeed, our assumptions show the existence of a point $v \in \operatorname{int} C \cap \operatorname{dom} \Phi$. The indicator function δ_C is then continuous at v, and so standard subdifferential calculus (see, e.g., [30, Proposition I.5.6]) implies that $\partial(\Phi + \delta_C)(\hat{v}) = \partial\Phi(\hat{v}) + \partial\delta_C(\hat{v})$. But $0 \in \partial(\Phi + \delta_C)(\hat{v})$ because $\hat{v} \in \arg\min_{v \in Y} \Phi(v) + \delta_C(v)$. This implies that $\partial\Phi(\hat{v}) \neq \emptyset$. Consequently also $\hat{v} \in \operatorname{dom} \Phi$, and $\Phi(\hat{v}) \in \mathbb{R}$.

Using the coercivity of Φ in Assumption A-Φ we then find that $\{v^k\}_{k=1}^{\infty}$ is bounded in Y, at least after moving to an unrelabelled tail of the sequence with $\Phi(v^k) \leq \Phi(\hat{v}) + 1$. Since Y is a dual space, the unit ball is weak* compact, and we deduce the existence of a subsequence, unrelabelled, and $v \in Y$ such that $v^k \rightharpoonup v$ weakly* in Y. By the weak* lower semicontinuity (Assumption A-Φ), we deduce

$$\Phi(v) \leq \liminf_{k \to \infty} \Phi(v^k) = \Phi(\hat{v}).$$

Since each $v^k \in C$, and C is closed, also $v \in C$. Therefore, by the strict convexity and the definition of $\hat{v}$, necessarily $v = \hat{v}$. Therefore $v^k \rightharpoonup \hat{v}$ weakly* in Y, and $\Phi(v^k) \to \Phi(\hat{v})$. Lemma 3 now shows that $v^k \to \hat{v}$ strongly in Y. □

5. Proofs of the main results

We now prove the existence, continuity, and non-degeneracy (interior solution) results of Section 3 through a series of lemmas and propositions, starting from general ones that are then specialised to provide the natural conditions presented in Section 3.

5.1. Existence and lower semicontinuity under lower bounds

Our principal tool for proving existence is the following proposition. We will in the rest of this section concentrate on proving the existence of the set $\mathcal{K}$ in the statement. We base this on the natural conditions of Section 3.

Proposition 2 *(Existence on compact parameter domain). Suppose Assumptions A-KA and A-Φ hold. With $\epsilon \geq 0$ and $\gamma \in (0, \infty]$ fixed, if there exists a compact set $\mathcal{K} \subset \operatorname{int} \mathcal{P}_\alpha^\infty$ with*

$$\inf_{\alpha \in \mathcal{P}_\alpha^\infty \setminus \mathcal{K}} F(u_{\alpha,\gamma,\epsilon}) > \inf_{\alpha \in \mathcal{P}_\alpha^\infty} F(u_{\alpha,\gamma,\epsilon}), \tag{19}$$

then there exists a solution $\alpha_{\gamma,\epsilon} \in \operatorname{int} \mathcal{P}_\alpha^\infty$ to ($\mathrm{P}^{\gamma,\epsilon}$). Moreover, the mapping

$$\mathcal{I}_{\gamma,\epsilon}(\alpha) := F(u_{\alpha,\gamma,\epsilon}),$$

is lower semicontinuous within $\operatorname{int} \mathcal{P}_\alpha^\infty$.

The proof depends on the following two lemmas that will be useful later on as well.

Lemma 5 *(Lower semicontinuity of the fidelity with varying parameters). Suppose Assumptions A-KA, A-Φ, and A-H hold. Let $u^k \overset{*}{\rightharpoonup} u$ weakly* in X, and $(\alpha^k, \gamma^k, \epsilon^k) \to (\alpha, \gamma, \epsilon) \in \operatorname{int} \mathcal{P}_\alpha^\infty \times (0, \infty] \times (0, \infty)$. Then*

$$J^{\gamma,\epsilon}(u; \alpha) \leq \liminf_{k \to \infty} J^{\gamma^k,\epsilon^k}(u^k; \alpha^k).$$

Proof. Let $\hat{\epsilon} > 0$ be such that $\alpha_j \geq \hat{\epsilon}$, $(j = 1, \dots, N)$. We then deduce for large k and some $C = C(\Omega, \gamma)$ that

$$\begin{aligned} \limsup_k J^{\gamma,\epsilon^k}(u^k; \hat{\epsilon}, \dots, \hat{\epsilon}) &\leq \limsup_k J^{\gamma^k,\epsilon^k}(u^k; \hat{\epsilon}, \dots, \hat{\epsilon}) + C \\ &\leq \limsup_k J^{\gamma^k,\epsilon^k}(u^k; \alpha^k) + C < \infty. \end{aligned} \tag{20}$$

Here we have assumed the final inequality to hold. This comes without loss of generality, because otherwise there is nothing to prove. Observe that this holds even if $\{\alpha^k\}_{k=1}^{\infty}$ is not bounded. In particular, if $\epsilon > 0$, restricting k to be large, we may assume that

$$C_1 := H(u^k) < \infty. \tag{21}$$

We recall that

$$J^{\gamma,\epsilon}(u; \alpha) := \epsilon H(u) + \Phi(Ku) + \sum_{j=1}^{N} \alpha_j \|A_j u\|_{\gamma,j}. \tag{22}$$

We want to show lower semicontinuity of each of the terms in turn. We start with the smoothing term. If $\epsilon > 0$, using (21), we write

$$\epsilon^k H(u^k) = (\epsilon^k - \epsilon) H(u^k) + \epsilon H(u^k) \leq (\epsilon^k - \epsilon) \frac{C_1}{2} + \epsilon H(u^k).$$

By the convergence $\epsilon^k \to \epsilon$, and the weak* lower semicontinuity of H, we find that

$$\epsilon H(u) \leq \liminf_{k \to \infty} \epsilon^k H(u^k). \tag{23}$$

If $\epsilon = 0$, we have

$$\epsilon H(u) = 0 \cdot \infty = 0,$$

while still

$$0 \leq \sup_k \epsilon^k H(u^k) < \infty.$$

Thus (23) follows.

The fidelity term $\Phi \circ K$ is weak* lower semicontinuous by the continuity of K and the weak* lower semicontinuity of Φ. It therefore remains to consider the terms in (22) involving both the regularisation parameters α, as well as the Huberisation parameter γ. Indeed using the dual formulation (6) of the Huberised norm, we have for some constant $C' = C'(\hat{\epsilon}, \Omega)$ that

$$\begin{aligned} \|A_j u^k\|_{\gamma^k,j} &= \sup_{\|\varphi(x)\| \leq 1} \int \varphi \, dA_j u^k - \frac{1}{2\gamma^k} \|\varphi\|^2 \, dx \\ &\geq \sup_{\|\varphi(x)\| \leq 1} \int_{\Omega} \varphi \, dA_j u^k + \frac{1}{2\gamma} \|\varphi\|^2 \, dx - C' |\gamma^{-1} - (\gamma^k)^{-1}| \\ &= \|A_j u^k\|_{\gamma,j} - C' |\gamma^{-1} - (\gamma^k)^{-1}|. \end{aligned}$$

Thus, if $\alpha \in \mathcal{P}_\alpha$, we get

$$\begin{aligned}\alpha_j^k \|A_j u^k\|_{\gamma^k,j} &= \alpha_j \|A_j u^k\|_{\gamma^k,j} + (\alpha_j^k - \alpha_j)\|A_j u^k\|_{\gamma^k,j} \\ &\geq \alpha_j \|A_j u^k\|_{\gamma^k,j} - |\alpha_j^k - \alpha_j| \|A_j u^k\|_{\gamma^k,j} \\ &\geq \alpha_j \|A_j u^k\|_{\gamma,j} - C'|\gamma^{-1} - (\gamma^k)^{-1}| - |\alpha_j^k - \alpha_j| \|A_j u^k\|_{\gamma^k,j}.\end{aligned}$$

It follows from (20) that the sequence $\{\|A_j u^k\|_{\gamma^k,j}\}_{k=1}^\infty$ is bounded in $\mathcal{M}(\Omega;\mathbb{R}^{m_j})$ for each $j = 1,\dots,N$. Thus

$$\liminf_{k\to\infty} \alpha_j^k \|A_j u^k\|_{\gamma^k,j} \geq \liminf_{k\to\infty} \alpha_j \|A_j u^k\|_{\gamma,j} \geq \alpha_j \|A_j u\|_{\gamma,j},$$

where the final step follows from Lemma 1.

It remains to consider the case that $\alpha \in \mathcal{P}_\alpha^\infty \setminus \mathcal{P}_\alpha$, i.e., when $\alpha_\ell = \infty$ for some ℓ. We may pick sequences $\{\beta_j^\ell\}_{\ell=1}^\infty$, $(j = 1,\dots,N)$, such that $\beta_j^\ell \nearrow \alpha_j$. Further, we may find $\{\beta_j^{k,\ell}\}_{\ell=1}^\infty$ such that $\beta_j^{k,\ell} \leq \alpha_j^k$ with $\alpha_j^k = \lim_{\ell\to\infty} \beta_j^{k,\ell}$ and $\beta_j^\ell = \lim_{k\to\infty} \beta_j^{k,\ell}$. Then, by the bounded case studied above

$$\liminf_{k\to\infty} \alpha_j^k \|A_j u^k\|_{\gamma^k,j} \geq \liminf_{k\to\infty} \beta_j^{k,\ell} \|A_j u^k\|_{\gamma^k,j} \geq \beta_j^\ell \|A_j u\|_{\gamma,j}.$$

But $\{\alpha_j^k \|A_j u^k\|_{\gamma^k,j}\}_{k=1}^\infty$ is bounded by (20), and

$$\liminf_{\ell\to\infty} \beta_j^\ell \|A_j u\|_{\gamma,j} \geq \alpha_j \|A_j u\|_{\gamma,j}.$$

