

**Weierstraß-Institut**  
**für Angewandte Analysis und Stochastik**  
**Leibniz-Institut im Forschungsverbund Berlin e. V.**

Preprint

ISSN 2198-5855

**Analysis and optimisation of a variational model for mixed  
Gaussian and Salt & Pepper noise removal**

Luca Calatroni<sup>1</sup>, Kostas Papafitsoros<sup>2</sup>

submitted: October 4, 2018

<sup>1</sup> CMAP, École Polytechnique  
Route de Saclay  
91128 Palaiseau Cedex  
France  
E-Mail: [luca.calatroni@polytechnique.edu](mailto:luca.calatroni@polytechnique.edu)

<sup>2</sup> Weierstrass Institute  
Mohrenstr. 39  
10117 Berlin  
Germany  
E-Mail: [kostas.papafitsoros@wias-berlin.de](mailto:kostas.papafitsoros@wias-berlin.de)

No. 2542  
Berlin 2018



---

2010 *Mathematics Subject Classification.* 94A08, 49K20, 49J20, 49J40, 49Q20, 65J20, 34K29.

*Key words and phrases.* Mixed noise removal, denoising, total variation, exact solutions, Gaussian noise, Salt & Pepper noise, bilevel optimisation, parameter learning.

The authors thank Prof. Juan Carlos De Los Reyes for his useful comments on Section 4. LC acknowledges the joint ANR/FWF Project “Efficient Algorithms for Nonsmooth Optimization in Imaging” (EANOI) FWF n. I1148 / ANR-12-IS01-0003, the Fondation Mathématique Jacques Hadamard (FMJH) and the RISE EU project NoMADS. KP acknowledges the support of the Einstein Foundation Berlin within the ECMath project CH12. Both authors would like to thank the Isaac Newton Institute for Mathematical Sciences for support and hospitality during the program “Variational Methods and Effective Algorithms for Imaging and Vision” when work on this paper was undertaken. This work was supported by EPSRC grant number EP/K032208/1.

Edited by  
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)  
Leibniz-Institut im Forschungsverbund Berlin e. V.  
Mohrenstraße 39  
10117 Berlin  
Germany

Fax: +49 30 20372-303  
E-Mail: [preprint@wias-berlin.de](mailto:preprint@wias-berlin.de)  
World Wide Web: <http://www.wias-berlin.de/>

# Analysis and optimisation of a variational model for mixed Gaussian and Salt & Pepper noise removal

Luca Calatroni, Kostas Papafitsoros

## Abstract

We analyse a variational regularisation problem for mixed noise removal that was recently proposed in [14]. The data discrepancy term of the model combines  $L^1$  and  $L^2$  terms in an infimal convolution fashion and it is appropriate for the joint removal of Gaussian and Salt & Pepper noise. In this work we perform a finer analysis of the model which emphasises on the balancing effect of the two parameters appearing in the discrepancy term. Namely, we study the asymptotic behaviour of the model for large and small values of these parameters and we compare it to the corresponding variational models with  $L^1$  and  $L^2$  data fidelity. Furthermore, we compute exact solutions for simple data functions taking the total variation as regulariser. Using these theoretical results, we then analytically study a bilevel optimisation strategy for automatically selecting the parameters of the model by means of a training set. Finally, we report some numerical results on the selection of the optimal noise model via such strategy which confirm the validity of our analysis and the use of popular data models in the case of “blind” model selection.

## 1 Introduction

Image denoising is a classical problem in imaging which is defined as the task of removing oscillations and interferences from a given image  $f$ . Given a regular image domain  $\Omega \subseteq \mathbb{R}^2$  the problem can be formulated mathematically as the task of retrieving a noise-free image  $u$  from its noisy measurement  $f$ , the latter being the result of a (possibly non-linear) degradation operator  $\mathcal{T}$ . In its general form, this problem can be written in the following way

$$\text{find } u \quad \text{such that} \quad f = \mathcal{T}(u).$$

Here the operator  $\mathcal{T}$  introduces noise in the image, not only in an additive way, and it is not to be confused with the forward operator in the context of inverse problems. In order to obtain a noise-free image, a classical technique consists of minimising an appropriate energy functional  $\mathcal{J}$  over a suitable Banach space  $X$  where the image functions are assumed to lie. In its general form the problem reads as follows:

$$\min_{u \in X} \{ \mathcal{J}(u) := R(u) + \lambda \Phi(u, f) \}. \quad (1.1)$$

Here,  $R(u)$  stands for the regularisation term encoding *a priori* information on the regularity of the solution,  $\Phi(u, f)$  for the data-fitting measure that depends on the statistical and physical assumptions in the data and  $\lambda > 0$  is a scalar parameter whose magnitude balances the regularisation against trust in the data. Since the seminal work of Rudin, Osher, Fatemi [52], a popular choice for  $R$  in (1.1) is  $R(u) = |Du|(\Omega)$ , the Total Variation (TV) seminorm [4], due to its ability of preserving salient structures in the image, i.e., edges, while removing noise at the same time. Here  $Du$  represents the distributional derivative of the function  $u \in \text{BV}(\Omega)$ , the space of functions of bounded variation, and

$|Du|(\Omega)$  is the total variation of this measure. In the recent years, also higher-order regularisation terms that improve upon TV-induced artefacts – notably, the creation of piecewise constant structures – have been proposed in the literature. Among those, the most prominent is the Total Generalized Variation (TGV) [9], see also [47] for a comprehensive review.

In this work we will mostly focus on the standard TV regularisation energy, as our work focuses on the choice of  $\Phi$  rather than of  $R$ . Classical data fidelity terms for denoising images with Gaussian or impulsive Salt & Pepper noise are based on the use of the  $L^2$  and  $L^1$  norm, respectively, i.e.,

$$\Phi_{L^2}(u, f) := \frac{1}{2} \int_{\Omega} (f - u)^2 dx \quad \text{or} \quad \Phi_{L^1}(u, f) := \int_{\Omega} |f - u| dx.$$

These discrepancies are statistically consistent with the assumptions on the noise since they can be derived as the MAP estimators of the underlying likelihood function [7].

There exists a considerable amount of work in the literature regarding the structures of solutions of variational problems with pure  $L^1$  or  $L^2$  fidelity terms. As a result, the differences between the effects that these terms have on the solutions of the corresponding denoising models are well understood. See for instance [1, 2, 3, 16, 17, 18, 20, 28, 36, 43, 51, 53] for TV regularisation and [10, 48, 49, 50, 54] for TGV. For example, in the case of TV regularisation, it is known that the use of  $L^2$  fidelity does not introduce new discontinuities in the solution, which is not the case for the  $L^1$  fidelity. Moreover, the  $L^1$  model is capable of exact data recovery, in contrast to the  $L^2$  one, where always some loss of contrast occurs.

## 1.1 Image denoising for noise mixtures

Due to different image acquisition and transmission faults, the given image  $f$  may be corrupted by a mixture of noise statistics. This is typical, for instance, whenever the presence of noise is due to electronic faults and/or photon-counting processes (such as in astronomy and microscopy applications) combined with actual damages in the receiving sensors, resulting in a lack of information transmittance in only a few image pixels (“burned pixels”). The modelling of  $\Phi$  in (1.1) is therefore expected to encode such noise combination. In this work we consider the special case when Gaussian and sparse Salt & Pepper noise are present in the data.

Several authors have considered previously such noise combination. In [32, 41], for instance, a combined model with a  $L^1 + L^2$  data fidelity and TV regularisation is considered for the joint removal of impulsive and Gaussian noise. Another approach is considered in [12] where two sequential steps having  $L^1$  and  $L^2$  as data fidelity are performed to remove the impulsive and the Gaussian component of the mixed noise, respectively. Framelet-based approaches combining  $L^1$  and  $L^2$  data fidelities in a discrete setting have also been proposed in [27, 55]. However, despite the observed good practical performance of the models described above, the use of an additive and sequential combination of  $L^1$  and  $L^2$  data fitting terms lacks a rigorous statistical interpretation in terms, for instance, of a MAP estimation.

We remark that the combination of other noise distributions such as, for instance, Gaussian and Poisson is also frequent in astronomy and microscopy applications and have been studied in several works such as, for instance, [8, 37, 42]. Such noise mixtures, however, are outside the scope of this work.

## 1.2 Infimal convolution modelling of data discrepancies

Recently in [14], a non-standard variational model for noise removal of mixtures of Salt & Pepper and Gaussian, and Gaussian and Poisson noise has been studied. The model, which will be referred to as TV–IC model, is based on the minimisation of an energy functional which is the sum of  $|Du|(\Omega)$  and an infimal convolution of single noise data discrepancy terms. Given two positive parameters  $\lambda_1, \lambda_2$  it reads:

$$\min_{u \in \text{BV}(\Omega)} |Du|(\Omega) + \Phi^{\lambda_1, \lambda_2}(u, f),$$

where the data fidelity  $\Phi^{\lambda_1, \lambda_2}(u, f)$  is defined as

$$\Phi^{\lambda_1, \lambda_2}(u, f) := \inf_v \lambda_1 \Phi_1(v) + \lambda_2 \Phi_2(v, f - u). \quad (1.2)$$

Here,  $\Phi_1, \Phi_2$  denote standard data fidelity terms typically used for single noise removal such as the  $L^1, L^2$  norm and the Kullback-Leibler functional. In the particular case of a mixture of Salt & Pepper and Gaussian noise, (1.2) specifies into  $\Phi_1(v) = \|v\|_{L^1(\Omega)}$  and  $\Phi_2(v, f - u) = \frac{1}{2} \|f - u - v\|_{L^2(\Omega)}^2$ . In this case, the minimisation in (1.2) is done over  $L^1(\Omega)$  and  $\Phi^{\lambda_1, \lambda_2}(u, f)$  reads

$$\Phi^{\lambda_1, \lambda_2}(u, f) = \min_{v \in L^1(\Omega)} \lambda_1 \|v\|_{L^1(\Omega)} + \frac{\lambda_2}{2} \|f - u - v\|_{L^2(\Omega)}^2, \quad f, u \in L^1(\Omega). \quad (1.3)$$

It can be easily checked that the minimisation in (1.3) is indeed well-defined. The reader should not be alerted by the fact that the  $L^1$  functions  $f, u$  appear in the  $L^2$  part of (1.3), as the variable  $v$  takes care of any non-integrability issue, see Proposition 2.1. In fact, as we are going to remark in the same proposition, the functional  $\Phi^{\lambda_1, \lambda_2}$  can be written equivalently as

$$\Phi^{\lambda_1, \lambda_2}(u, f) = \int_{\Omega} \varphi(f(x) - u(x)) dx,$$

where  $\varphi$  is the well-known Huber-regularisation of the absolute value function. As a consequence, the functional  $\Phi^{\lambda_1, \lambda_2}$  can be simply seen as a Huberised  $L^1$  norm.

In [14], it was shown that the data discrepancy (1.3) corresponds to the joint MAP estimator for a denoising problem featuring a mixture of Laplace and Gaussian noise distributions. The effectiveness of the model for the removal of such noise mixture as well as the additional property of decomposing the noise into its sparse (Salt & Pepper) and distributed (Gaussian) component was there confirmed with extended numerical examples.