Thus lower semicontinuity follows. □

Lemma 6 *(Convergence of reconstructions away from boundary). Suppose Assumptions A-KA, A-Φ, and A-H hold. Let* $(\alpha^k, \gamma^k, \epsilon^k) \to (\alpha, \gamma, \epsilon)$ *in* $\operatorname{int} \mathcal{P}_\alpha^\infty \times (0,\infty] \times [0,\infty)$. *Then we can find* $u_{\alpha,\gamma,\epsilon} \in \arg\min J^{\gamma,\epsilon}(\,\cdot\,;\alpha)$ *and extract a subsequence satisfying*

$$J^{\gamma^k,\epsilon^k}(u_{\alpha^k,\gamma^k,\epsilon^k};\alpha^k) \to J^{\gamma,\epsilon}(u_{\alpha,\gamma,\epsilon};\alpha), \tag{24a}$$

$$u_{\alpha^k,\gamma^k,\epsilon^k} \overset{*}{\rightharpoonup} u_{\alpha,\gamma,\epsilon} \quad \textit{weakly* } X, \quad \textit{and} \tag{24b}$$

$$Ku_{\alpha^k,\gamma^k,\epsilon^k} \to Ku_{\alpha,\gamma,\epsilon} \quad \textit{strongly in } Y. \tag{24c}$$

Proof. By Lemma 5, we have

$$\liminf_{k\to\infty} J^{\gamma^k,\epsilon^k}(u_{\alpha^k,\gamma^k,\epsilon^k};\alpha^k) \geq J^{\gamma,\epsilon}(u_{\alpha,\gamma,\epsilon};\alpha) = \min_{u\in X} J^{\gamma,\epsilon}(u;\alpha).$$

We also want the opposite inequality

$$\limsup_{k\to\infty} J^{\gamma^k,\epsilon^k}(u_{\alpha^k,\gamma^k,\epsilon^k};\alpha^k) \leq J^{\gamma,\epsilon}(u_{\alpha,\gamma,\epsilon};\alpha). \tag{25}$$

Let $\delta > 0$. If $\epsilon = 0$, we use Assumption A-*H* on $u = u_{\alpha,\gamma,\epsilon}$, to produce u^δ. Otherwise, we set $u^\delta = u_{\alpha,\gamma,\epsilon}$. In either case

$$J^{\gamma,\epsilon}(u^\delta;\alpha) \leq J^{\gamma,\epsilon}(u_{\alpha,\gamma,\epsilon};\alpha) + \delta.$$

In particular $A_j u^\delta = 0$ if $\alpha_j = \infty$. Then for large enough k we obtain

$$\begin{aligned} J^{\gamma^k,\epsilon^k}(u_{\alpha^k,\gamma^k,\epsilon^k};\alpha^k) &\le J^{\gamma^k,\epsilon^k}(u^\delta;\alpha^k) \\ &\le J^{\gamma,\epsilon}(u^\delta;\alpha) + \delta \\ &\le J^{\gamma,\epsilon}(u_{\alpha,\gamma,\epsilon};\alpha) + 2\delta. \end{aligned} \tag{26}$$

Since $\delta > 0$ was arbitrary, this proves (25) and consequently (24a), that is

$$\begin{aligned} \lim_{k\to\infty} J^{\gamma^k,\epsilon^k}(u_{\alpha^k,\gamma^k,\epsilon^k};\alpha^k) &= \min_{u\in X} J^{\gamma,\epsilon}(u;\alpha) = J(u_{\alpha,\gamma,\epsilon};\alpha) \\ &\le J^{\gamma,\epsilon}(0;\alpha) = \epsilon H(0) + \Phi(0) < \infty. \end{aligned} \tag{27}$$

Minding Proposition 1, this allows us to extract a subsequence of $\{u_{\alpha,\gamma^k\epsilon^k}\}_{k=1}^\infty$, unrelabelled, and convergent weakly* in X to some

$$\widetilde{u} \in \operatorname*{arg\,min}_{u\in X} J(u;\alpha).$$

We may choose $u_{\alpha,\gamma,\epsilon} := \widetilde{u}$. This shows (24b).

If K is compact, we may further assume that $Ku_{\alpha^k,\gamma^k,\epsilon^k} \to Ku_{\alpha,\gamma,\epsilon}$ strongly in Y, showing (24c). If K is not compact, Φ is continuous and strongly convex by Assumption A-Φ, and we still have $\Phi(Ku_{\alpha^k,\gamma^k,\epsilon^k}) \to \Phi(Ku_{\alpha,\gamma,\epsilon})$. The assumptions of Lemma 3 are therefore satisfied. This shows (24c). □

From the lower semicontinuity Lemma 5, we immediately obtain the following standard result.

Theorem 3 *(Existence of solutions to the reconstruction sub-problem). Let $\Omega \subset \mathbb{R}^n$ be a bounded open domain. Suppose Assumptions A-KA and A-Φ hold, and that $\alpha \in \operatorname{int} \mathcal{P}_\alpha^\infty$, $\epsilon \ge 0$, and $\gamma \in (0,\infty]$. Then ($D^{\gamma,\epsilon}$) admits a minimiser $u_{\alpha,\gamma,\epsilon} \in X \cap \operatorname{dom} \epsilon H$.*

Proof. By Lemma 5, fixing $(\alpha^k,\gamma^k,\epsilon^k) = (\alpha,\gamma,\epsilon)$, the functional $J^{\gamma,\epsilon}(\,\cdot\,;\alpha)$ is lower semicontinuous with respect to weak* convergence in X. So we just have to establish a weak* convergent minimising sequence. Towards this end, we let $\{u^k\}_{k=1}^\infty \subset \operatorname{dom} \epsilon H$ be a minimising sequence for ($D^{\gamma,\epsilon}$). We may assume without loss of generality that $\sup_k J^{\gamma,\epsilon}(u^k;\alpha) < \infty$. By Proposition 1 and the inequality $J^{\gamma,0} \le J^{\gamma,\epsilon}$, we deduce $\sup_k \|u^k\|_X < \infty$. After possibly switching to a subsequence, unrelabelled, we may therefore assume $\{u^k\}_{k=1}^\infty$ weakly* convergent in X to some $\hat{u} \in X$. This proves the claim. □

Proof of Proposition 2. Let us take a sequence $\{\alpha^k\}_{k=1}^\infty$ convergent to $\alpha \in \operatorname{int} \mathcal{P}_\alpha^\infty$. Application of Lemma 6 with $\gamma^k = \gamma$ and $\epsilon^k = \varepsilon$ and the weak* lower semicontinuity of F immediately show the lower semicontinuity of $\mathcal{I}_{\gamma,\epsilon}$ within $\operatorname{int} \mathcal{P}_\alpha^\infty$.

Finally, if $\{\alpha^k\}_{k=1}^\infty$ is a minimising sequence for ($P^{\gamma,\epsilon}$), by assumption we may take it to lie in $\mathcal{K}$. By the compactness of $\mathcal{K}$, we may assume the sequence convergent to some $\alpha \in \mathcal{K}$. By the lower semicontinuity established above, $\hat{u} = u_{\alpha,\gamma,\epsilon}$ is a solution to ($P^{\gamma,\epsilon}$). □

5.2. Towards Γ-convergence and continuity of the solution map

The next lemma, immediate from the previous one, will form the first part of the proof of continuity of the solution map. As its condition, we introduce a stronger form of (19) that is uniform over a range of ϵ and γ.

Lemma 7 *(Γ-lower limit of the cost map in terms of regularisation). Suppose Assumptions A-KA, A-Φ, and A-H hold. Let $\mathcal{K} \subset \operatorname{int} \mathcal{P}_\alpha^\infty$ be compact. Then*

$$\mathcal{I}_{\gamma,\epsilon}(\alpha) \leq \liminf_{(\alpha',\gamma',\epsilon')\to(\alpha,\gamma,\epsilon)} \mathcal{I}_{\gamma',\epsilon'}(\alpha'), \tag{28}$$

when the convergence is within $\mathcal{K} \times [\bar\gamma, \infty] \times [0, \bar\epsilon]$.

Proof. Consequence of (24b) of Lemma 6 and the weak* lower semicontinuity of F. □

The next lemma will be used to get partial strong convergence of minimisers as we approach $\partial\mathcal{P}_\alpha$. This will then be used to derive simplified conditions for this not happening. This result is the counterpart of Lemma 6 that studied convergence of reconstructions away from the boundary, and depends on the additional Assumption A-δ. This is the only place where we use the assumption, and replacing this lemma by one with different assumptions would allow us to remove Assumption A-δ.

Lemma 8 *(Convergence of reconstructions at the boundary). Suppose Assumptions A-KA, A-Φ, and A-δ hold, and that Φ is strongly convex. Suppose $\{(\alpha^k, \gamma^k, \epsilon^k)\}_{k=1}^\infty \in \operatorname{int} \mathcal{P}_\alpha^\infty \times (0,\infty] \times [0, \bar\epsilon]$ satisfies $\alpha^k \to \alpha \in \partial\mathcal{P}_\alpha^\infty$. If $\bar\epsilon = 0$ or Assumption A-H holds and $\bar\epsilon > 0$ is small enough, then $Ku_{\alpha^k,\gamma^k,\epsilon^k} \to f$ strongly in Y.*

Proof. We denote for short $u^k := u_{\alpha^k,\gamma^k,\epsilon^k}$, and note that f is unique by the strong convexity of Φ. Since $\bar\alpha \in \partial\mathcal{P}_\alpha^\infty$, there exists an index $\ell \in \{1, \dots, N\}$ such that $\alpha_\ell^k \to 0$. We let ℓ be the first such index, and pick arbitrary $\delta > 0$. We take $\bar f_{\delta,\ell}$ as given by Assumption A-δ, observing that the construction still holds with Huberisation, that is, for any $\gamma \in (0,\infty]$ and in particular any $\gamma = \gamma^k$, we have

$$\begin{aligned} \Phi(K\bar f_{\delta,\ell}) &< \delta + \Phi(f), \\ \|A_\ell \bar f_{\delta,\ell}\|_{\gamma,\ell} \leq \|A_\ell \bar f_{\delta,\ell}\|_\ell &< \infty, \quad \text{and,} \\ \sum_{j\neq\ell} \|A_j \bar f_{\delta,\ell}\|_{\gamma,j} &= 0. \end{aligned}$$

If we are aiming for $\bar\epsilon > 0$, let us also pick $\widetilde f_{\delta,\ell}$ by application of Assumption A-H to $u = \bar f_{\delta,\ell}$. Otherwise, with $\bar\epsilon = 0$, let us just set $\widetilde f_{\delta,\ell} = \bar f_{\delta,\ell}$. Since

$$u^k \in \operatorname*{arg\,min}_{u\in X} J^{\gamma^k,\epsilon^k}(u;\alpha^k),$$

we have

$$\begin{aligned} \Phi(Ku^k) \leq J^{\gamma^k,\epsilon^k}(u^k;\alpha^k) &\leq J^{\gamma^k,\epsilon^k}(\widetilde f_{\delta,\ell};\alpha^k) \leq J^{\gamma^k,\epsilon^k}(\bar f_{\delta,\ell};\alpha^k) + \delta \\ &= \epsilon^k H(\bar f_{\delta,\ell}) + \Phi(K\bar f_{\delta,\ell}) + R^{\gamma^k}(A\bar f_{\delta,\ell};\alpha^k) + \delta \\ &\leq \epsilon^k H(\bar f_{\delta,\ell}) + \Phi(f) + \alpha_\ell^k \|A_\ell \bar f_{\delta,\ell}\|_\ell + 2\delta. \end{aligned}$$