Note that a Huber smoothing of the  $L^1$  norm has previously been considered in order to apply fast second order minimisation algorithms such as semismooth Newton method in [33]. Also, in the purely discrete setting, smoothed TV– $L^1$  models have been studied in [45] for exact histogram specification. Similar models (among which also Huber-type) were also considered in [5], where the authors obtained bounds on the infinity norm of the difference between data and solutions.

**Our contribution.** In this work we examine in depth the similarities and the differences between the TV–IC model and the pure TV– $L^1, \text{TV–}L^2$  ones. We first provide detailed asymptotic results as  $\lambda_1$  or  $\lambda_2$  tend either to infinity or to zero and describe how the solution of the model varies in these cases. Note that these results are proved for a general regularisation term. Secondly, in order to have a better insight on the type of solutions one can expect, we do a fine scale analysis of the one-dimensional TV regularised model by computing exact solutions for simple data functions  $f$ . Up to our knowledge, this is the first time that the effect of the Huberised  $L^1$  fidelity term is studied in the continuous setting.

In the second part of the paper, we focus on the optimal selection of the parameters  $\lambda_1, \lambda_2$  appearing in (1.3). In order to do that, we consider a bilevel optimisation approach [13, 24] which in its general formulation reads

$$\begin{aligned} & \min_{\lambda_1, \lambda_2 \geq 0} F(u_{\lambda_1, \lambda_2}) \\ \text{subject to } & u_{\lambda_1, \lambda_2} \in \operatorname{argmin}_{u \in \operatorname{BV}(\Omega)} |Du|(\Omega) + \Phi^{\lambda_1, \lambda_2}(u, f). \end{aligned} \quad (1.4)$$

Here  $F$  denotes a cost functional which measures how far the solution  $u_{\lambda_1, \lambda_2}$  is from some ground truth (training) image. The parameters  $\lambda_1, \lambda_2$  selected within this framework are therefore those producing the closest reconstruction  $u_{\lambda_1, \lambda_2}$  to the training image, see Section 4.1 for more details. We perform a rigorous analysis on the existence of solutions of (1.4) as well as a regularised version of it, proving the Fréchet differentiability of the solution map  $\mathcal{S} : (\lambda_1, \lambda_2) \mapsto u_{\lambda_1, \lambda_2}$  and the existence of adjoint states. This allows to derive a handy characterisation of the gradient of the reduced form of  $F$  which can be used for efficient numerical implementations. Our analysis justifies rigorously the formal Lagrangian approach considered in [14, Section 7].

We conclude our study with some numerical experiments connecting the analysis on the structure of solutions discussed above with the problem of learning the optimal noise model for a given noisy image with unknown noise intensity. Our numerical findings show that in case of pure Salt & Pepper and Gaussian denoising, the bilevel optimisation approach applied to the TV–IC model computes optimal parameters which enforces pure  $L^1$  fidelity. In the case of noise mixture, a combination of  $L^1$  and  $L^2$  data fitting is preferred. Interestingly, in the case of pure Gaussian, it is not the pure  $L^2$  data fitting that is selected but still a combination of  $L^1$  and  $L^2$ , indicating the benefit of the use of  $L^1$  discrepancy even in the case of Gaussian noise.

We emphasise that the two parts of the paper are intrinsically connected, see for instance Propositions 4.2 and 4.3. In the former, by making use of the analytical results of the first part, we show with a help of a counterexample, that in order to show existence of solutions for the bilevel optimisation problem (1.4), it is necessary to enforce an upper bound on the parameters  $\lambda_1, \lambda_2$ . The existence of solutions in this case, is shown in Proposition 4.3 also by making use of the results of the first part of the paper.

Overall, this study motivates further the use of the TV–IC model and in general the use of the infimal convolution based fidelity term, by (i) describing the structure of the expected solutions and (ii) by proposing an automated parameter selection strategy making this model more flexible and applicable to mixed denoising problems.

## 2 Analysis of the $L^1$ – $L^2$ IC model: characterisation and asymptotics

We start this section by observing that the  $L^1$ – $L^2$  infimal convolution term can be equivalently formulated as a Huberised  $L^1$  norm. This provides an interesting motivation on its effectiveness in the removal of mixed Salt & Pepper and Gaussian noise. As usual,  $\Omega \subseteq \mathbb{R}^d$  denotes an open, bounded, connected domain with Lipschitz boundary.

**Proposition 2.1.** *Let  $\lambda_1, \lambda_2 > 0$ ,  $f, u \in L^1(\Omega)$  and consider the data fitting term,  $\Phi^{\lambda_1, \lambda_2}(u, f)$  defined in (1.3). Then*

$$\Phi^{\lambda_1, \lambda_2}(u, f) = \int_{\Omega} \varphi(f(x) - u(x)) dx, \quad (2.1)$$

where for  $t \in \mathbb{R}$

$$\varphi(t) = \begin{cases} \lambda_1 |t| - \frac{\lambda_1^2}{2\lambda_2}, & \text{if } |t| \geq \frac{\lambda_1}{\lambda_2}, \\ \frac{\lambda_2}{2} |t|^2, & \text{if } |t| < \frac{\lambda_1}{\lambda_2}. \end{cases} \quad (2.2)$$

*Proof.* The proof is straightforward, in view of

$$\Phi^{\lambda_1, \lambda_2}(u, f) = \min_{v \in L^1(\Omega)} \int_{\Omega} \left( \lambda_1 |v(x)| + \frac{\lambda_2}{2} (f(x) - u(x) - v(x))^2 \right) dx,$$

and noticing that the minimisation in the definition of  $\Phi^{\lambda_1, \lambda_2}$  can be considered pointwise. Immediate calculations show that the optimal  $v$  can be computed explicitly as

$$v_{opt}(x) = \begin{cases} f(x) - u(x) - \frac{\lambda_1}{\lambda_2} \frac{f(x) - u(x)}{|f(x) - u(x)|}, & \text{if } |f(x) - u(x)| \geq \frac{\lambda_1}{\lambda_2}, \\ 0, & \text{if } |f(x) - u(x)| < \frac{\lambda_1}{\lambda_2}. \end{cases} \quad (2.3)$$

□

From the formulation (2.1)–(2.2) one sees that the infimal convolution of  $L^1$  and  $L^2$  data fidelities coincides with a smoothed  $L^1$  norm, see Figure 1. This is in fact well-known in the context of optimisation in Hilbert spaces [6], where an explicit expression like (2.3) is often used to compute the soft-thresholding operators. Furthermore, let us point out here that Proposition 2.1 above is analogous to a similar result about the Huberised total variation functional, which has also been shown to be equivalent to a corresponding infimal convolution functional involving  $L^1$  and  $L^2$  norms, see [11] and [30].

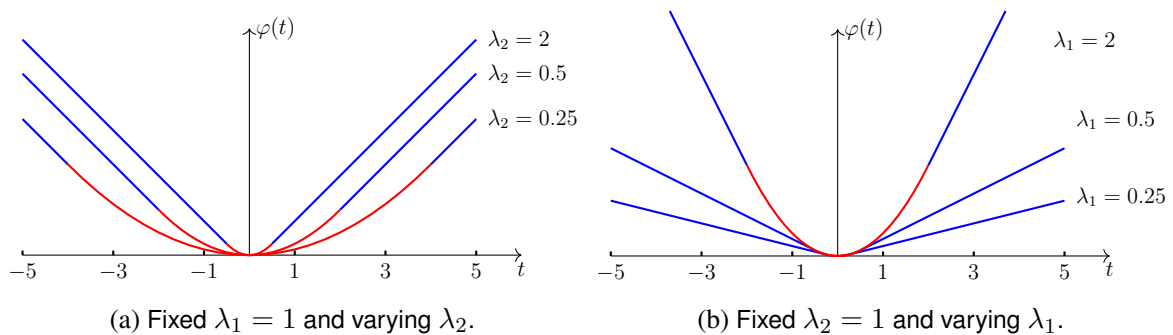


Figure 1: Example plots of the Huber function  $\varphi$ .

The formulation (2.1)–(2.2) of  $\Phi^{\lambda_1, \lambda_2}$  provides an interesting insight on its interpretation and motivates its effectiveness for mixed noise removal. When  $|f(x) - u(x)|$  is large, the noise component is interpreted as Salt & Pepper by the model and then  $\Phi^{\lambda_1, \lambda_2}$  behaves locally as  $\|f - u\|_{L^1(\Omega)}$ . On the other hand, if  $|f(x) - u(x)|$  is small, then the model assumes that the noise is Gaussian and enforces a data fidelity  $\Phi^{\lambda_1, \lambda_2}$  locally equal to  $\|f - u\|_{L^2(\Omega)}^2$ , see Figure 2 for a visualisation.

## 2.1 Asymptotic behaviour

In this section, we investigate the asymptotic behaviour of the  $L^1$ – $L^2$  IC model. Here, we do not need to restrict to the TV regulariser, but we can consider a more general regularisation functional  $J$  with the following properties:

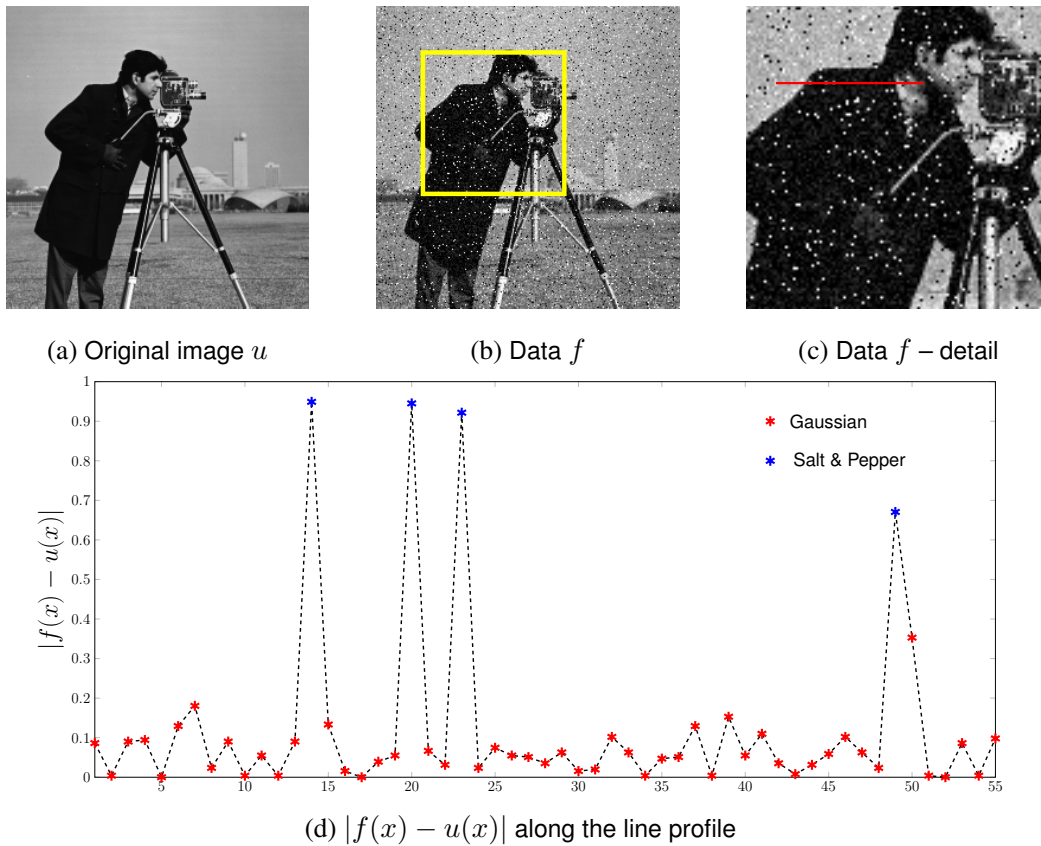


Figure 2: Interpretation of the  $L^1$ – $L^2$  IC term based on the formulation (2.1)–(2.2). First row: original image, noisy version and detail with line profile (red). Second row: difference  $|f - u|$  along the line profile. The fidelity functional  $\Phi^{\lambda_1, \lambda_2}$  locally acts like  $\|f - u\|_{L^1}$  at points with high values of  $|f - u|$  (blue asterisks), thus assuming these points to be corrupted by Salt & Pepper noise. On the other hand,  $\Phi^{\lambda_1, \lambda_2}$  acts locally as  $\|f - u\|_{L^2}^2$  at points with low values of  $|f - u|$  (red asterisks), identifying these points as corrupted by Gaussian noise. **Parameters:** Gaussian variance  $\sigma^2 = 0.005$ , density of pixels corrupted by Salt & Pepper noise  $d = 5\%$ .



- (i)  $J : L^1(\Omega) \rightarrow \mathbb{R} \cup \{\infty\}$  is positive, proper, convex, lower semicontinuous with respect to the strong convergence in  $L^1(\Omega)$ .
- (ii) There exist constants  $C_1, C_2 > 0$  such that  $C_1|Du|(\Omega) \leq J(u) \leq C_2|Du|(\Omega)$  for every  $u \in \text{BV}(\Omega)$ .

Classical regularisers such as TV, Huber-TV and TGV of any order, satisfy the above properties. Note that this is also true for a large class of structural TV-type functionals that are commonly used in inverse problems, see [31].

We are interested in the following general problem:

$$\min_{\substack{u \in \text{BV}(\Omega) \\ v \in L^1(\Omega)}} J(u) + \lambda_1 \|v\|_{L^1(\Omega)} + \frac{\lambda_2}{2} \|f - u - v\|_{L^2(\Omega)}^2, \quad (2.4)$$

which is a more general version of the TV–IC model for Gaussian and Salt & Pepper noise removal

$$\min_{\substack{u \in \text{BV}(\Omega) \\ v \in L^1(\Omega)}} |Du|(\Omega) + \lambda_1 \|v\|_{L^1(\Omega)} + \frac{\lambda_2}{2} \|f - u - v\|_{L^2(\Omega)}^2. \quad (2.5)$$

The well-posedness of (2.5) has been studied in [14]. For the more general model (2.4) existence of minimisers  $(u^*, v^*) \in \text{BV}(\Omega) \times L^1(\Omega)$  follows from a simple application of the direct method of calculus of variations. Note however, that the solution  $u^*$  is not necessary unique as  $\Phi^{\lambda_1, \lambda_2}(\cdot, f)$  is not strictly convex ( $J$  is not necessarily strictly convex either). The same holds for  $v^*$  due to formula (2.3) which connects it with  $u^*$ .

One natural question one may ask is in what degree we can expect to recover the single noise models by sending the parameters  $\lambda_1, \lambda_2$  (or their ratio) to infinity. In the following we answer this question by taking advantage of the formulation (2.1)–(2.2) and using some  $\Gamma$ -convergence arguments. Firstly, we extend a corresponding proposition that was shown in [14, Proposition 5.1], to the general regulariser case, adjusted for our purposes.

**Proposition 2.2.** *Let  $(u^*, v^*) \in \text{BV}(\Omega) \times L^1(\Omega)$  be an optimal pair for (2.4). Then, the following assertions hold:*

- (i) *If  $\lambda_1 \rightarrow \infty, f \in L^1(\Omega)$  then  $v^* \rightarrow 0$  in  $L^1(\Omega)$ .*
- (ii) *If  $\lambda_2 \rightarrow \infty, f \in L^2(\Omega)$  then  $\|f - u^* - v^*\|_{L^2(\Omega)} \rightarrow 0$ . If in addition  $\lambda_1$  is fixed, then the same result holds with  $f \in L^1(\Omega)$ .*
- (iii) *If both  $\lambda_1, \lambda_2 \rightarrow \infty$  and  $f \in L^2(\Omega)$  then (i) holds and we have that  $u^* \rightarrow f$  in  $L^1(\Omega)$ . If  $f \in \text{BV}(\Omega)$  this convergence is also weakly\* in  $\text{BV}(\Omega)$ .*

*Proof.* (i) We notice that

$$\lambda_1 \|v^*\|_{L^1(\Omega)} \leq J(u) + \lambda_1 \|v\|_{L^1(\Omega)} + \frac{\lambda_2}{2} \|f - u - v\|_{L^2(\Omega)}^2, \quad \forall u \in \text{BV}(\Omega), v \in L^1(\Omega),$$

which by setting  $v = f - u$ , implies

$$\lambda_1 \|v^*\|_{L^1(\Omega)} \leq J(u) + \lambda_1 \|f - u\|_{L^1(\Omega)}, \quad \forall u \in \text{BV}(\Omega), v \in L^1(\Omega). \quad (2.6)$$

Given  $\epsilon > 0$ , we can find  $u_\epsilon \in C_c^\infty(\Omega)$  such that  $\|f - u_\epsilon\|_{L^1(\Omega)} < \epsilon$ . Thus (2.6) becomes

$$\|v^*\|_{L^1(\Omega)} \leq \frac{1}{\lambda_1} J(u_\epsilon) + \epsilon \Rightarrow \limsup_{\lambda_1 \rightarrow \infty} \|v^*\|_{L^1(\Omega)} \leq \epsilon.$$

Since  $\epsilon$  was arbitrary, the result follows.

(ii) In this case we have that for every  $u \in \text{BV}(\Omega)$  and  $v \in L^1(\Omega)$

$$\frac{\lambda_2}{2} \|f - u^* - v^*\|_{L^2(\Omega)}^2 \leq J(u) + \lambda_1 \|v\|_{L^1(\Omega)} + \frac{\lambda_2}{2} \|f - u - v\|_{L^2(\Omega)}^2, \quad (2.7)$$

which by setting  $v = 0$ , implies

$$\frac{\lambda_2}{2} \|f - u^* - v^*\|_{L^2(\Omega)}^2 \leq J(u) + \frac{\lambda_2}{2} \|f - u\|_{L^2(\Omega)}^2, \quad \forall u \in \text{BV}(\Omega).$$

Then, proceeding as in step (i) the result follows. Notice that if we assume that  $\lambda_1$  is fixed (or more generally bounded from above) and by merely assuming  $f \in L^1(\Omega)$ , we can have the same result by setting  $v = f$  and  $u = 0$  in (2.7).

(iii) Notice that if both  $\lambda_1, \lambda_2 \rightarrow \infty$  and  $f \in L^2(\Omega)$  then from (i), (ii) we get that  $\|v^*\|_{L^1(\Omega)} \rightarrow 0$  and  $\|f - u^* - v^*\|_{L^1(\Omega)} \rightarrow 0$  and hence an application of the triangle inequality implies that  $u^* \rightarrow f$  in  $L^1(\Omega)$ . If in addition  $f \in \text{BV}(\Omega)$ , then we have that  $|Du^*|(\Omega)$  is uniformly bounded by setting  $v = 0, u = f$  in

$$C_1 |Du^*|(\Omega) \leq J(u^*) \leq J(u) + \lambda_1 \|v\|_{L^1(\Omega)} + \frac{\lambda_2}{2} \|f - u - v\|_{L^2(\Omega)}^2, \quad \forall u \in \text{BV}(\Omega), v \in L^1(\Omega).$$

From compactness in  $\text{BV}(\Omega)$  and the fact that  $\|f - u^*\|_{L^1(\Omega)} \rightarrow 0$  we infer that  $u^* \rightarrow f$  weakly\* in  $\text{BV}(\Omega)$ .  $\square$

In what follows, we refine the result above and prove convergence of the minimisers of (2.4) to the minimisers of the single noise models. To do so, we apply  $\Gamma$ -convergence arguments [21] to the IC term  $\Phi^{\lambda_1, \lambda_2}$ .

**Proposition 2.3.** *Let  $f \in L^1(\Omega)$  and let us define the functional  $F^{\lambda_1, \lambda_2} : L^1(\Omega) \rightarrow \mathbb{R}^+$  by  $F^{\lambda_1, \lambda_2}(u) := \Phi^{\lambda_1, \lambda_2}(u, f)$ . Then*

- (i) *For any fixed  $\lambda_1$ ,  $F^{\lambda_1, \lambda_2}$   $\Gamma$ -converges to  $F_1(\cdot) := \lambda_1 \|f - \cdot\|_{L^1(\Omega)}$  as  $\lambda_2 \rightarrow \infty$ .*
- (ii) *For any fixed  $\lambda_2$ ,  $F^{\lambda_1, \lambda_2}$   $\Gamma$ -converges to  $\bar{F}_2(\cdot)$  as  $\lambda_1 \rightarrow \infty$ , where for every  $u \in L^1(\Omega)$ ,  $\bar{F}_2$  is defined as*

$$\bar{F}_2(u) := \frac{\lambda_2}{2} \|f - u\|_{L^2(\Omega)}^2 := \begin{cases} \frac{\lambda_2}{2} \|f - u\|_{L^2(\Omega)}^2, & \text{if } f - u \in L^2(\Omega), \\ +\infty, & \text{if } f - u \in L^1(\Omega) \setminus L^2(\Omega). \end{cases}$$

*Proof.* For (i), let  $(\lambda_2^{(n)})_{n \in \mathbb{N}}$  be a sequence with  $\lambda_2^{(n)} \rightarrow \infty$  and set  $F^n := F^{\lambda_1, \lambda_2^{(n)}}$ . We notice that  $F^n$  converges uniformly to  $\lambda_1 \|f - \cdot\|_{L^1(\Omega)}$ . Indeed, for  $u \in L^1(\Omega)$ , we have

$$\begin{aligned} |F^n(u) - \lambda_1 \|f - u\|_{L^1(\Omega)}| &= \left| \int_{\Omega} \varphi(f - u) - \lambda_1 |f - u| dx \right| \\ &\leq \int_{|f-u| \geq \frac{\lambda_1}{\lambda_2^{(n)}}} \frac{\lambda_1^2}{2\lambda_2^{(n)}} dx + \int_{|f-u| < \frac{\lambda_1}{\lambda_2^{(n)}}} \left| \frac{\lambda_2^{(n)}}{2} |f - u|^2 - \lambda_1 |f - u| \right| dx \\ &\leq \int_{\Omega} \frac{\lambda_1^2}{2\lambda_2^{(n)}} dx + \int_{\Omega} \frac{\lambda_1^2}{2\lambda_2^{(n)}} dx + \int_{\Omega} \frac{\lambda_1^2}{\lambda_2^{(n)}} dx \rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Since the last limit is independent of  $u$ , the convergence of the functionals is indeed uniform. Moreover  $F^n$  is continuous with respect to the  $L^1$  topology, see for instance [14]. Thus from [21, Proposition 5.2] we immediately get that  $F^n$   $\Gamma$ -converges to  $F_1$  as  $n \rightarrow \infty$ .