Observe that it is no problem if some index $\alpha_j^k = \infty$, because by definition as a minimiser u^k achieves smaller value than $\bar f_{\delta,\ell}$ above, and for the latter $\|A_j \bar f_{\delta,\ell}\|_{\gamma^k,j} = 0$. Choosing $\bar\epsilon > 0$ small enough, it follows for $\epsilon \in [0, \bar\epsilon]$ that

$$0 \leq \Phi(Ku^k) - \Phi(f) \leq 2\delta + \alpha_\ell^k \|A_\ell \bar f_{\delta,\ell}\|_\ell.$$

Choosing k large enough, we thus see that

$$0 \leq \Phi(Ku^k) - \Phi(f) \leq (2 + \alpha_\ell)\delta.$$

Letting $\delta \searrow 0$, we see that $\Phi(Ku^k) \to \Phi(f)$. Lemma 4 with $C = Y$ therefore shows that $Ku^k \to f$ strongly in Y. □

5.3. Minimality and co-coercivity

Our remaining task is to show the existence of $\mathcal{K}$ for (19), and of a uniform $\mathcal{K}$ – see (33) below – for the application of Lemma 7. When the fidelity and cost functionals satisfy some additional conditions, we will now reduce this to the existence of $\widetilde{\alpha} \in \operatorname{int} \mathcal{P}_\alpha$ satisfying $F(u_{\widetilde{\alpha}}) < F(\bar{f})$ for a specific $\bar{f} \in K^{-1}f$. So far, we have made no reference to the data, the ground-truth f_0 or the corrupted measurement data f. We now assume this in an abstract way, and need a type of source condition, called minimality, relating the ground truth f_0 to the noisy data f. We will get back to how this is obtained later.

Definition 2. Let $p > 0$. We say that $\bar{v} \in X$ is (K, p)-*minimal* if there exists $C \geq 0$ and $\varphi_{\bar{v}} \in Y^*$ such that

$$F(u) - F(\bar{v}) \geq (\varphi_{\bar{v}} | K(u - \bar{v})) - \frac{C}{p}\|K(u - \bar{v})\|_Y^p.$$

Remark 8. If we can take $C = 0$, then the final condition just says that $K^*\varphi_{\bar{v}} \in \partial F(\bar{v})$. This is a rather strong property. Also, instead of $t \mapsto t^p$, we could in the following proofs use any strictly increasing energy $\psi : [0, \infty) \to [0, \infty)$, $\psi(0) = 0$.

To deal with the smoothing term ϵH with $\epsilon > 0$, we also need co-coercivity; for the justification of the term for the condition in (29) below, more often seen in the context of monotone operators, we refer to the equivalences in [4, Theorem 18.15].

Definition 3. We say that F is (K, p)-*co-coercive at* $(u^*, \lambda^*) \in X \times X^*$, $\lambda^* \in \partial F(u^*)$, if

$$F(u) - F(u^*) \leq (\lambda^* | u - u^*) + \frac{C}{p}\|K(u - u^*)\|_Y^p, \quad (u \in X). \tag{29}$$

If F is (K, p)-co-coercive at (u, λ) for every $u \in X$ and $\lambda \in \partial F(u)$, we say that F is (K, p)-co-coercive. If $p = 2$, we say that F is simply K-co-coercive.

Remark 9. In essence, K-co-coercivity requires $F = F_0 \circ K$ and usual (I-)co-coercivity of F_0.

Lemma 9. *Suppose $\bar{v} \in X$ is (K, p)-minimal. If $\{u^k\}_{k=1}^\infty \subset X$ satisfies $Ku^k \to Kv$ in Y, then*

$$F(\bar{v}) \leq \liminf_{k \to \infty} F(u^k). \tag{30}$$

If, moreover, F is (K, p)-co-coercive at $(\bar{v}, K^*\varphi_{\bar{v}})$, then

$$F(\bar{v}) = \lim_{k \to \infty} F(u^k). \tag{31}$$

Proof. For (30), we use the (K, p)-minimality of $\bar{v}$ to obtain

$$F(u^k) - F(\bar{v}) \geq (\varphi_{\bar{v}} | Ku^k - v) - \frac{C}{p}\|Ku^k - v\|_Y^p, \quad (k = 1, 2, 3, \ldots).$$

Taking the limit, it follows that

$$\liminf_{k\to\infty} F(u^k) \geq F(\bar{v}).$$

If we additionally have the (K,p)-co-coercivity at $(\bar{v}, K^*\varphi_{\bar{v}})$, then, likewise

$$F(u^k) - F(\bar{v}) \leq (\varphi_{\bar{v}}|Ku^k - v) + \frac{C}{2}\|Ku^k - v\|_Y^p, \quad (k = 1,2,3,\ldots).$$

From this we immediately get

$$\limsup_{k\to\infty} F(u^k) \leq F(\bar{v}). \qquad \square$$

Proposition 3 *(One-point conditions under co-coercivity). Suppose Assumptions A-KA, A-Φ and A-δ hold, and that Φ is* strongly *convex. If $\bar{f}$ is (K,q)-minimal and*

$$\begin{array}{l} \text{some } \widetilde{\alpha} \in \operatorname{int} \mathcal{P}_\alpha \quad \text{and} \\ \quad u_{\widetilde{\alpha}} \in \arg\min_{u\in X} J(u;\widetilde{\alpha}) \quad \text{satisfy} \end{array} \quad \begin{cases} F(u_{\widetilde{\alpha}}) < F(\bar{f}), & \text{and} \\ u_\alpha \text{ is } (K,p)\text{-minimal}, \end{cases} \tag{32}$$

then there exist $\bar{\gamma}, \bar{\epsilon} > 0$ such that the following hold.

(i) *For each $\epsilon \in [0, \hat{\epsilon}]$ and $\gamma \in [\bar{\gamma}, \infty]$ there exists a compact set $\mathcal{K} \subset \operatorname{int} \mathcal{P}_\alpha^\infty$ such that (19) holds.*
(ii) *If, moreover, Assumption A-H holds and F is (K,p)-co-coercive for any $p > 0$, then there exists a compact set $\mathcal{K} \subset \operatorname{int} \mathcal{P}_\alpha^\infty$ such that*

$$\inf_{\alpha\in\mathcal{P}_\alpha^\infty\setminus\mathcal{K}} F(u_{\alpha,\gamma,\epsilon}) > \inf_{\alpha\in\mathcal{P}_\alpha^\infty} F(u_{\alpha,\gamma,\epsilon}), \quad (\gamma \in [\bar{\gamma}, \infty], \epsilon \in [0, \bar{\epsilon}]). \tag{33}$$

In both cases, the existence of $\mathcal{K}$ says that every solution $\bar{\alpha}$ to $(\mathrm{P}^{\gamma,\epsilon})$ satisfies $\bar{\alpha} \in \mathcal{K}$.

Proof. We note that f is unique by the strong convexity of Φ. Let us first prove (i). In fact, let us pick $\bar{\bar{\gamma}}, \bar{\bar{\epsilon}} > 0$ and assume with $\widetilde{\alpha}$ fixed that

$$u_{\widetilde{\alpha},\gamma,\epsilon} \in \arg\min_{u\in X} J^{\gamma,\epsilon}(u;\widetilde{\alpha}) \quad \text{satisfy} \quad F(u_{\widetilde{\alpha},\gamma,\epsilon}) < F(\bar{f}), \quad (\gamma \in [\bar{\bar{\gamma}}, \infty], \epsilon \in [0, \bar{\bar{\epsilon}}]). \tag{34}$$

We want to show the existence of a compact set $\mathcal{K} \subset \mathcal{P}_\alpha^\infty$ such that solutions $\hat{\alpha}$ to $(\mathrm{P}^{\gamma,\epsilon})$ satisfy $\hat{\alpha} \in \mathcal{K}$ whenever $(\gamma,\epsilon) \in [\bar{\gamma}, \infty] \times [0, \bar{\epsilon}]$ for $\bar{\gamma} \in [\bar{\bar{\gamma}}, \infty)$ and $\bar{\epsilon} \in (0, \bar{\bar{\epsilon}}]$ to be determined during the course of the proof. We thus let $(\alpha^k, \gamma^k, \epsilon^k) \in \mathcal{P}_\alpha^\infty \times [\bar{\gamma}, \infty] \times [0, \bar{\epsilon}]$. Since this set is compact, we may assume that $\alpha^k \to \hat{\alpha} \in \mathcal{P}_\alpha^\infty$, and $\epsilon^k \to \hat{\epsilon}$, and $\gamma^k \to \hat{\gamma}$. Suppose $\hat{\alpha} \in \partial\mathcal{P}_\alpha^\infty$. By Lemma 8 then $Ku^k \to f$ strongly in Y for small enough $\bar{\epsilon}$, with no conditions on $\bar{\gamma}$. Further by the (K,q)-minimality of $\bar{f}$ and Lemma 9 then

$$F(\bar{f}) \leq \liminf_{k\to\infty} F(Ku^k). \tag{35}$$

If we fix $\gamma^k := \gamma$ and ϵ^k, and pick $\{\alpha^k\}_{k=1}^\infty$ is a minimising sequence for $(\mathrm{P}^{\gamma,\epsilon})$, we find that (35) is in contradiction to (32). Necessarily then $\hat{\alpha} \in \operatorname{int} \mathcal{P}_\alpha^\infty$. By the lower semicontinuity result of Proposition 2, $\hat{\alpha}$ therefore has to solve $(\mathrm{P}^{\gamma,\epsilon})$. We have proved (i), because, if $\mathcal{K}$ did not exist, we could choose $\alpha^k \to \hat{\alpha} \in \partial\mathcal{P}_\alpha$.

If $\gamma^k \to \hat{\gamma}$, $\epsilon^k \to \hat{\epsilon}$, and α^k solves $(\mathrm{P}^{\gamma,\epsilon})$ for $(\gamma,\epsilon) = (\gamma^k, \epsilon^k)$, then $\alpha^k \in \operatorname{int} \mathcal{P}_\alpha^\infty$ by (i). Now (35) is in contradiction to (34). Therefore (ii) holds if (34) holds.