For (ii), we now set  $F^n := F^{\lambda_1^{(n)}, \lambda_2}$  with  $\lambda_1^{(n)} \rightarrow \infty$  and observe that  $F^n$  converges pointwise to  $\overline{F}_2$ . Indeed if  $f - u \in L^1(\Omega) \setminus L^2(\Omega)$  then

$$F^n(u) \geq \int_{|f-u| < \frac{\lambda_1^{(n)}}{\lambda_2}} \frac{\lambda_2}{2} |f - u|^2 dx \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

On the other hand, if  $f - u \in L^2(\Omega)$ , proceeding as before we have

$$\begin{aligned} \left| F^n(u) - \frac{\lambda_2}{2} \|f - u\|_{L^2(\Omega)}^2 \right| &\leq \int_{|f-u| \geq \frac{\lambda_1^{(n)}}{\lambda_2}} \lambda_1^{(n)} |f - u| - \frac{(\lambda_1^{(n)})^2}{2\lambda_2} dx + \int_{|f-u| \geq \frac{\lambda_1^{(n)}}{\lambda_2}} \frac{\lambda_2}{2} |f - u|^2 dx \\ &\leq \int_{|f-u| \geq \frac{\lambda_1^{(n)}}{\lambda_2}} \frac{3}{2} \lambda_2 |f - u|^2 dx \rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

If  $(\lambda_1^{(n)})_{n \in \mathbb{N}}$  is increasing, it can be easily verified that the sequence  $(F^n)_{n \in \mathbb{N}}$  is increasing. Moreover the  $L^2$  norm is lower semicontinuous with respect to the strong  $L^1$  topology. Hence from [21, Remark 5.5], we have that  $F^n$   $\Gamma$ -converges to  $\overline{F}_2$ . In the case where  $(\lambda_1^{(n)})_{n \in \mathbb{N}}$  is non-monotonically going to infinity, we can find an increasing subsequence and then the result follows from the Urysohn property of  $\Gamma$ -convergence, see [21, Proposition 8.3].  $\square$

As a corollary, we obtain the following result on the convergence of minimisers.

**Corollary 2.4** (Convergence to single noise models). *The following two results hold:*

- (i) *Let  $f \in L^1(\Omega)$ ,  $\lambda_1 > 0$  fixed and  $(\lambda_2^{(n)})_{n \in \mathbb{N}}$  with  $\lambda_2^{(n)} \rightarrow \infty$ . If  $(u_n)_{n \in \mathbb{N}} \subseteq \text{BV}(\Omega)$  is a sequence of minimisers of (2.4), then every subsequence of  $(u_n)_{n \in \mathbb{N}}$  has a weak\* in  $\text{BV}(\Omega)$  cluster point which is a minimiser of*

$$\min_{u \in \text{BV}(\Omega)} J(u) + \lambda_1 \|f - u\|_{L^1(\Omega)}. \quad (2.8)$$

*Moreover if the solution  $u^*$  of (2.8) is unique, then  $u_n \rightarrow u^*$  weakly\* in  $\text{BV}(\Omega)$  and*

$$J(u_n) + \Phi^{\lambda_1, \lambda_2^{(n)}}(u_n, f) \rightarrow J(u^*) + \lambda_1 \|f - u^*\|_{L^1(\Omega)}. \quad (2.9)$$

- (ii) *Let  $f \in L^2(\Omega)$ ,  $\lambda_2 > 0$  fixed and  $(\lambda_1^{(n)})_{n \in \mathbb{N}}$  with  $\lambda_1^{(n)} \rightarrow \infty$ . If  $(u_n)_{n \in \mathbb{N}} \subseteq \text{BV}(\Omega)$  is a sequence of minimisers of (2.4), then  $u_n \rightarrow u^*$  weakly\* in  $\text{BV}(\Omega)$ , where  $u^*$  is the unique minimiser of*

$$\min_{u \in \text{BV}(\Omega)} J(u) + \frac{\lambda_2}{2} \|f - u\|_{L^2(\Omega)}^2. \quad (2.10)$$

*Moreover*

$$J(u_n) + \Phi^{\lambda_1^{(n)}, \lambda_2}(u_n, f) \rightarrow J(u^*) + \frac{\lambda_2}{2} \|f - u^*\|_{L^2(\Omega)}^2.$$

*Proof.* (i) Since the functional  $J$  is lower semicontinuous with respect to  $L^1$ , then [21, Proposition 6.25] we have that the minimising functionals  $\Phi^{\lambda_1, \lambda_2^{(n)}}(\cdot, f) + J(\cdot)$  also  $\Gamma$ -converge to the functional

in (2.8). Moreover, note that  $(u_n)_{n \in \mathbb{N}}$  is uniformly bounded in  $BV(\Omega)$ . Since  $u_n$  is a minimiser we have in fact that for every  $u \in BV(\Omega)$  and every  $v \in L^1(\Omega)$ , the following three inequalities hold

$$C_1 |Du_n|(\Omega) \leq J(u_n) \leq J(u) + \lambda_1 \|v\|_{L^1(\Omega)} + \frac{\lambda_2^{(n)}}{2} \|f - u - v\|_{L^2(\Omega)}^2, \quad (2.11)$$

$$\frac{\lambda_2^{(n)}}{2} \|f - u_n - v_n\|_{L^2(\Omega)}^2 \leq J(u) + \lambda_1 \|v\|_{L^1(\Omega)} + \frac{\lambda_2^{(n)}}{2} \|f - u - v\|_{L^2(\Omega)}^2, \quad (2.12)$$

$$\lambda_1 \|v_n\|_{L^1(\Omega)} \leq J(u) + \lambda_1 \|v\|_{L^1(\Omega)} + \frac{\lambda_2^{(n)}}{2} \|f - u - v\|_{L^2(\Omega)}^2. \quad (2.13)$$

By setting  $u = 0$ ,  $v = f$  in (2.11) one obtains a uniform bound for the sequence  $(|Du_n|(\Omega))_{n \in \mathbb{N}}$ . From (2.12) and from the fact that  $\Omega$  is bounded, one obtains a uniform bound for  $(\|f - u_n - v_n\|_{L^1(\Omega)})_{n \in \mathbb{N}}$ . Similarly, from (2.13), a uniform bound on  $(\|v_n\|_{L^1(\Omega)})_{n \in \mathbb{N}}$  is obtained and this means that  $(\|u_n\|_{L^1(\Omega)})_{n \in \mathbb{N}}$  is also bounded. Thus every subsequence of  $u_n$  has a cluster point in  $BV(\Omega)$  with respect to the weak\* topology, which must be a minimiser of (2.8), [21, Corollary 7.20]. Furthermore if (2.8) has a unique minimiser, then every subsequence of  $u_n$  has a further subsequence that converges to  $u$  weakly\* in  $BV(\Omega)$ . Thus, in this case  $u_n$  converges to  $u$  weakly\* in  $BV(\Omega)$  and moreover (2.9) holds by [21, Corollary 7.20].

(ii) The  $\Gamma$ -convergence of the energies follows as above. By setting,  $v = u = 0$  in (2.11) one obtains a uniform bound on  $(|Du_n|(\Omega))_{n \in \mathbb{N}}$  and similarly as in (i) a bound on  $L^1(\Omega)$  and consequently in  $BV(\Omega)$  is obtained for  $u$ . The rest of the proof follows as in (i), bearing in mind that the solution of (2.10) is unique. □

We summarise our findings so far in Table 1, where we have combined the results of Proposition 2.2 and Corollary 2.4.

	$\lambda_1 \rightarrow \infty$ $f \in L^1(\Omega)$	$\lambda_1 \rightarrow \infty$ $\lambda_2$ fixed $f \in L^2(\Omega)$	$\lambda_2 \rightarrow \infty$ $f \in L^2(\Omega)$ (or $f \in L^1(\Omega)$ & $\lambda_1$ fixed)	$\lambda_2 \rightarrow \infty$ $\lambda_1$ fixed $f \in L^1(\Omega)$ $J_{-\lambda_1} L^1$ has !sol. $u^*$	$\lambda_1 \rightarrow \infty$ $\lambda_2 \rightarrow \infty$ $f \in L^2(\Omega)$
$v$	$v \rightarrow 0$ in $L^1(\Omega)$	$v \rightarrow 0$ in $L^1(\Omega)$	$\ f - u - v\ _{L^2(\Omega)} \rightarrow 0$	$v \rightarrow f - u^*$ in $L^1(\Omega)$	$v \rightarrow 0$ in $L^1(\Omega)$
$u$	cannot say	$u \rightarrow$ solution $J_{-\frac{\lambda_2}{2}} L^2$ , $w^*$ in $BV(\Omega)$	cannot say	$u \rightarrow u^*, w^*$ in $BV(\Omega)$	$u \rightarrow f$ in $L^1(\Omega)$ , ( $w^*$ in $BV(\Omega)$ ) if $f \in BV(\Omega)$

Table 1: Summary of all the asymptotic results concerning the solution pair  $u, v$  of (2.4) when one or both parameters  $\lambda_1$  and  $\lambda_2$  are let to infinity.

In the case of bounded data and TV regularisation, the results obtained above can be refined. We first recall the following well-known result, see [19, Lemma 3.5].

**Proposition 2.5.** *Let  $u$  be a solution of (2.5), with  $f \in L^\infty(\Omega)$ . Then the following maximum principle holds:*

$$\operatorname{ess\,inf}_{x \in \Omega} f(x) \leq \operatorname{ess\,inf}_{x \in \Omega} u(x) \leq \operatorname{ess\,sup}_{x \in \Omega} u(x) \leq \operatorname{ess\,sup}_{x \in \Omega} f(x).$$

We can now prove the following result for the TV–IC minimisation problem (2.5).

**Proposition 2.6.** *Suppose that  $f \in L^\infty(\Omega)$  and the parameters  $\lambda_1, \lambda_2 > 0$  satisfy the following condition*

$$\frac{\lambda_1}{\lambda_2} \geq 2\|f\|_\infty. \tag{2.14}$$

Then, if  $u$  is a solution of (2.5), there holds

$$\Phi^{\lambda_1, \lambda_2}(u, f) = \frac{\lambda_2}{2}\|f - u\|_{L^2(\Omega)}^2. \tag{2.15}$$

As a result, the problem (2.5) is equivalent to a standard TV- $L^2$  minimisation problem.

*Proof.* This is a direct consequence of Proposition 2.5 and the formulation (2.1)–(2.2) of  $\Phi^{\lambda_1, \lambda_2}(u, f)$ . Indeed, using a translation argument and Proposition 2.5 one shows directly that for the solution  $u$  of (2.5), it holds  $\|f - u\|_\infty \leq 2\|f\|_\infty$ . Thus if (2.14) holds, (2.1) implies that  $\varphi(f - u) = \frac{\lambda_2}{2}|f - u|^2$  so that (2.15) holds as well.  $\square$

We note here that the adaptation of Proposition 2.5 and, consequently, of Proposition 2.6 to other widely used regularisers is not immediate. For instance, it remains an open problem to show that the solution  $u$  of the TGV- $L^2$  problem with data  $f \in L^\infty(\Omega)$  is also an  $L^\infty$  function, see for instance the corresponding discussion in [54]. However, in dimension one this fact is true when  $f \in \text{BV}(\Omega)$ , by taking advantage of the estimate  $\|u\|_{L^\infty(\Omega)} \leq C\|u\|_{\text{BV}(\Omega)}$ , see for instance [49, Proposition 2].