It remains to verify (34) for $\bar{\bar{\epsilon}} > 0$ small enough and $\bar{\bar{\gamma}} > 0$ large enough. By Lemma 6, we may find a sequence $\epsilon^k \searrow 0$ and $\gamma^k \nearrow \infty$ such that $Ku_{\widetilde{\alpha},\gamma^k,\epsilon^k} \to K\widetilde{u}_{\widetilde{\alpha}}$ for some $\widetilde{u}_{\widetilde{\alpha}} \in \arg\min J(\cdot;\widetilde{\alpha})$. Since Φ is strictly convex, and both $u_{\widetilde{\alpha}}, \widetilde{u}_{\widetilde{\alpha}} \in \arg\min_{u\in X} J(u;\widetilde{\alpha})$, we find that $K\widetilde{u}_{\widetilde{\alpha}} = Ku_{\widetilde{\alpha}}$. Recalling the (K,p)-minimality and -co-coercivity at $(u_{\widetilde{\alpha}}, K^*\varphi_{u_{\widetilde{\alpha}}})$, Lemma 9 and (32) now yield

$$\limsup_{k\to\infty} F(u_{\widetilde{\alpha},\gamma^k,\epsilon^k}) = F(u_{\widetilde{\alpha}}) < F(\bar{f}).$$

Since we may repeat the above arguments on arbitrary sequences $(\gamma^k,\epsilon^k) \to (\infty,0)$, we conclude that (34) holds for small enough $\bar{\epsilon} > 0$ and large enough $\bar{\gamma} > 0$. □

We now show the Γ-convergence of the cost map, and as a consequence the outer semicontinuity of the solution map. For an introduction to Γ-convergence, we refer to [9,21].

Proposition 4 *(Γ-convergence of the cost map and continuity of the solution map). Suppose Assumptions A-KA, A-Φ, and A-H hold along with (33). Suppose, moreover, that F is (K,p)-co-coercive, and every solution $u_{\alpha,\gamma,\epsilon}$ to $(D^{\gamma,\epsilon})$ is (K,p)-minimal with $\alpha \in \mathcal{K}$ and $(\gamma,\epsilon) \in [\bar{\gamma},\infty] \times [0,\bar{\epsilon}]$. Then*

$$\mathcal{I}_{\gamma',\epsilon'}|\mathcal{K} \xrightarrow{\Gamma} \mathcal{I}_{\gamma,\epsilon}|\mathcal{K} \tag{36}$$

when $(\gamma',\epsilon'),(\gamma,\epsilon) \in [\bar{\gamma},\infty] \times [0,\bar{\epsilon}]$ and $\mathcal{K}$ is as in (33). Moreover, the solution map

$$\mathcal{S}(\gamma,\epsilon) = \operatorname*{arg\,min}_{\alpha\in\mathcal{P}_\alpha^\infty} \mathcal{I}_{\gamma,\epsilon}(\alpha)$$

is outer semicontinuous within $[\bar{\gamma},\infty] \times [0,\bar{\epsilon}]$.

Proof. Lemma 7 shows the Γ-lower limit (28). We still have to show the Γ-upper limit. This means that given $\hat{\alpha} \in \mathcal{K}$ and $(\gamma^k,\epsilon^k) \to (\gamma,\epsilon)$ within $[0,\bar{\epsilon}] \times [\bar{\gamma},\infty]$, we have to show the existence of a sequence $\{\alpha^k\}_{k=1}^\infty \subset \mathcal{K}$ such that

$$\mathcal{I}_{\gamma,\epsilon}(\hat{\alpha}) \geq \limsup_{k\to\infty} \mathcal{I}_{\gamma^k,\epsilon^k}(\alpha^k).$$

We claim that we can take $\alpha^k = \hat{\alpha}$. With $u^k := u_{\alpha^k,\gamma^k,\epsilon^k}$, Lemma 6 gives a subsequence satisfying $Ku^k \to K\hat{u}$ strongly with $\hat{u}$ a minimiser of $J^{\gamma,\epsilon}(\cdot;\alpha)$. We just have to show that

$$F(\hat{u}) = \lim_{k\to\infty} F(u^k). \tag{37}$$

Since F is (K,p)-co-coercive, and $\hat{u}$ by assumption (K,p)-minimal, this follows from Lemma 9.

We have therefore established the Γ-convergence of $\mathcal{I}_{\gamma',\epsilon'}|\mathcal{K}$ to $\mathcal{I}_{\gamma,\epsilon}|\mathcal{K}$ as $(\gamma',\epsilon') \to (\gamma,\epsilon)$ within $[\bar{\gamma},\infty] \times [0,\bar{\epsilon}]$. Our assumption (33) says that the family $\{\mathcal{I}_{\gamma',\epsilon'} \mid (\gamma',\epsilon') \in [\bar{\gamma},\infty] \times [0,\bar{\epsilon}]\}$ is equi-mildly coercive in the sense of [9]. Therefore, by the properties of Γ-convergence, see [9, Theorem 1.12], the solution map is outer semicontinuous. □

5.4. The L^2-squared fidelity with $(K,2)$-co-coercive cost

In what follows, we seek to prove (32) for the L^2-squared fidelity with $(K,2)$-co-coercive cost functionals by imposing more natural conditions derived from (38) in the next lemma.

Lemma 10 *(Natural conditions for L^2-squared 2-co-coercive case). Suppose Assumptions A-KA and A-δ hold. Let Y be a Hilbert space, $f \in \mathcal{R}(K)$, and*

$$\Phi(v) = \frac{1}{2}\|f - v\|_Y^2.$$

Then (32) holds if $\bar{f}$ is $(K,2)$-minimal, F is $(K,2)$-co-coercive at $(\bar{f}, K^*\varphi_{\bar{f}})$ with $K^*\varphi_{\bar{f}} \in \partial F(\bar{f})$, and

$$T(f;\bar{\alpha}) > T(f - t\varphi_{\bar{f}};\bar{\alpha}) \tag{38}$$

for some $\bar{\alpha} \in \operatorname{int} \mathcal{P}_\alpha$ and $t \in (0, 1/C]$, where C is the co-coercivity constant.

Here we recall the definition of $T(\,\cdot\,;\bar{\alpha})$ from (7).

Proof. Let $\alpha \in \operatorname{int} \mathcal{P}_\alpha$. We have from the co-coercivity (29) that

$$F(\bar{f}) - F(u_\alpha) \geq -(K^*\varphi_{\bar{f}}|u_\alpha - \bar{f}) - \frac{C}{2}\|Ku_\alpha - f\|_Y^2.$$

Using the definition of the subdifferential,

$$F(u) - F(\bar{f}) \geq (K^*\varphi_{\bar{f}}|u - \bar{f}), \quad (u \in X).$$

Summing, therefore

$$F(u) - F(u_\alpha) \geq (K^*\varphi_{\bar{f}}|u - u_\alpha) - \frac{C}{2}\|Ku_\alpha - f\|_Y^2, \quad (u \in X). \tag{39}$$

Setting $u = \bar{f}$, we deduce

$$F(\bar{f}) - F(u_\alpha) \geq (\varphi_{\bar{f}}|f - Ku_\alpha) - \frac{C}{2}\|Ku_\alpha - f\|_Y^2. \tag{40}$$

Let $\alpha = t\bar{\alpha}$ for some $t > 0$. Since $\Phi \circ K$ is continuous with $\operatorname{dom}(\Phi \circ K) = X$, the optimality conditions for u_α solving ($\mathrm{D}^{\gamma,\epsilon}$) state [30, Proposition I.5.6]

$$0 \in K^*(Ku_{t\bar{\alpha}} - f) + tA^*[\partial R(\,\cdot\,;\bar{\alpha})](Au_{t\bar{\alpha}}). \tag{41}$$

Because $u_{t\bar{\alpha}}$ solves ($\mathrm{D}^{\gamma,\epsilon}$), we have $R(Au_{t\bar{\alpha}};\bar{\alpha}) = T(Ku_{t\bar{\alpha}};\bar{\alpha})$. By Lemma 11 below, therefore

$$0 \in K^*(Ku_{t\bar{\alpha}} - f) + t\psi_t, \quad \psi_t \in [\partial T(K\,\cdot\,;\bar{\alpha})](u_{t\bar{\alpha}}). \tag{42}$$

Multiplying by $(K^\dagger)^*$ we deduce $f - Ku_{t\bar{\alpha}} = t(K^\dagger)^*\psi_t$, so that referring back to (40), and using the definition of the subdifferential, we get for any $t > 0$ the estimate

$$\begin{aligned}
F(\bar{f}) - F(u_{t\bar{\alpha}}) &\geq (tK^\dagger\varphi_{\bar{f}}|\psi_t) - \frac{C}{2}\|Ku_{t\bar{\alpha}} - f\|_Y^2 \\
&= (u_{t\bar{\alpha}} - (\bar{f} - tK^\dagger\varphi_{\bar{f}})|\psi_t) + (\bar{f} - u_{t\bar{\alpha}}|\psi_t) - \frac{C}{2}\|Ku_{t\bar{\alpha}} - f\|_Y^2 \\
&\geq T(Ku_{t\bar{\alpha}};\bar{\alpha}) - T(K(\bar{f} - tK^\dagger\varphi_{\bar{f}});\bar{\alpha}) \\
&\quad + (\bar{f} - u_{t\bar{\alpha}}|\psi_t) - \frac{C}{2}\|Ku_{t\bar{\alpha}} - f\|_Y^2.
\end{aligned}$$

Since $u_{t\bar{\alpha}}$ solves ($\mathrm{D}^{\gamma,\epsilon}$) for $\alpha = t\bar{\alpha}$, using (42), we have

$$(\bar{f} - u_{t\bar{\alpha}}|\psi_t) = \frac{1}{t}\|Ku_{t\bar{\alpha}} - f\|_Y^2 \geq 2\left(T(f;\bar{\alpha}) - T(Ku_{t\bar{\alpha}};\bar{\alpha})\right).$$

It follows

$$F(\bar{f}) - F(u_{t\bar{\alpha}}) \geq T(K\bar{f};\bar{\alpha}) - T(K(\bar{f} - tK^\dagger\varphi_{\bar{f}});\bar{\alpha}) + \frac{t^{-1} - C}{2}\|Ku_{t\bar{\alpha}} - f\|_Y^2. \tag{43}$$

We see that (38) implies (32) if $0 < t \leq C^{-1}$. □

Lemma 11. *Suppose $\psi \in [\partial R(\,\cdot\,;\bar{\alpha})](Au)$ with $A^*\psi \in \mathcal{R}(K^*)$, and that we obtain $R(Au;\bar{\alpha}) = T(Ku;\bar{\alpha})$. Then $A^*\psi \in [\partial T(K\,\cdot\,;\bar{\alpha})](u)$.*

Proof. Let $\lambda \in Y$ be such that $K^*\lambda = A^*\psi$. By the definition of the subdifferential, we have

$$R(Au'';\bar{\alpha}) - R(Au;\bar{\alpha}) \geq \langle \lambda, K(u' - u)\rangle, \quad (u'' \in X).$$

Minimising over $u'' \in X$ with $Ku'' = Ku'$ for some $u' \in X$, and using $R(Au;\bar{\alpha}) = T(Ku;\bar{\alpha})$, we deduce

$$T(Ku';\bar{\alpha}) - T(Ku;\bar{\alpha}) \geq \langle \lambda, K(u' - u)\rangle, \quad (u' \in X).$$

Thus

$$T(Ku';\bar{\alpha}) - T(Ku;\bar{\alpha}) \geq \langle A^*\psi, u' - u\rangle, \quad (u' \in X).$$

This proves the claim. □

Summarising the developments so far, we may state:

Proposition 5. *Suppose Assumption A-KA holds Assumption A-δ. Let Y be a Hilbert space, $f \in Y \cap \mathcal{R}(K)$, and*

$$\Phi(v) = \frac{1}{2}\|f - v\|_Y^2.$$

If $\bar{f}$ is $(K,2)$-minimal, F is $(K,2)$-co-coercive at $(\bar{f}, K^*\varphi_{\bar{f}})$, and

$$T(f;\bar{\alpha}) > T(f - t\varphi_{\bar{f}};\bar{\alpha}) \tag{44}$$

for some $\bar{\alpha} \in \operatorname{int}\mathcal{P}_\alpha$ and $t \in (0, 1/C]$, then the claims of Proposition 3 hold.

Proof. It is easily checked that Assumption A-Φ holds. Lemma 10 then verifies the remaining conditions of Proposition 3, which shows the existence of $\mathcal{K}$ in both cases. Finally, Proposition 2 shows the existence of $\hat{\alpha} \in \operatorname{int}\mathcal{P}_\alpha^\infty$ solving ($\mathrm{P}^{\gamma,\epsilon}$). For the continuity of the solution map, we refer to Proposition 4. □

5.5. L^2-squared fidelity with L^2-squared cost

We may finally finish the proof of our main result on the L^2 fidelity $\Phi(v) := \frac{1}{2}\|f - v\|_Y^2$, $Y = L^2(\Omega;\mathbb{R}^d)$, with the L^2-squared cost functional $F(u) = \frac{1}{2}\|K_0u - f_0\|_Z^2$.

Proof of Theorem 1. We have to verify the conditions of Proposition 5, primarily the $(K,2)$-minimality of $\bar{f}$, the K-cocoercivity of F, and (44). Regarding minimality and co-coercivity, we write $F = F_0 \circ K_0$, where $F_0(v) = \frac{1}{2}\|v - f_0\|_Z^2$. Then for any $v, v' \in Z$, we have

$$F_0(v') - F_0(v) = \langle v' - v, v - f_0\rangle + \frac{1}{2}\|v' - v\|^2_{L^2(\Omega)}.$$

From this $(I,2)$-co-coercivity of F_0 with $C = 1$ is clear, as is the $(I,2)$-minimality with regard to F_0 of every $v \in Y$. By extension, F is easily seen to be $(K_0,2)$-co-coercive, and every $u \in X$ $(K_0,2)$-minimal. Using (8), $(K,2)$-co-coercivity of F with $C = C_0$ is immediate, as is the $(K,2)$-minimality of every $u \in X$.

Regarding (44), we need to find $\varphi_{\bar{f}}$ such that $K^*\varphi_{\bar{f}} = \nabla F(\bar{f})$. We have

$$K_0^*(K_0\bar{f} - f_0) = \nabla F(\bar{f}).$$

From this we observe that $\varphi_{\bar{f}}$ exists, because (8) implies $\mathcal{N}(K) \subset \mathcal{N}(K_0)$, and hence $\mathcal{R}(K^*) \subset \mathcal{R}(K_0^*)$. Here $\mathcal{N}$ and $\mathcal{R}$ stand for the nullspace and range, respectively. Setting $K^*\varphi_{\bar{f}} = K_0^*(K_0\bar{f} - f_0)$ and using $KK^\dagger = I$ on $\mathcal{R}(K)$, we thus find that

$$\varphi_{\bar{f}} = (K_0K^\dagger)^*(K_0\bar{f} - f_0).$$

Observe that since $\mathcal{N}(K) \subset \mathcal{N}(K_0)$, this expression does not depend on the choice of $\bar{f} \in K^{-1}f$. Following Remark 2, we can replace $\bar{f} = K^\dagger f$. It follows that (10) implies (44). □

Remark 10. Provided that Φ satisfies Assumptions A-Φ, A-δ, and A-H, it is easy to extend Lemma 10 and consequently Theorem 1 to the case

$$\Phi(v) = \frac{1}{2}\|v\|^2_{\hat{Y}} - (f|v)_{Y^*,Y}, \quad (v \in Y),$$

where $\hat{Y} \supset Y$ is a Hilbert space, $f \in Y^*$, and Y still a reflexive Banach space. As $Y \subset \hat{Y} = \hat{Y}^* \subset Y^*$, in this case, we still have

$$\nabla\Phi(v) = v - f \in Y^*.$$

In particular

$$\nabla\Phi(Ku) = K^*(Ku - f) \in X^*.$$

Therefore the expression (41) still holds, which is the only place where we needed the specific form of Φ.

Example 13 *(Bingham flow).* As a particular case of this remark, we take $\hat{Y} = Y = H_0^1(\Omega)$. Then $Y^* = H^{-1}(\Omega)$. With $f \in L^2(\Omega)$, the Riesz representation theorem allows us to write

$$\int_\Omega fv\,dx = (\widetilde{f}|v)_{H^{-1}(\Omega),H_0^1(\Omega)},$$

for some $\widetilde{f} \in H^{-1}(\Omega)$, which we may identify with f. Therefore, Theorem 1 can be extended to cover the Bingham flow of Example 8. In particular, we get the same condition for interior solutions as in Corollary 1, namely

$$\mathrm{TV}(f) > \mathrm{TV}(f_0).$$

5.6. A more general technique for the L^2-squared fidelity

We now study another technique that does not require $(K,2)$-minimality and $(K,2)$-co-coercivity at $\bar{f}$. We still however require Φ to be the L^2-squared fidelity, and $u_{\bar{\alpha}}$ to be (K,p)-minimal.

Lemma 12 *(Natural conditions for the general L^2-squared case). Suppose Assumptions A-KA and A-δ hold. Let Y a Hilbert space, $f \in Y \cap \mathcal{R}(K)$, and*

$$\Phi(v) = \frac{1}{2}\|f - v\|_Y^2.$$

The claims of Proposition 3 hold if for some $\bar{\alpha} \in \operatorname{int} \mathcal{P}_\alpha$ and $t > 0$ both $\bar{f}$ and the solution $u_{\bar{\alpha}}$ are (K,p)-minimal and

$$T(Ku_{\bar{\alpha}}; \bar{\alpha}) > T(f - t\varphi_{u_{\bar{\alpha}}}; \bar{\alpha}). \tag{45}$$

Remark 11. Setting $\bar{\alpha} = s\bar{\bar{\alpha}}$ in (45), we see employing the lower semicontinuity Lemma 2 that the former is implied by

$$T(f; \bar{\alpha}) > \limsup_{s \searrow 0} T(f - t\varphi_{s\bar{\bar{\alpha}}}; \bar{\alpha}). \tag{46}$$

Here we use the shorthand $\varphi_{s\bar{\bar{\alpha}}} := \varphi_{u_{s\bar{\bar{\alpha}}}}$. The difficulty is going to the limit, because we do not generally have any reasonable form of convergence of $\{\varphi_{s\bar{\bar{\alpha}}}\}_{s>0}$. If we did indeed have $\varphi_{s\bar{\bar{\alpha}}} \to \varphi_{\bar{\bar{f}}}$, then (46) and consequently (50) would be implied by the condition (44) we derived using $(K,2)$-co-coercivity. We will in the next subsection go to the limit with finite-dimensional functionals that are not $(K,2)$-co-coercive and hence the earlier theory does not apply.

Proof. Let us observe that (32) holds if for some $\bar{\alpha} > 0$ and $C > 0$, we can find a (K,p)-minimal

$$u_{\bar{\alpha}} \in \operatorname*{arg\,min}_{u \in X} J(u; \bar{\alpha}),$$

satisfying

$$(\varphi_{\bar{\alpha}} | f - Ku_{\bar{\alpha}}) > 0. \tag{47}$$

Here we denote for short $\varphi_{\bar{\alpha}} := \varphi_{u_{\bar{\alpha}}}$, recalling that $K^*\varphi_{\bar{\alpha}} \in \partial F(u_{\bar{\alpha}})$. Indeed, by the definition of the subdifferential, the minimality of $u_{\bar{\alpha}}$, and (47), we deduce

$$F(\bar{f}) \geq F(u_{\bar{\alpha}}) + (\varphi_{\bar{\alpha}} | f - Ku_{\bar{\alpha}}) > F(u_{\bar{\alpha}}). \tag{48}$$

This shows (32).