In view of the Proposition above, one sees that in the case of TV regularisation, the Gaussian noise model can be recovered simply by fixing either  $\lambda_2$  and setting  $\lambda_1$  large enough or by fixing  $\lambda_1$  and setting  $\lambda_2$  small enough. In Figure 3 we graphically depict this behaviour.

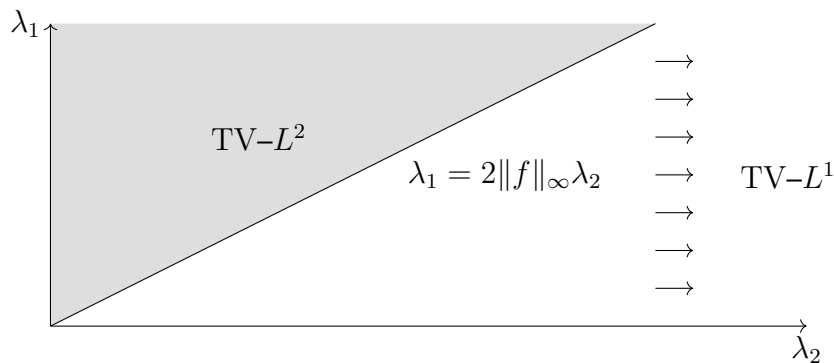


Figure 3: If  $\frac{\lambda_1}{\lambda_2} \geq 2\|f\|_\infty$  then (2.5) is equivalent to the TV- $L^2$  problem, see Proposition 2.6. By fixing  $\lambda_1$  and sending  $\lambda_2$  to infinity the solution  $u$  converges to a solution of an TV- $L^1$  in the sense of Corollary 2.4.

## 2.2 Convergence of the parameters to zero and non-exact recovery of the data

We now study the asymptotic behaviour of the model when the parameters are sent to zero. For this analysis and for the sake of simplicity, we focus on the TV minimisation model (2.5) but the results can be easily extended to the general regulariser case. First, we recall the definition of the *mean*  $u_\Omega$  and

median values  $\{u^\Omega\}$  of an  $L^1$  function  $u$  defined by:

$$u_\Omega := \int_\Omega u \, dx,$$

$$u^\Omega \in \operatorname{argmin}_{c \in \mathbb{R}} \int_\Omega |u - c| \, dx.$$

**Remark 2.7.** Note that the median value is not necessarily unique. Moreover, if  $u \in L^2(\Omega)$  then  $u_\Omega = \operatorname{argmin}_{c \in \mathbb{R}} \int_\Omega |u - c|^2 \, dx$ .

We have the following result:

**Proposition 2.8.** Let  $f \in L^1(\Omega)$  and  $(\lambda_1^{(n)})_{n \in \mathbb{N}}, (\lambda_2^{(n)})_{n \in \mathbb{N}}$  two sequences such that

$$\lambda_1^{(n)} \rightarrow 0 \quad \text{and} \quad \frac{\lambda_1^{(n)}}{\lambda_2^{(n)}} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (2.16)$$

Then, denoting by  $(u_n)_{n \in \mathbb{N}}$  the sequence of the corresponding solutions of (2.5) we have that

$$u_n \rightarrow f^\Omega \quad \text{weakly}^* \text{ in } \operatorname{BV}(\Omega).$$

By this we mean that every subsequence of  $(u_n)_{n \in \mathbb{N}}$  has a further subsequence converging to a median of  $f$ .

*Proof.* Let  $(u_n)_{n \in \mathbb{N}} \subseteq \operatorname{BV}(\Omega)$  be a sequence of the corresponding solutions for the parameters  $(\lambda_1^{(n)}, \lambda_2^{(n)})$ . Notice that the sequence  $(u_n)_{n \in \mathbb{N}}$  is uniformly bounded in  $\operatorname{BV}(\Omega)$ . Indeed the bound on TV is obtained again from (2.11), while for the  $L^1$  bound, we first observe that:

$$-\frac{\lambda_1^2}{2\lambda_2} |\Omega| + \lambda_1 \|f - u\|_{L^1(\Omega)} \leq \int_\Omega \varphi(f - u) \, dx, \quad \forall u \in L^1(\Omega). \quad (2.17)$$

Using (2.17), we get

$$-\frac{(\lambda_1^{(n)})^2}{2\lambda_2^{(n)}} |\Omega| + \lambda_1^{(n)} \|f - u_n\|_{L^1(\Omega)} \leq \int_\Omega \varphi(f - u_n) \, dx \leq \lambda_1^{(n)} \|f\|_{L^1(\Omega)} \Rightarrow$$

$$-\frac{\lambda_1^{(n)}}{2\lambda_2^{(n)}} |\Omega| + \|f - u_n\|_{L^1(\Omega)} \leq \|f\|_{L^1(\Omega)},$$

where the bound is obtained using (2.16). Thus every subsequence of  $(u_n)_{n \in \mathbb{N}}$  has a further (not relabelled) subsequence converging to an element  $u^* \in \operatorname{BV}(\Omega)$ . We will show that  $u^*$  is a median of  $f$ . Notice first that since  $\lambda_1^{(n)} \rightarrow 0$ , then from (2.11) we get that  $|Du_n|(\Omega) \rightarrow 0$ . Thus, from the lower semicontinuity of total variation and the fact that  $\Omega$  is connected we get that  $u^*$  is a constant. By thus setting  $u = c \in \mathbb{R}$  and  $v = f - c$  along with (2.17) in

$$|Du_n|(\Omega) + \int_\Omega \varphi(f - u_n) \, dx \leq |Du|(\Omega) + \lambda_1^{(n)} \|v\|_{L^1(\Omega)} + \frac{\lambda_2^{(n)}}{2} \|f - u - v\|_{L^2(\Omega)}^2,$$

we get

$$\begin{aligned} -\frac{(\lambda_1^{(n)})^2}{2\lambda_2^{(n)}}|\Omega| + \lambda_1^{(n)}\|f - u_n\|_{L^1(\Omega)} &\leq \lambda_1^{(n)}\|f - c\|_{L^1(\Omega)} &\Rightarrow \\ -\frac{\lambda_1^{(n)}}{\lambda_2^{(n)}}|\Omega| + \|f - u_n\|_{L^1(\Omega)} &\leq \|f - c\|_{L^1(\Omega)} &\Rightarrow \text{(by taking limits)} \\ \|f - u^*\|_{L^1(\Omega)} &\leq \|f - c\|_{L^1(\Omega)}. \end{aligned}$$

Hence since  $u^*$  is constant and  $c \in \mathbb{R}$  was arbitrary, we have that  $u$  is a median of  $f$ .  $\square$

Similarly, we have the following result:

**Proposition 2.9.** *Let  $f \in L^2(\Omega)$  and  $(\lambda_1^{(n)})_{n \in \mathbb{N}}, (\lambda_2^{(n)})_{n \in \mathbb{N}}$  two sequences such that*

$$\lambda_2^{(n)} \rightarrow 0 \quad \text{and} \quad \frac{\lambda_2^{(n)}}{\lambda_1^{(n)}} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

*Then for the corresponding solutions  $(u_n)_{n \in \mathbb{N}}$  of (2.5) we have that*

$$u_n \rightarrow f_\Omega \quad \text{weakly}^* \text{ in } \text{BV}(\Omega).$$

*Proof.* The proof follows the same steps as in Proposition 2.8. First, observe that the sequence of solutions  $(u_n)_{n \in \mathbb{N}}$  is bounded in  $\text{BV}(\Omega)$ . Indeed, from (2.11) we have that  $|Du_n|(\Omega) \rightarrow 0$ . From (2.13) we further get that  $(v_n)_{n \in \mathbb{N}}$  is bounded in  $L^1$  and from (2.12) we get that  $(f - u_n - v_n)_{n \in \mathbb{N}}$  is bounded in  $L^1$ . Thus, from the triangle inequality we have that  $(u_n)_{n \in \mathbb{N}}$  is bounded in  $L^1$ . Hence, there exists a subsequence of  $(u_n)_{n \in \mathbb{N}}$  that converges to a function  $u^*$  weakly\* in  $\text{BV}(\Omega)$  with  $u^*$  being a constant. It remains to show that  $u^*$  is the mean value of  $f$ . As before, we have for an arbitrary  $c \in \mathbb{R}$

$$\begin{aligned} |Du_n|(\Omega) + \int_{\Omega} \varphi(f - u_n) dx &\leq \frac{\lambda_2^{(n)}}{2}\|f - c\|_{L^2(\Omega)}^2 &\Rightarrow \\ \frac{1}{2} \int_{|f - u_n| < \frac{\lambda_1^{(n)}}{\lambda_2^{(n)}}} |f - u_n|^2 dx &\leq \frac{1}{2}\|f - c\|_{L^2(\Omega)}^2 &\Rightarrow \text{(using Fatou's Lemma)} \\ \frac{1}{2}\|f - u^*\|_{L^2(\Omega)}^2 &\leq \frac{1}{2}\|f - c\|_{L^2(\Omega)}^2. \end{aligned}$$

Since  $u$  is a constant and  $c \in \mathbb{R}$  was arbitrary the proof is complete.  $\square$

The following proposition states that with the  $L^1$ – $L^2$  infimal convolution fidelity model we can never expect exact recovery of the data. This is similar to the pure  $L^2$  model, see also [19, Proposition 4.1].

**Proposition 2.10.** *Let  $f \in L^1(\Omega)$  and  $u^*$  to be a solution of the minimisation problem (2.5). Then  $u^* = f$  if and only if  $f$  is a constant.*

*Proof.* One direction is straightforward. Suppose now that  $f$  is a solution of (2.5). Note that in this case necessarily we must have  $f \in \text{BV}(\Omega) \subseteq L^{d^*}(\Omega)$ , where  $d^* = d/(d-1)$ , see [4]. It follows that for every  $0 < \epsilon < 1$ , the function  $f_\epsilon := \epsilon f$  is suboptimal. Thus we have

$$|Df|(\Omega) \leq \epsilon |Df|(\Omega) + \Phi^{\lambda_1, \lambda_2}(u, \epsilon f) \implies 0 \leq (\epsilon - 1)|Df|(\Omega) + \Phi^{\lambda_1, \lambda_2}(u, \epsilon f).$$

We continue

$$\begin{aligned}
& 0 \leq (\epsilon - 1)|Df|(\Omega) + \Phi^{\lambda_1, \lambda_2}(u, \epsilon f) \\
& 0 \leq (\epsilon - 1)|Df|(\Omega) + \int_{|f-\epsilon f| \geq \frac{\lambda_1}{\lambda_2}} \lambda_1 |f - \epsilon f| - \frac{\lambda_1^2}{2\lambda_2} dx + \int_{|f-\epsilon f| < \frac{\lambda_1}{\lambda_2}} \frac{\lambda_2}{2} |f - \epsilon f|^2 dx + \implies \\
& 0 \geq |Df|(\Omega) - \int_{|f| \geq \frac{\lambda_1}{\lambda_2|1-\epsilon|}} \lambda_1 |f| dx + \int_{|f| \geq \frac{\lambda_1}{\lambda_2|1-\epsilon|}} \frac{\lambda_1^2}{2\lambda_2|1-\epsilon|} dx - |1-\epsilon| \int_{|f| < \frac{\lambda_1}{\lambda_2|1-\epsilon|}} \frac{\lambda_2}{2} |f|^2 dx.
\end{aligned} \tag{2.18}$$