We need to show that (45) implies (47). As in the proof of Lemma 10, we deduce by application of Lemma 11 that

$$0 \in K^*(Ku_{\bar{\alpha}} - f) + \psi_{\bar{\alpha}}, \quad \text{for some} \quad \psi_{\bar{\alpha}} \in [\partial T(K\,\cdot\,; \bar{\alpha})](u_{\bar{\alpha}}). \tag{49}$$

Then

$$(\bar{f} - u_{\bar{\alpha}} | \psi_{\bar{\alpha}}) = (\bar{f} - u_{\bar{\alpha}} | K^*(f - Ku_{\bar{\alpha}})) = \|f - Ku_{\bar{\alpha}}\|_Y^2 \geq 0.$$

Multiplying (49) by $(K^\dagger)^*$ and using this estimate, we deduce for any $t > 0$ that

$$\begin{aligned}(\varphi_{\bar\alpha}|f - Ku_{\bar\alpha}) &\geq t^{-1}(u_{\bar\alpha} - (u_{\bar\alpha} - tK^\dagger\varphi_{\bar\alpha})|\psi_{\bar\alpha})\\ &= t^{-1}(u_{\bar\alpha} - (\bar f - tK^\dagger\varphi_{\bar\alpha})|\psi_{\bar\alpha}) + t^{-1}(\bar f - u_{\bar\alpha}|\psi_{\bar\alpha})\\ &\geq t^{-1}(u_{\bar\alpha} - (\bar f - tK^\dagger\varphi_{\bar\alpha})|\psi_{\bar\alpha})\\ &\geq t^{-1}\bar\alpha\left(T(Ku_{\bar\alpha};\bar\alpha) - T(K(\bar f - tK^\dagger\varphi_{\bar\alpha});\bar\alpha)\right).\end{aligned}$$

The last step follows from the definition of $\partial T(K\,\cdot\,;\bar\alpha)$. This proves that (45) implies (47). □

Summing up the developments so far, we may in contrast to Proposition 5 that depended on $\bar f$ and co-coercivity, state:

Proposition 6. *Suppose Assumptions A-KA and A-δ hold. Let Y be a Hilbert space, $f \in Y \cap \mathcal{R}(K)$, and*

$$\Phi(v) = \frac{1}{2}\|f - v\|_Y^2.$$

If for some $\bar{\bar\alpha} \in \operatorname{int}\mathcal{P}_\alpha$, $t > 0$, the solution $u_{\bar{\bar\alpha}}$ is (K,p)-minimal with

$$T(Ku_{\bar{\bar\alpha}};\bar{\bar\alpha}) > T(f - t\varphi_{u_{\bar{\bar\alpha}}};\bar{\bar\alpha}), \tag{50}$$

then there exist $\bar\gamma > 0$ and $\bar\epsilon > 0$ such that the following hold.

(i) *For each $\epsilon \in [0,\bar\epsilon]$ and $\gamma \in [\bar\gamma,\infty]$ there exists a compact set $\mathcal{K} \subset \operatorname{int}\mathcal{P}_\alpha^\infty$ such that (19) holds. In particular there exists a solution $\hat\alpha \in \operatorname{int}\mathcal{P}_\alpha^\infty$ to ($\mathrm{P}^{\gamma,\epsilon}$).*

(ii) *If, moreover, Assumption A-H holds, there exists a compact set $\mathcal{K} \subset \operatorname{int}\mathcal{P}_\alpha^\infty$ such that (33) holds and the solution map $\mathcal{S}$ is outer semicontinuous within $[\bar\gamma,\infty] \times [0,\bar\epsilon]$.*

Proof. It is easily checked that Assumption A-Φ holds. Lemma 12 then verifies the remaining conditions of Proposition 3, which shows the existence of $\mathcal{K}$ in both cases. Finally, Proposition 2 shows the existence of $\hat\alpha \in \operatorname{int}\mathcal{P}_\alpha^\infty$ solving ($\mathrm{P}^{\gamma,\epsilon}$). For the continuity of the solution map, we refer to Proposition 4. □

5.7. L^2-squared fidelity with Huberised L^1-type cost

We now study the Huberised total variation cost functional. We cannot in general prove that solutions u_α for small α are better than f. Consider, for example f_0 a step function, and f a noisy version without the edge destroyed. The solution u_α *might* smooth out the edge, and then we might have $\|Du_\alpha - Df\|_{\mathcal{M}} \approx \|Du_\alpha\|_{\mathcal{M}} + \|Df\|_{\mathcal{M}} > \|Df_0 - Df\|_{\mathcal{M}} \approx 0$. This destroys all hope of verifying the conditions of Lemma 12 in the general case. If we however modify the set of test functions in the definition of $L^1_\eta\nabla$ to be discrete we can prove this bound. Alternatively, we could assume uniformly bounded divergence from the family of test functions. We have left this case out for simplicity, and prove our results for general L^1 costs with finite-dimensional Z.

Lemma 13 *(Conditions for L^1_η cost). Suppose Z is a reflexive Banach space, and $K_0 : X \to Z$ is linear and bounded, and satisfies (8). Then, whenever (44) holds, $F(u) := F_{L^1_\eta}(K_0u)$ is $(K,1)$-co-coercive and (50) holds for some $\bar{\bar\alpha} \in \operatorname{int}\mathcal{P}_\alpha$ with both $u_{\bar{\bar\alpha}}$ and $\bar f$ being $(K,1)$-minimal.*

Proof. Denote

$$B := \{\lambda \in Z^* \mid \|\lambda\|_{Z^*} \leq 1\}.$$

We first verify $(I,1)$-co-coercivity of $F_{L^1_\eta}$. Let $v, v^* \in Z$, and $\lambda^* \in B$ be such that $\lambda^* \in \partial F_{L^1_\eta}(v^*)$. Clearly λ^* achieves the maximum for $F_{L^1_\eta}(v^*)$. Let $\lambda \in B$ achieve the maximum for $F_{L^1_\eta}(v)$. Then

$$\begin{aligned} F_{L^1_\eta}(v) - F_{L^1_\eta}(v^*) &\leq (\lambda|v - f_0) - (\lambda|v^* - f_0) \\ &= (\lambda^*|v - v^*) + (\lambda - \lambda^*|v - v^*) \\ &\leq (\lambda^*|v - v^*) + 2 \sup_{\lambda' \in B} \|\lambda'\| \|v - v^*\| \\ &\leq (\lambda^*|v - v^*) + 2\|v - v^*\|. \end{aligned} \tag{51}$$

This proves $(I,1)$-co-coercivity of $F_{L^1_\eta}$. $(K,1)$-co-coercivity of F now follows similarly to the argument in the proof of Theorem 1, using (8).

Analogously, taking the triangle inequality in (51) in the opposite direction, we show that every $u \in X$ is $(K,1)$-minimal. Therefore, in particular both $\bar{f}$ and $u_{\bar{\alpha}}$ are $(K,1)$-minimal

To verify (50), it is enough to verify (46). Similarly to the proof of Theorem 1, using (8), we verify that $K^*\varphi_{s\bar{\alpha}} \in \partial F(u_{s\bar{\alpha}})$ exists, and

$$\varphi_{s\bar{\alpha}} \in (K^\dagger)^* \partial F(u_{s\bar{\alpha}}) = (K_0 K^\dagger)^* \lambda_{s\bar{\alpha}},$$

where $\lambda_{s\bar{\alpha}} \in B$ achieves the maximum for $F_0(K_0 u_{s\bar{\alpha}} - f_0)$. In fact, $\varphi_{s\bar{\alpha}} \in \mathcal{R}(K)$. If this would not hold, we could find $v \perp \mathcal{R}(K)$ such that

$$0 < (v|\varphi_{s\bar{\alpha}}) = (K_0 K^\dagger v|\lambda_{s\bar{\alpha}}) \leq \|K_0 K^\dagger v\| \|\lambda_{s\bar{\alpha}}\| \leq C\|KK^\dagger v\| \|\lambda_{s\bar{\alpha}}\|.$$

But, for any $v' \in \mathcal{R}(K)$,

$$(v'|KK^\dagger v) = ((KK^\dagger)^* v'|v) = (v'|v) = 0.$$

Therefore $\|KK^\dagger v\| = 0$, and we reach a contradiction unless $\lambda_{s\bar{\alpha}} = 0$, that is $\varphi_{s\bar{\alpha}} = 0 \in \mathcal{R}(K)$.

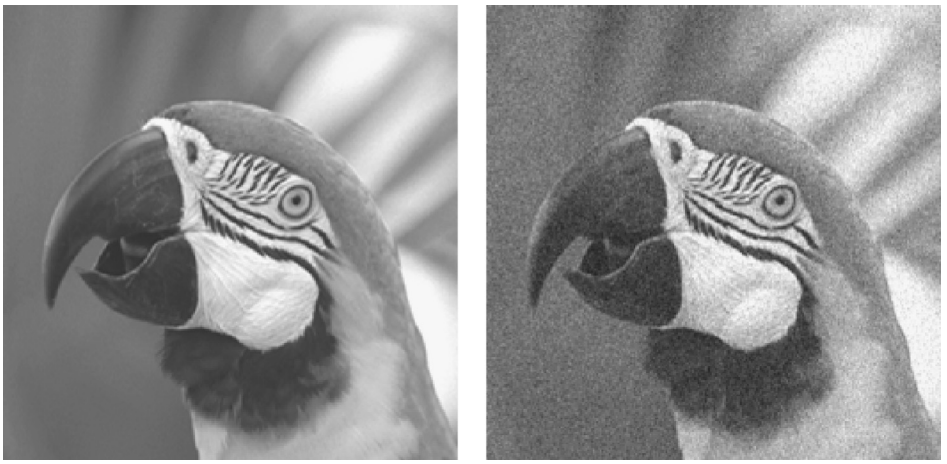
As is easily verified, $\partial F_{L^1_\eta}$ is outer semicontinuous with respect to strong convergence in the domain Z and weak* convergence in the codomain Z^*. That is, given $v^k \to v$ and $z^k \overset{*}{\rightharpoonup} z$ with $z^k \in \partial F_{L^1_\eta}(u^k)$, we have $z \in \partial F_{L^1_\eta}(v)$. By Lemma 8 and (8), we have $K_0 u^k \to K_0 \bar{f}$ strongly in Z. Since B is bounded, we may therefore find a sequence $s^k \searrow 0$ with $\lambda_{s^k\bar{\alpha}} \overset{*}{\rightharpoonup} \lambda_0 \in \partial F(\bar{f})$ weakly* in Z^*. Since by assumption K_0 is compact, then also K_0^* is compact [40, Theorem 4.19]. Consequently $\varphi_{s^k\bar{\alpha}} \to \varphi_0 := (K_0 K^\dagger)^* \lambda_0$ strongly in Y after possibly moving to an unrelabelled subsequence. Let us now consider the right hand side of (50) for $\bar{\bar{\alpha}} = s^k\bar{\alpha}$. Since $f = K\bar{f}$, and we have proved that $\varphi_{u_{s^k\bar{\alpha}}} \in \mathcal{R}(K)$, Lemma 2 shows that

$$\lim_{k\to 0} T(f - t\varphi_{u_{s^k\bar{\alpha}}}; \bar{\alpha}) = T(f - t\varphi_0; \bar{\alpha}).$$

Minding the discussion surrounding (46), we observe that choosing $\bar{\bar{\alpha}} = s^k\bar{\alpha}$ for large enough $k > 0$, (46) is implied by (44). □

Proof of Theorem 2. From the proof of Lemma 13, we observe that (44), can be expanded as

$$T(f; \bar{\alpha}) > T(f - t(K_0 K^\dagger)^* \lambda_0; \bar{\alpha}),$$



(a) Original image **(b)** Noisy image

Fig. 1. Parrot test image.

where $\lambda_0 \in V$ with $\lambda_0 \in \partial F_{L^1_\eta}(K_0\bar{f})$. As in the proof of Theorem 1, this is in fact independent of the choice of $\bar{f}$, so may replace $\bar{f} = K^\dagger f$. Thus $\lambda_0 \in \partial F_{L^1_\eta}(K_0 K^\dagger f)$. By Lemma 13, the conditions of Proposition 6 are satisfied, so we may apply it together with Proposition 2 to conclude the proof. □

Remark 12. The considerations of Remark 10 also apply to Lemma 12 and consequently Theorem 2. That is, the results hold for the cost

$$\Phi(v) = \frac{1}{2}\|v\|^2_{\hat{Y}} - (f|v)_{Y^*,Y}, \quad (v \in Y), \tag{52}$$

where $\hat{Y} \supset Y$ is a Hilbert space, $f \in Y^*$, and Y a reflexive Banach space. Indeed, again the specific form of Φ was only used for the optimality condition (49), which is also satisfied by the form (52).

6. Numerical verification and insight

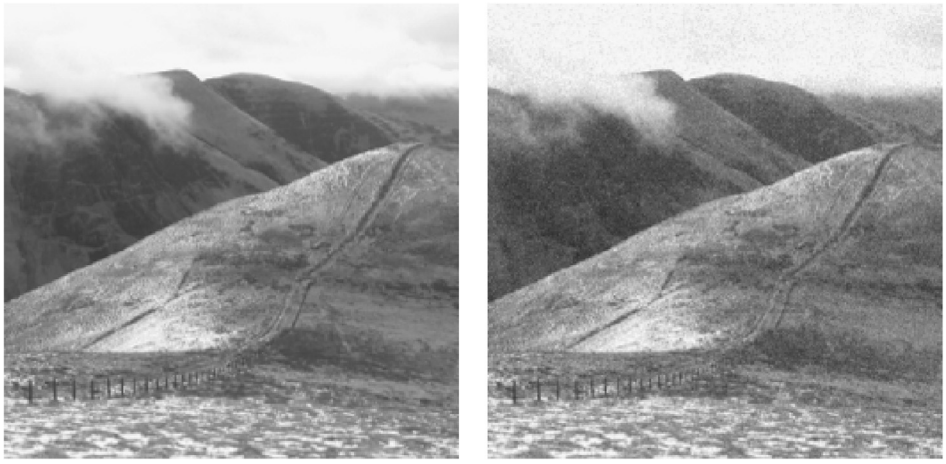
In order to verify the above theoretical results, and to gain further insight into the cost map $\mathcal{I}_{\gamma,\epsilon}$, we computed the values for a grid of values of $\vec{\alpha}$, for both TV and TGV^2 denoising, and L^2_2 and $L^1_\eta\nabla$ cost functionals. This we did for two different images, the parrot image depicted in Fig. 1 and the Scottish southern uplands image depicted in Fig. 2. The results are visualised in Figs. 3 and 4, respectively. For TV, the parameter range was

$$\alpha \in U := \{0.001, 0.01, 0.02, \ldots 0.5\}/n$$

(altogether 51 values), where $n = 256$ is the edge length of the rectangular test image. For TGV^2 the parameter range was $\vec{\alpha} \in U \times (U/n)$. We set $\gamma = 100$, $\epsilon = 1\text{E}^{-10}$, and computed the denoised image $u_{\alpha,\gamma,\epsilon}$ by the SSN denoising algorithm that we report separately in [24] with more extensive numerical comparisons and further applications.

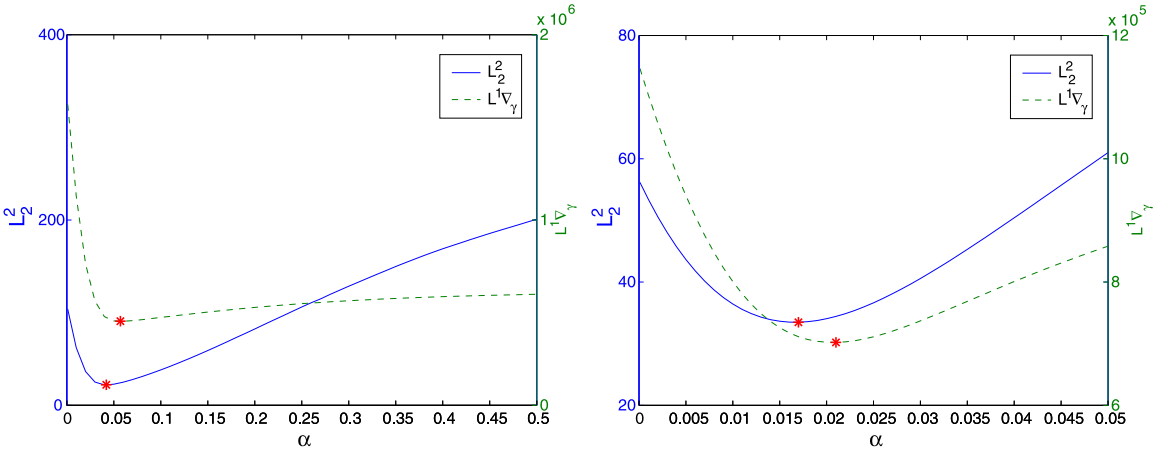
As we can see, the optimal $\vec{\alpha}$ clearly seems to lie away from the boundary of the parameter domain $\mathcal{P}_\alpha$, confirming the theoretical studies for the squared L^2 cost L^2_2, and the discrete version of the Huberised TV cost $L^1_\eta\nabla$. The question remains: do these results hold for the full Huberised TV?

We further observe from the numerical landscapes that the cost map $\mathcal{I}_{\gamma,\epsilon}$ is roughly quasiconvex in the variable α for both TV and TGV^2. In the β variable of TGV^2 the same does not seem to hold, as around



(a) Original image **(b)** Noisy image

Fig. 2. Uplands test image.



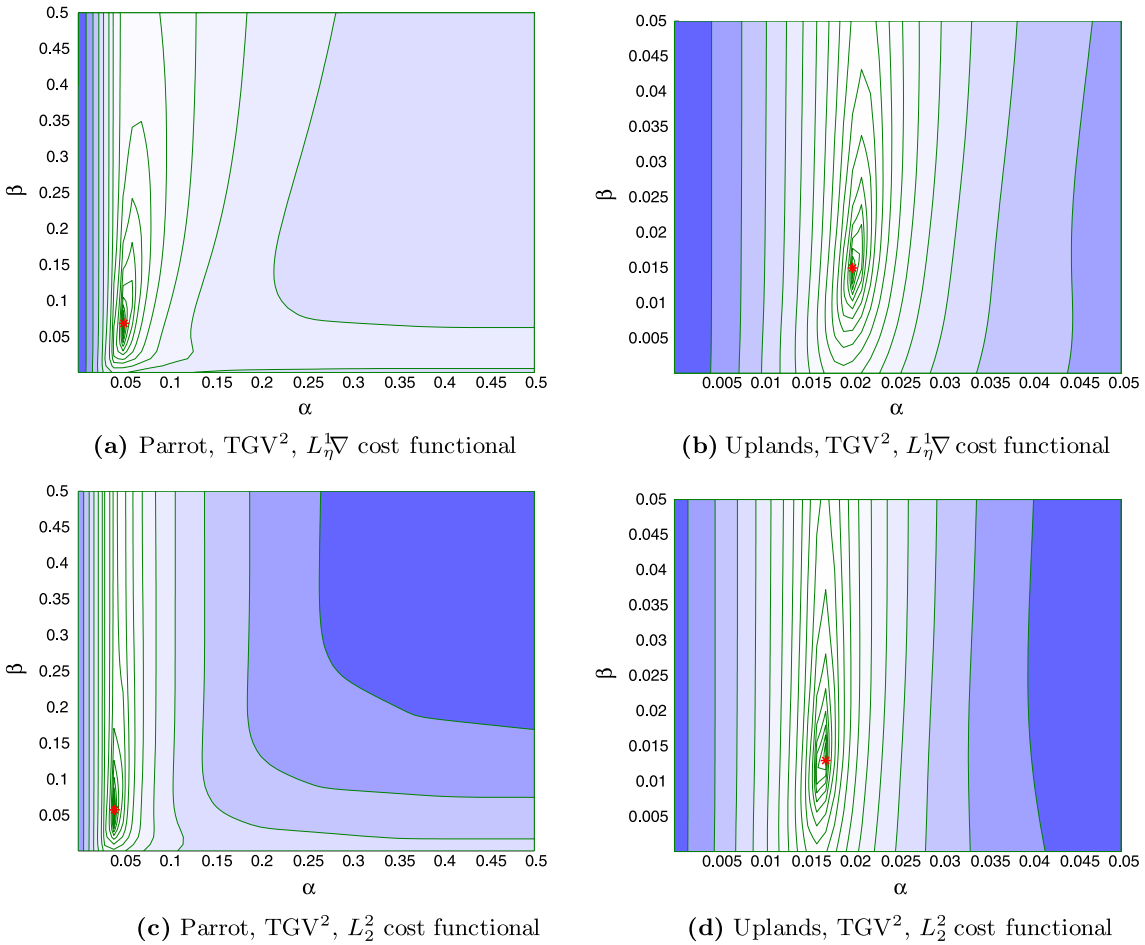
(a) Parrot, TV, all cost functionals **(b)** Uplands, TV, all cost functionals

Fig. 3. Cost functional value versus α for TV denoising, for both the parrot and uplands test images, for both L_2^2 and $L_\eta^1\nabla$ cost functionals. For the parrot image, optimal α from [24] for the initialisation $0.1/n$, resp. $(1/n^2, 0.1/n)$, is indicated by an asterisk. For the landscape image, optimal α from [24] for the initialisation $0.01/n$, resp. $(0.1/n^2, 0.01/n)$, is indicated by an asterisk.

the optimal solution the level sets tend to expand along α as β increases, until starting to reach their limit along β. However, the level sets around the optimal solution also tend to be very elongated on the β axes. This suggests that TGV^2 is reasonably robust with respect to choice of β, as long as it is in the right range.