Now working with each one of the first three terms in (2.18), using dominated convergence, we have

$$\lim_{\epsilon \rightarrow 1} \int_{|f| \geq \frac{\lambda_1}{\lambda_2|1-\epsilon|}} |f| dx = 0, \tag{2.19}$$

$$\lim_{\epsilon \rightarrow 1} \int_{|f| \geq \frac{\lambda_1}{\lambda_2|1-\epsilon|}} \frac{\lambda_1^2}{2\lambda_2|1-\epsilon|} dx = \frac{\lambda_1}{2} \lim_{t \rightarrow \infty} \int_{|f| \geq t} t dx \leq \frac{\lambda_1}{2} \lim_{t \rightarrow \infty} \int_{|f| \geq t} |f| dx = 0, \tag{2.20}$$

and finally

$$\begin{aligned}
& \lim_{\epsilon \rightarrow 1} |1-\epsilon| \int_{|f| < \frac{\lambda_1}{\lambda_2|1-\epsilon|}} \frac{\lambda_2}{2} |f|^2 dx = \frac{\lambda_2}{2} \lim_{\epsilon \rightarrow 1} |1-\epsilon| \int_{|f| < \frac{\lambda_1}{\lambda_2|1-\epsilon|}} |f|^{\frac{d}{d-1}} |f|^{\frac{d-2}{d-1}} \\
& \leq \frac{\lambda_2}{2} \left( \frac{\lambda_1}{\lambda_2} \right)^{\frac{d-2}{d-1}} \lim_{\epsilon \rightarrow 1} \frac{|1-\epsilon|}{|1-\epsilon|^{\frac{d-2}{d-1}}} \int_{|f| < \frac{\lambda_1}{\lambda_2|1-\epsilon|}} |f|^{\frac{d}{d-1}} dx \leq \frac{\lambda_2}{2} \left( \frac{\lambda_1}{\lambda_2} \right)^{\frac{d-2}{d-1}} \|f\|_{L^{d^*}(\Omega)}^{d^*} \lim_{\epsilon \rightarrow 1} |1-\epsilon|^{\frac{1}{d-1}} = 0.
\end{aligned} \tag{2.21}$$

By combining (2.19), (2.20), (2.21) with (2.18) we get that  $|Df|(\Omega) = 0$ , and since  $\Omega$  is connected,  $f$  is a constant function. Note that the calculations above assumed that  $d > 1$  but if  $d = 1$  then  $f \in L^\infty(\Omega) \subseteq L^2(\Omega)$  and the analogous calculation to (2.21) follows more easily.  $\square$

It is clear from all the analysis above that the TV-IC model, at least when  $f \in L^\infty(\Omega)$ , can reproduce the TV- $L^2$  solutions, but as far as the TV- $L^1$  solutions are concerned, these are only (guaranteed to be) recovered in the limit  $\lambda_2 \rightarrow \infty$ , see again Figure 3.

## 2.3 The one-homogeneous analogue

We conclude this section by briefly presenting an alternative form of (1.3), i.e., its one-homogeneous analogue, by which the TV- $L^1$  solutions can also be recovered for finite parameters. This discussion is motivated by some analogous results in [11]. We define:

$$\Phi_{1-hom}^{\lambda_1, \lambda_2}(u, f) := \min_{v \in L^1(\Omega)} \lambda_1 \|v\|_{L^1(\Omega)} + \lambda_2 \|f - u - v\|_{L^2(\Omega)},$$

which, via a straightforward computation gives

$$\begin{aligned}
& \min \left( \lambda_1, \frac{\lambda_2}{2|\Omega|^{1/2}} \right) \|f - u\|_{L^1(\Omega)} \leq \min_{v \in L^1(\Omega)} \lambda_1 \|v\|_{L^1(\Omega)} + \frac{\lambda_2}{2|\Omega|^{1/2}} \|f - u - v\|_{L^1(\Omega)} \\
& \leq \min_{v \in L^1(\Omega)} \lambda_1 \|v\|_{L^1(\Omega)} + \frac{\lambda_2}{2} \|f - u - v\|_{L^2(\Omega)} \leq \lambda_1 \|f - u\|_{L^1(\Omega)}.
\end{aligned}$$



Hence, it is clear that

$$\Phi_{1-hom}^{\lambda_1, \lambda_2}(u, f) = \lambda_1 \|f - u\|_{L^1(\Omega)}, \quad \text{if } \frac{\lambda_1}{\lambda_2} \leq \frac{1}{2|\Omega|^{1/2}},$$

Thus, under such choice the TV– $L^1$  model can be recovered. We state here without any proof the relationship between these two versions. Let us define the following sets for  $f \in L^1(\Omega)$ :

$$\begin{aligned} S_{IC} &= \left\{ u^* \in \text{BV}(\Omega) : u^* = \operatorname{argmin}_{u \in \text{BV}(\Omega)} |Du|(\Omega) + \Phi^{\lambda_1, \lambda_2}(u, f), \text{ for some } \lambda_1, \lambda_2 > 0 \right\}, \\ S_{IC}^{1-hom} &= \left\{ u^* \in \text{BV}(\Omega) : u^* = \operatorname{argmin}_{u \in \text{BV}(\Omega)} |Du|(\Omega) + \Phi_{1-hom}^{\lambda_1, \lambda_2}(u, f), \text{ for some } \lambda_1, \lambda_2 > 0 \right\}, \\ S_{L^1} &= \left\{ u^* \in \text{BV}(\Omega) : u^* = \operatorname{argmin}_{u \in \text{BV}(\Omega)} |Du|(\Omega) + \lambda_1 \|f - u\|_{L^1(\Omega)}, \text{ for some } \lambda_1 > 0 \right\}. \end{aligned}$$

Then one can show by using similar techniques as in [11] that

$$S_{IC}^{1-hom} = S_{IC} \cup S_{L^1} \quad \text{and that, in general, } S_{L^1} \setminus S_{IC} \neq \emptyset.$$

### 3 Exact solutions

In order to get more insights about the relationship of the TV–IC model with the pure TV– $L^1$  and TV– $L^2$  models, we compute in this section some exact solutions for simple one dimensional data functions  $f$ . In particular, we set here  $\Omega = (-2L, 2L)$  for some  $L > 0$ , and we consider as data  $f$  the following step function

$$f(x) = \begin{cases} 0, & \text{if } x \in (-2L, -L) \cup (L, 2L), \\ h, & \text{if } x \in [-L, L], \end{cases} \quad (3.1)$$

where  $h > 0$ . Using similar techniques as in [10, 11, 48, 51], we can easily show using primal-dual optimality conditions, that a function  $u \in \text{BV}(\Omega)$  is a solution of (2.5) if and only if there exists a function  $v \in H_0^1(\Omega)$  such that

$$v' = \begin{cases} \lambda_1 \frac{f-u}{|f-u|}, & \text{if } |f-u| \geq \frac{\lambda_1}{\lambda_2}, \\ \lambda_2(f-u), & \text{if } |f-u| < \frac{\lambda_1}{\lambda_2}, \end{cases} \quad (3.2)$$

$$v \in \text{Sgn}(Du), \quad (3.3)$$

where

$$\text{Sgn}(Du) = \left\{ v \in L^\infty(\Omega) \cap L^\infty(\Omega, Du) : \|v\|_\infty \leq 1, v = \frac{dDu}{d|Du|}, |Du|\text{-a.e.} \right\}.$$

Here  $\frac{dDu}{d|Du|}$  denotes the Radon–Nikodým density of  $Du$  with respect to  $|Du|$ . Compared to the aforementioned references, the only difference here is the right-hand side of (3.2) which is the subdifferential of  $\Phi^{\lambda_1, \lambda_2}$  evaluated at  $f - u$ . With the help of the optimality conditions above, we are able to compute analytically all the solutions to the problem (2.5) for the data (3.1), and for all combinations of the parameters  $\lambda_1, \lambda_2$ . Note that similarly to pure  $L^1$  and  $L^2$  cases one can show that no new jump discontinuities are created for the solution  $u$ , which will be constant in the areas where  $\{f \neq u\}$ . Thus

all the solutions will be either constants or piecewise constants with jumps at  $x = -L$  and  $x = L$  which must also have the same orientations with the jumps of  $f$ .

We first examine the case  $u = \frac{h}{2}$ , i.e., the mean value of  $f$ . In such case we have  $|f - u| = \frac{h}{2}$  everywhere and thus if  $\frac{h}{2} \leq \frac{\lambda_1}{\lambda_2}$  then  $v' = \lambda_2(f - u)$  everywhere. In order for the condition (3.2) to hold we must also have  $\lambda_2 \leq \frac{2}{hL}$ . Note that if  $\frac{h}{2} \leq \frac{\lambda_1}{\lambda_2}$  and  $\lambda_2 < \frac{2}{hL}$ , then the solution must be constant otherwise one can check that in every case the condition (3.3) would be violated. One can further show in this case that if  $u$  is a constant with  $\frac{h}{2} < \frac{\lambda_1}{\lambda_2}$ , the only possibility is  $u = \frac{h}{2}$ .

Suppose now that  $\frac{h}{2} \geq \frac{\lambda_1}{\lambda_2}$  and also  $\lambda_1 < \frac{1}{L}$ . Observe that in this case, every constant function  $u = c$  with  $\frac{\lambda_1}{\lambda_2} \leq c \leq h - \frac{\lambda_1}{\lambda_2}$  satisfies (3.2)–(3.3). In that case we have  $|f - u| \geq \frac{\lambda_1}{\lambda_2}$  everywhere. Notice again that no other constant function is a solution. By contradiction, that would mean that  $|f - u| > \frac{\lambda_1}{\lambda_2}$  and  $|f - u| < \frac{\lambda_1}{\lambda_2}$  on  $(-2L, L) \cup (L, 2L)$  and  $(-L, L)$  respectively (or vice versa). With the help of (3.2) and the fact that  $v \in H_0^1(\Omega)$  one would then arrive to a contradiction. Furthermore, one can check that discontinuous solutions cannot occur in this case either.

We concentrate now on the case  $\frac{h}{2} \geq \frac{\lambda_1}{\lambda_2}$  and  $\lambda_1 = \frac{1}{L}$ . Note that the constant functions  $u = c$  with  $\frac{\lambda_1}{\lambda_2} \leq c \leq h - \frac{\lambda_1}{\lambda_2}$  are solutions in this case as well. However, one can also verify that the following family of discontinuous functions are also solutions:

$$u(x) = \begin{cases} c_1, & \text{if } x \in (-2L, -L), \\ h - d, & \text{if } x \in [-L, L], \\ c_2, & \text{if } x \in (L, 2L), \end{cases} \quad \frac{\lambda_1}{\lambda_2} \leq c_i < h - d \leq h - \frac{\lambda_1}{\lambda_2}, \quad i = 1, 2.$$

It can be checked similarly as before that no other solutions can occur.

Finally, we consider the case  $\lambda_1 > \frac{1}{L}$  and  $\lambda_2 \geq \frac{2}{hL}$ . We claim that in that case the unique solution is given by

$$u(x) = \begin{cases} \frac{1}{L\lambda_2}, & \text{if } x \in (-2L, -L) \cup (L, 2L), \\ h - \frac{1}{L\lambda_2}, & \text{if } x \in [-L, L]. \end{cases}$$

One can similarly check that no other solution is possible. We summarise our findings in the following proposition.

**Proposition 3.1.** *Let  $\Omega = (-2L, 2L)$  and  $f \in \text{BV}(\Omega)$  being the jump function given by (3.1). Then the solutions  $u$  to the TV–IC minimisation problem*

$$\min_{u \in \text{BV}(\Omega)} |Du|(\Omega) + \Phi^{\lambda_1, \lambda_2}(u, f),$$

are given by the following formulae:

(i) *If  $\frac{h}{2} < \frac{\lambda_1}{\lambda_2}$  and  $\lambda_2 \leq \frac{2}{hL}$ , then the solution is unique and given by*

$$u = \frac{h}{2}.$$

(ii) *If  $\frac{h}{2} \geq \frac{\lambda_1}{\lambda_2}$  and  $\lambda_1 < \frac{1}{L}$ , then there exist infinitely many constant solutions given by*

$$u = c, \quad \frac{\lambda_1}{\lambda_2} \leq c \leq h - \frac{\lambda_1}{\lambda_2}.$$

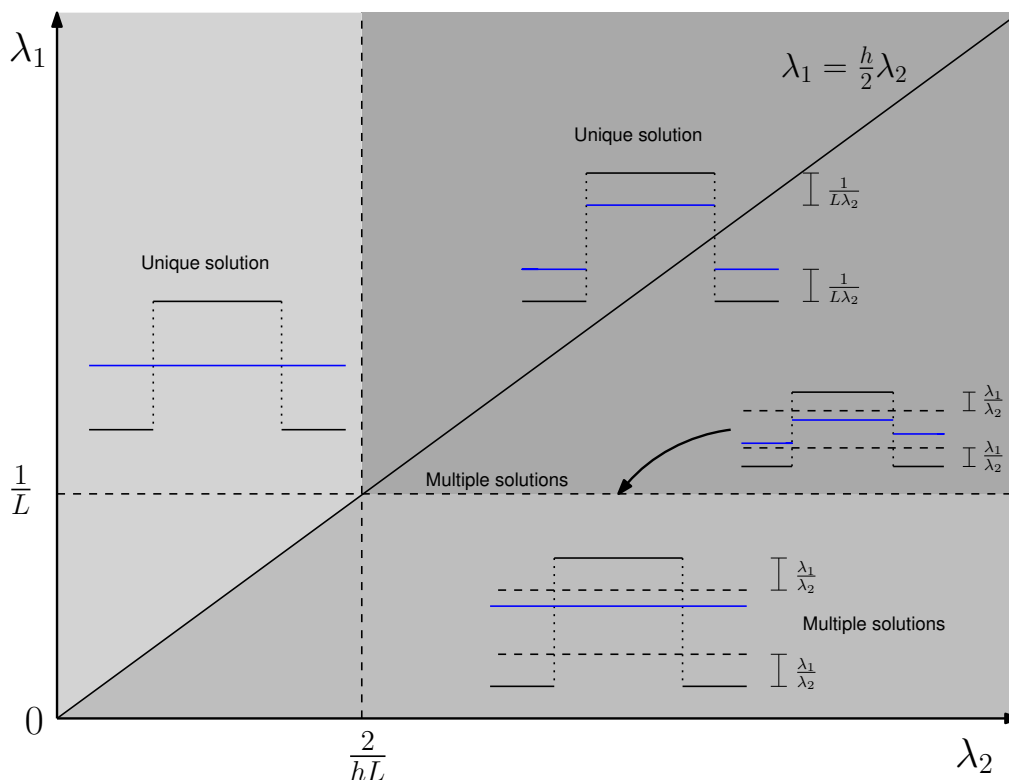


Figure 4: Visualisation of all the possible solutions to the TV–IC problem (2.5) for data (3.1) and for all the possible combinations of the parameters  $\lambda_1$  and  $\lambda_2$ , see Proposition 3.1.

(iii) If  $\frac{h}{2} \geq \frac{\lambda_1}{\lambda_2}$  and  $\lambda_1 = \frac{1}{L}$ , then there exist infinitely many solutions given by

$$u(x) = \begin{cases} c_1, & \text{if } x \in (-2L, -L), \\ h - d, & \text{if } x \in [-L, L], \\ c_2, & \text{if } x \in (L, 2L), \end{cases} \quad \frac{\lambda_1}{\lambda_2} \leq c_i \leq h - d \leq h - \frac{\lambda_1}{\lambda_2}, \quad i = 1, 2.$$

(iv) If  $\lambda_2 > \frac{2}{hL}$  and  $\lambda_1 > \frac{1}{L}$ , then the solution is unique and given by

$$u(x) = \begin{cases} \frac{1}{L\lambda_2}, & \text{if } x \in (-2L, -L) \cup (L, 2L), \\ h - \frac{1}{L\lambda_2}, & \text{if } x \in [-L, L]. \end{cases}$$

A visualisation of these solutions is depicted in Figure 4. Observe that for large enough ratio  $\frac{\lambda_1}{\lambda_2}$  all the TV– $L^2$  solutions are recovered as Proposition 2.6 predicts – compare also Figures 3 and 4. Moreover, observe that as  $\lambda_1$  and  $\frac{\lambda_1}{\lambda_2}$  goes to zero, the solutions indeed converge to a median of  $f$ , as shown in Proposition 2.8. Note however, that apart from some medians, the solutions of the TV– $L^1$  model are not recovered. More precisely, the ones that perfectly fit the data in the whole domain or part of it, i.e.,  $u = h$ ,  $u = 0$  and  $u = f$  cannot be obtained here. This is also in accordance to Proposition 2.10.

## 4 Automatic selection of parameters

We describe now a *bilevel optimisation* strategy for the estimation of optimal parameters  $\lambda_1$  and  $\lambda_2$  in the TV–IC model (2.5) based on the use of training sets, see [13, 24, 26]. This approach has been heuristically considered for the TV–IC model in [14, Section 7] with little theoretical justification. To fill this gap, we prove in this section existence results for the solution of the bilevel minimisation problem and for the corresponding adjoint problem, thus making the derivation the optimality system in [14] rigorous.

We point out that in [40, 41] an adaptive optimisation approach has been proposed for the automatic selection of parameters when a linear combination of  $L^1$  and  $L^2$  data fidelities is considered. However, differently to our setting, in that approach the noise level is assumed to be known.

### 4.1 Bilevel optimisation

Learning approaches have become very popular over the recent years due to their ability of combining data- and model-driven algorithms for the optimal design of imaging models. In particular, bilevel optimisation techniques have been proposed by several different authors in discrete [38, 39, 46, 56] and functional [13, 24, 25, 26, 34, 35] settings as a tool to estimate the “best” variational image restoration model within a certain class by means of training examples. Typically, such examples consist of images obtained in standard acquisition settings, and thus corrupted by noise with equal (unknown) intensity, paired with their corresponding versions ideally acquired in a very low-noise setting. In medical imaging, for instance, such training set can be provided by means of real and/or simulated phantoms. Note that in order to make the estimation robust, a large training set is often desirable; for that, stochastic optimisation techniques and sampling approaches can be used to reduce the computational costs, see, e.g., [15].

The general bilevel optimisation problem can be formulated as:

$$\min_{\lambda \in [0, \infty)^m} F(u_\lambda) \quad (4.1)$$

subject to:

$$u_\lambda \in \operatorname{argmin}_{u \in X} \{J(u, \lambda) := |Du|(\Omega) + \Phi^\lambda(u, f)\}, \quad (4.2)$$

where  $\lambda = (\lambda_1, \dots, \lambda_m) \in [0, \infty)^m$  are the parameters to optimise and  $F \geq 0$  is an appropriate quality measure which is minimised under the constraint that the function  $u_\lambda$  is a solution of the denoising model (4.2) in a suitable function space  $X$ .

In [13, 24] this approach is used to estimate the optimal parameters  $\lambda$  in the case when single data models  $\Phi_i$ ,  $i = 1, \dots, m$  are linearly combined, i.e., when

$$\Phi^\lambda(u, f) = \sum_{i=1}^m \lambda_i \Phi_i(u, f).$$

There, theoretical results showing existence of minima and the adjoint state for the problem (4.1)–(4.2) are shown and Newton-type methods are proposed for its efficient numerical solution. Similar results and algorithms are further studied in [25, 26] for the estimation of optimal parameters of higher-order regularisers (e.g., TGV) combined with Gaussian fidelity.

In the following we set  $m = 2$  and consider the problem of estimating the optimal parameters  $\lambda_1$  and  $\lambda_2$  in (2.5).

**General framework:** The non-smooth TV–IC bilevel problem reads:

$$\begin{aligned} & \min_{\lambda_1, \lambda_2 \geq 0} F(u_{\lambda_1, \lambda_2}) \\ \text{subject to } & u_{\lambda_1, \lambda_2} \in \operatorname{argmin}_{u \in \operatorname{BV}(\Omega)} |Du|(\Omega) + \Phi^{\lambda_1, \lambda_2}(u, f), \end{aligned} \quad (4.3)$$

where  $\Phi^{\lambda_1, \lambda_2}$  is the IC fidelity (1.3). We follow [13, 24, 25, 26] and introduce an appropriate smoothing of the TV semi-norm combined with a further quadratic smoothing. This is crucial for the following proofs and for the design of the gradient-based optimisation algorithms we intend to use.

For  $\epsilon \ll 1$ , we then consider the following regularised version of (4.3):

$$\begin{aligned} & \min_{\lambda_1, \lambda_2 \geq 0} F(u_{\lambda_1, \lambda_2}) \\ \text{subject to } & u_{\lambda_1, \lambda_2} \in \operatorname{argmin}_{u \in H^1(\Omega)} \frac{\epsilon}{2} \|u\|_{H^1(\Omega)}^2 + \|\nabla u\|_{\gamma, L^1(\Omega)} + \Phi_{\epsilon, \gamma}^{\lambda_1, \lambda_2}(u, f). \end{aligned} \quad (4.4)$$

Here, we denote by  $\|\nabla u\|_{\gamma, L^1(\Omega)} = \int_{\Omega} |\nabla u|_{\gamma} dx$ , with  $|\cdot|_{\gamma}$  being a smooth Huber-type regularisation depending on a parameter  $\gamma > 0$  whose  $C^1$ -derivative reads:

$$h_{\gamma}(z) := \begin{cases} \frac{z}{|z|} & \text{if } \gamma|z| - 1 \geq \frac{1}{2\gamma}, \\ \frac{z}{|z|} \left(1 - \frac{\gamma}{2} \left(1 - \gamma|z| + \frac{1}{2\gamma}\right)^2\right) & \text{if } \gamma|z| - 1 \in \left(-\frac{1}{2\gamma}, \frac{1}{2\gamma}\right), \\ \gamma z & \text{if } \gamma|z| - 1 \leq -\frac{1}{2\gamma}. \end{cases} \quad (4.5)$$

Note that  $|\cdot|_{\gamma}$  has one degree higher regularity than the classical Huber function  $\varphi$  in (2.2). This higher-order Huber-type smoothing has been previously used in [13, 26] for similar bilevel problems since it endows the problem (4.4) with further regularity, compare Theorem 4.6.

We then similarly regularise the IC fidelity term as:

$$\Phi_{\epsilon, \gamma}^{\lambda_1, \lambda_2}(u, f) := \min_{v \in L^2(\Omega)} \left\{ \mathcal{G}_{\epsilon, \gamma}^{\lambda_1, \lambda_2}(v, u, f) := \frac{\epsilon}{2} \|v\|_{L^2(\Omega)}^2 + \lambda_1 \|v\|_{\gamma, L^1(\Omega)} + \frac{\lambda_2}{2} \|f - u - v\|_{L^2(\Omega)}^2 \right\}, \quad (4.6)$$

where  $\|v\|_{\gamma, L^1(\Omega)}$  is defined analogously as above. For simplicity, from now on, we will assume that  $f \in L^2(\Omega)$ .