7. A data statement for the EPSRC

This is a theoretical mathematics paper, and any data used merely serves as a demonstration of mathematically proven results. Moreover, photographs that are for all intents and purposes statistically comparable to the ones used for the final experiments, can easily be produced with a digital camera, or downloaded from the internet. This will provide a better evaluation of the results than the use of exactly the same data as we used.



(a) Parrot, TGV^2, $L^1_\eta\nabla$ cost functional

(b) Uplands, TGV^2, $L^1_\eta\nabla$ cost functional

(c) Parrot, TGV^2, L^2_2 cost functional

(d) Uplands, TGV^2, L^2_2 cost functional

Fig. 4. Cost functional value versus α for TGV^2 denoising, for both the parrot and uplands test images, for both L^2_2 and $L^1_\eta\nabla$ cost functionals. The illustrations are contour plots of function value versus $\vec{\alpha} = (\beta, \alpha)$. For the parrot image, optimal $\vec{\alpha}$ from [24] for the initialisation $0.1/n$, resp. $(1/n^2, 0.1/n)$, is indicated by an asterisk. For the landscape image, optimal $\vec{\alpha}$ from [24] for the initialisation $0.01/n$, resp. $(0.1/n^2, 0.01/n)$, is indicated by an asterisk.

Acknowledgments

In Cambridge, this project has been supported by King Abdullah University of Science and Technology (KAUST) Award No. KUK-I1-007-43, EPSRC grants Nr. EP/J009539/1 “Sparse & Higher-order Image Restoration”, and Nr. EP/M00483X/1 “Efficient computational tools for inverse imaging problems”. In Quito, the project has been supported by the Escuela Politécnica Nacional de Quito under award PIS 12-14 and the MATHAmSud project SOCDE “Sparse Optimal Control of Differential Equations”. When in Quito, T. Valkonen was moreover supported by a Prometeo scholarship of the Senescyt (Ecuadorian Ministry of Science, Technology, Education, and Innovation).

References

[1] B. Adcock, A.C. Hansen, B. Roman, G. Teschke, Generalized sampling: stable reconstructions, inverse problems and compressed sensing over the continuum, Adv. Imaging Electron Phys. 182 (2014) 187–279.
[2] L. Ambrosio, N. Fusco, D. Pallara, Functions of Bounded Variation and Free Discontinuity Problems, Oxford University Press, 2000.
[3] V. Barbu, Analysis and Control of Nonlinear Infinite Dimensional Systems, Academic Press, New York, 1993.
[4] H. Bauschke, P. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, CMS Books Math., Springer, ISBN 9781441994677, 2011.

[5] M. Benning, L. Gladden, D. Holland, C.-B. Schönlieb, T. Valkonen, Phase reconstruction from velocity-encoded MRI measurements—a survey of sparsity-promoting variational approaches, J. Magn. Reson. 238 (2014) 26–43.
[6] M. Bergounioux, Optimal control of problems governed by abstract elliptic variational inequalities with state constraints, SIAM J. Control Optim. 36 (1) (1998) 273–289, http://dx.doi.org/10.1137/S0363012996302615.
[7] M. Bergounioux, F. Mignot, Optimal control of obstacle problems: existence of Lagrange multipliers, ESAIM Control Optim. Calc. Var. 5 (2000) 45–70 (electronic).
[8] L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, L. Tenorio, B. van Bloemen Waanders, K. Willcox, Y. Marzouk, Large-Scale Inverse Problems and Quantification of Uncertainty, vol. 712, John Wiley & Sons, 2011.
[9] A. Braides, Gamma-Convergence for Beginners, Oxford Lecture Ser. Math. Appl., Oxford University Press, ISBN 9780198507840, 2002.
[10] K. Bredies, M. Holler, Regularization of linear inverse problems with total generalized variation, J. Inverse Ill-Posed Probl. 6 (22) (2014) 871–913, http://dx.doi.org/10.1515/jip-2013-0068.
[11] K. Bredies, K. Kunisch, T. Pock, Total generalized variation, SIAM J. Imaging Sci. 3 (2011) 492–526, http://dx.doi.org/10.1137/090769521.
[12] K. Bredies, T. Valkonen, Inverse problems with second-order total generalized variation constraints, in: Proceedings of the 9th International Conference on Sampling Theory and Applications, SampTA, 2011, Singapore, 2011, http://iki.fi/tuomov/mathematics/SampTA2011.pdf.
[13] T. Bui-Thanh, K. Willcox, O. Ghattas, Model reduction for large-scale systems with high-dimensional parametric input space, SIAM J. Sci. Comput. 30 (6) (2008) 3270–3288.
[14] L. Calatroni, J.C. de los Reyes, C.-B. Schönlieb, Dynamic sampling schemes for optimal noise learning under multiple nonsmooth constraints, in: System Modeling and Optimization, Springer-Verlag, 2014, pp. 85–95.
[15] A. Chambolle, P.-L. Lions, Image recovery via total variation minimization and related problems, Numer. Math. 76 (1997) 167–188, ISSN 0029-599X.
[16] G. Chavent, Nonlinear Least Squares for Inverse Problems: Theoretical Foundations and Step-by-Step Guide for Applications, Springer Science & Business Media, 2010.
[17] Y. Chen, T. Pock, H. Bischof, Learning ℓ_1-based analysis and synthesis sparsity priors using bi-level optimization, in: Workshop on Analysis Operator Learning vs. Dictionary Learning, NIPS 2012, 2012.
[18] Y. Chen, T. Pock, R. Ranftl, H. Bischof, Revisiting loss-specific training of filter-based MRFs for image restoration, in: 35th German Conference on Pattern Recognition, GCPR, 2013.
[19] Y. Chen, R. Ranftl, T. Pock, Insights into analysis operator learning: from patch-based sparse models to higher order MRFs, IEEE Trans. Image Process. 23 (3) (2014) 1060–1072.
[20] J. Chung, M.I. Español, T. Nguyen, Optimal regularization parameters for general-form Tikhonov regularization, arXiv:1407.1911.
[21] G. Dal Maso, An Introduction to Γ-Convergence, Progress in Nonlinear Differential Equations and Their Applications, Birkhäuser, Boston, ISBN 9780817636791, 1993.
[22] J.C. de los Reyes, On the optimization of steady Bingham flow in pipes, in: Recent Advances in Optimization and Its Applications in Engineering, Springer, 2010, pp. 379–388.
[23] J.C. de los Reyes, C.-B. Schönlieb, Image denoising: learning the noise model via nonsmooth PDE-constrained optimization, Inverse Probl. Imaging 7 (4) (2013).
[24] J.C. de los Reyes, C.-B. Schönlieb, T. Valkonen, Bilevel parameter learning for higher-order total variation regularisation models, arXiv:1508.07243.
[25] S. Delladio, Lower semicontinuity and continuity of functions of measures with respect to the strict convergence, Proc. Roy. Soc. Edinburgh Sect. A 119 (3–4) (1991) 265–278.
[26] F. Demengel, Fonctions à hessien borné, Ann. Inst. Fourier 34 (2) (1984) 155–190.
[27] S. Dempe, A.B. Zemkoho, KKT reformulation and necessary conditions for optimality in nonsmooth bilevel optimization, SIAM J. Optim. 24 (4) (2014) 1639–1669, http://dx.doi.org/10.1137/130917715.
[28] J. Domke, Generic methods for optimization-based modeling, in: International Conference on Artificial Intelligence and Statistics, 2012, pp. 318–326.
[29] J. Domke, Learning graphical model parameters with approximate marginal inference, arXiv:1301.3193.
[30] I. Ekeland, R. Temam, Convex Analysis and Variational Problems, SIAM, 1999.
[31] J. Fehrenbach, M. Nikolova, G. Steidl, P. Weiss, Bilevel image denoising using Gaussianity tests, in: Scale Space and Variational Methods in Computer Vision, in: Lecture Notes in Comput. Sci., vol. 9087, 2015, pp. 117–128.
[32] E. Haber, L. Horesh, L. Tenorio, Numerical methods for experimental design of large-scale linear ill-posed inverse problems, Inverse Probl. 24 (5) (2008) 055012, http://dx.doi.org/10.1088/0266-5611/24/5/055012.
[33] E. Haber, L. Horesh, L. Tenorio, Numerical methods for the design of large-scale nonlinear discrete ill-posed inverse problems, Inverse Probl. 26 (2) (2010) 025002.
[34] E. Haber, L. Tenorio, Learning regularization functionals—a supervised training approach, Inverse Probl. 19 (3) (2003) 611.
[35] M. Hintermüller, G. Stadler, An infeasible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration, SIAM J. Sci. Comput. 28 (1) (2006) 1–23.
[36] J. Kristensen, F. Rindler, Relaxation of signed integral functionals in BV, Calc. Var. Partial Differential Equations 37 (1–2) (2010) 29–62, http://dx.doi.org/10.1007/s00526-009-0250-5, ISSN 0944-2669.
[37] K. Kunisch, T. Pock, A bilevel optimization approach for parameter learning in variational models, SIAM J. Imaging Sci. 6 (2) (2013) 938–983.
[38] F. Rindler, G. Shaw, Strictly continuous extensions of functionals with linear growth to the space BV, preprint, 2013.
[39] R.T. Rockafellar, R.J.-B. Wets, Variational Analysis, Springer, ISBN 3540627723, 1998.
[40] W. Rudin, Functional Analysis, Int. Ser. Monogr. Pure Appl. Math., McGraw-Hill, ISBN 9780070619883, 2006.

[41] M.F. Tappen, Utilizing variational optimization to learn Markov random fields, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007, CVPR'07, IEEE, 2007, pp. 1–8.
[42] M.F. Tappen, C. Liu, E.H. Adelson, W.T. Freeman, Learning Gaussian conditional random fields for low-level vision, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007, CVPR'07, IEEE, 2007, pp. 1–8.
[43] T. Valkonen, The jump set under geometric regularisation. Part 2: higher-order approaches, arXiv:1407.2334.
[44] T. Valkonen, A primal-dual hybrid gradient method for non-linear operators with applications to MRI, Inverse Probl. 30 (5) (2014) 055012, http://dx.doi.org/10.1088/0266-5611/30/5/055012, http://iki.fi/tuomov/mathematics/nl-pdhgm.pdf.
[45] T. Valkonen, K. Bredies, F. Knoll, Total generalised variation in diffusion tensor imaging, SIAM J. Imaging Sci. 6 (1) (2013) 487–525, http://dx.doi.org/10.1137/120867172, http://iki.fi/tuomov/mathematics/dtireg.pdf.
[46] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612, http://dx.doi.org/10.1109/TIP.2003.819861.