Inspired by [13, 26], we focus on two main choices of  $F$ . Namely, we consider the  $L^2$  cost corresponding to Peak Signal to Noise Ratio (PSNR) optimisation

$$F_{L^2}(u_{\lambda_1, \lambda_2}) := \|u_{\lambda_1, \lambda_2} - \tilde{u}\|_{L^2(\Omega)}^2, \quad \text{for training data } \tilde{u} \in L^2(\Omega), \quad (4.7)$$

and the Huberised TV cost, which is related to quality measures that are more adjusted to actual human perception, such as the Structural Similarity Index (SSIM):

$$F_{L^2 D}(u_{\lambda_1, \lambda_2}) := \|D(u_{\lambda_1, \lambda_2} - \tilde{u})\|_{\mathcal{M}, \gamma}, \quad \text{for training data } \tilde{u} \in \operatorname{BV}(\Omega). \quad (4.8)$$

Here,  $\|Du\|_{\mathcal{M}, \gamma} := \int_{\Omega} |\nabla u|_{\gamma} dx + |D^s u|(\Omega)$  so that the smooth Huber-type regularisation is applied on the absolutely continuous part of  $|Du|$ . For the abstract formulation of (4.4) in terms of a general  $F$ , we refer the reader to [26].

**Remark 4.1.** Note that if  $\lambda_1 = 0$  and/or  $\lambda_2 = 0$ , then  $\Phi^{\lambda_1, \lambda_2}(u, f) = 0$  for every  $u \in L^1(\Omega)$ . In that case every constant function  $c$  is a minimiser of the lower level problem of (4.3). Thus we trivially have  $\inf_{\lambda_1, \lambda_2 \geq 0} F(u_{\lambda_1, \lambda_2}) < \infty$  where  $F$  is taken to be either  $F_{L^2}$  or  $F_{L^2 D}$ .

## 4.2 Well-posedness of the TV–IC bilevel problem

We now discuss the well-posedness of the bilevel problems (4.3) and (4.4). For this type of problems, it is a common practice to impose an extra box constraint on the parameters  $\lambda_1, \lambda_2$  in order to ensure existence of solutions (see [13]), although generalisations to unbounded intervals are also possible [26]. The following proposition says that for problem (4.3) (i.e., with no upper bound constraints on the parameter domain) existence of solutions may fail.

**Proposition 4.2.** *There exist data  $f \in L^2(\Omega)$  and training data  $\tilde{u} \in L^2(\Omega)$  such that the non-smooth bilevel problem (4.3) does not have a solution for the cost function  $F_{L^2}$ .*

*Proof.* Take  $f$  to be any non-constant function in  $BV(\Omega) \cap L^2(\Omega)$  with the property that there exists  $\lambda^*$  such that

$$f = \operatorname{argmin}_{u \in BV(\Omega)} |Du|(\Omega) + \lambda_1 \|f - u\|_{L^1(\Omega)}$$

for all  $\lambda_1 \geq \lambda^*$ . We note that there are a plethora of such functions, in particular this holds for any one-dimensional function in  $BV(\Omega) \cap L^\infty(\Omega)$ , see also [18]. Now set  $\tilde{u} = f$ . Then from Proposition 2.10, we have that  $F_{L^2}(u_{\lambda_1, \lambda_2}) > 0$  for every  $\lambda_1, \lambda_2 \geq 0$ . Now fix  $\lambda_1 \geq \lambda^*$  and let  $\lambda_2^{(n)} \rightarrow \infty$  as  $n \rightarrow \infty$ . Then according to Corollary 2.4 we have that  $u_{\lambda_1, \lambda_2^{(n)}} \rightarrow f$  weakly\* in  $BV(\Omega)$  as  $n \rightarrow \infty$ . The function  $f$  can in fact be chosen such that this convergence is even uniform, see for instance the example in Proposition 3.1. Then it is clear that  $F_{L^2}(u_{\lambda_1, \lambda_2^{(n)}}) \rightarrow 0$ , which means that

$$\inf_{\lambda_1, \lambda_2 \geq 0} F_{L^2}(u_{\lambda_1, \lambda_2}) = 0.$$

Since  $F(u_{\lambda_1, \lambda_2}) > 0$  for every  $\lambda_1, \lambda_2 \geq 0$ , we have that for such choices of  $f$  and  $\tilde{u}$ , the bilevel problem (4.3) does not have solutions.  $\square$

Similarly, we can use the example of Proposition 3.1 to show that the bilevel problem (4.3) with  $F_{L^1_\gamma D}$  cost may not have a solution either, in the case when  $\lambda_1$  and  $\lambda_2$  are unbounded.

On the other hand, in the following Proposition we show that the existence of solutions of (4.3) is always guaranteed whenever box constraints on  $\lambda_1, \lambda_2$  are considered.

**Proposition 4.3** (Well-posedness of (4.3) with box constraints). *The bilevel problem (4.3) with the extra box constraints  $0 \leq \lambda_i \leq L_i$ ,  $L_i > 0$  for  $i = 1, 2$  admits an optimal solution  $(\hat{\lambda}_1, \hat{\lambda}_2)$  for both choices of cost functionals (4.7)–(4.8).*

*Proof.* Let  $(\lambda_1^{(n)}, \lambda_2^{(n)}) \in C := \{(\lambda_1, \lambda_2) : 0 \leq \lambda_1 \leq L_1, 0 \leq \lambda_2 \leq L_2\}$  be a minimising sequence for (4.3). Let us denote by  $u_n := u_{\lambda_1^{(n)}, \lambda_2^{(n)}}$  the corresponding solution to the lower level problem corresponding to the parameter pair  $(\lambda_1^{(n)}, \lambda_2^{(n)})$ .

First suppose that after some index  $n_0$  at least one of the terms  $(\lambda_1^{(n)})_{n \in \mathbb{N}}$  and  $(\lambda_2^{(n)})_{n \in \mathbb{N}}$  is zero. This means that  $u_n = c_n$  are constants for  $n \geq n_0$ . Now if  $F = F_{L^2}$ , due to the coercivity of this functional we have that  $(c_n)_{n \in \mathbb{N}}$  is bounded, so  $(u_n)_{n \in \mathbb{N}}$  is bounded in  $BV(\Omega)$ . If  $F = F_{L^1_\gamma}$  then it is obvious that  $F(u_n) = \|D\tilde{u}\|_{\mathcal{M}, \gamma}$  for  $n \geq 0$ . Thus in this case every constant function trivially solves the bilevel problem (4.3).

We can then assume that  $\lambda_1^{(n)}, \lambda_2^{(n)} > 0$  for every  $n \in \mathbb{N}$ . We claim again that the sequence  $u_n$  is bounded in  $BV(\Omega)$ . Indeed, we have that for every  $n \in \mathbb{N}$  we can bound the TV term as

$$|Du_n|(\Omega) \leq |Du_n|(\Omega) + \Phi^{\lambda_1^{(n)}, \lambda_2^{(n)}}(u_n, f) \leq \Phi^{\lambda_1^{(n)}, \lambda_2^{(n)}}(0, f) \leq \frac{L_2}{2} \|f\|_{L^2(\Omega)}^2.$$

To bound  $u_n$  in  $L^1(\Omega)$ , we separate the two cases depending on whether the sequence  $\frac{\lambda_1^{(n)}}{\lambda_2^{(n)}}$  is bounded or not.

If this sequence is bounded by some  $K > 0$ , we observe that  $u_n$  is also a minimiser of

$$\min_{u \in \text{BV}(\Omega)} \Phi^{1, \frac{\lambda_2^{(n)}}{\lambda_1^{(n)}}}(u, f) + \frac{1}{\lambda_1^{(n)}} |Du|(\Omega),$$

which implies that

$$\Phi^{1, \frac{\lambda_2^{(n)}}{\lambda_1^{(n)}}}(u_n, f) \leq \Phi^{1, \frac{\lambda_2^{(n)}}{\lambda_1^{(n)}}}(0, f).$$

Then we have the following successive bounds

$$\begin{aligned} \int_{\Omega} |f - u_n| dx &= \int_{|f - u_n| < \frac{\lambda_1^{(n)}}{\lambda_2^{(n)}}} |f - u_n| dx + \int_{|f - u_n| \geq \frac{\lambda_1^{(n)}}{\lambda_2^{(n)}}} \frac{\lambda_1^{(n)}}{2\lambda_2^{(n)}} dx \\ &+ \int_{|f - u_n| \geq \frac{\lambda_1^{(n)}}{\lambda_2^{(n)}}} |f - u_n| - \frac{\lambda_1^{(n)}}{2\lambda_2^{(n)}} dx \leq K|\Omega| + \frac{1}{2} \int_{\Omega} |f - u_n| dx + \Phi^{1, \frac{\lambda_2^{(n)}}{\lambda_1^{(n)}}}(u_n, f), \end{aligned}$$

whence we get:

$$\begin{aligned} \frac{1}{2} \int_{\Omega} |f - u_n| dx &\leq K|\Omega| + \Phi^{1, \frac{\lambda_2^{(n)}}{\lambda_1^{(n)}}}(0, f) \\ &\leq K|\Omega| + \int_{|f| \geq \frac{\lambda_1^{(n)}}{\lambda_2^{(n)}}} |f| dx + \frac{\lambda_2^{(n)}}{2\lambda_1^{(n)}} \int_{|f| < \frac{\lambda_1^{(n)}}{\lambda_2^{(n)}}} |f|^2 dx \leq K|\Omega| + \int_{\Omega} |f| dx + \frac{K}{2} |\Omega|. \end{aligned}$$

Suppose now that the sequence  $\frac{\lambda_1^{(n)}}{\lambda_2^{(n)}}$  is unbounded. This means that there exists a (non-relabelled) subsequence  $\frac{\lambda_2^{(n)}}{\lambda_1^{(n)}} \rightarrow 0$ . Since  $\lambda_1^{(n)}$  is bounded, this further implies that  $\lambda_2^{(n)} \rightarrow 0$ . Then from Proposition 2.9 we get that  $u_n \rightarrow f_{\Omega}$  weakly\* in  $\text{BV}(\Omega)$  and that, in particular,  $u_n$  is bounded in  $L^1(\Omega)$ . So in both cases we have that  $u_n$  is bounded in  $\text{BV}(\Omega)$ .

Having this combined with the boundedness of the sequence  $(\lambda_1^{(n)}, \lambda_2^{(n)})_{n \in \mathbb{N}}$  implies that we can then extract a further non-relabelled subsequence  $(\lambda_1^{(n)}, \lambda_2^{(n)}, u_n)_{n \in \mathbb{N}}$  converging weakly\* in  $\mathbb{R} \times \mathbb{R} \times \text{BV}(\Omega)$  to a limit point  $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{u})$ . In particular, this entails that  $u_n \rightarrow \hat{u}$  strongly in  $L^1(\Omega)$ . Similarly as in the proof of Proposition 2.3 one can now show that the sequence  $J(\cdot, \lambda_1^{(n)}, \lambda_2^{(n)})$   $\Gamma$ -converges to  $J(\cdot, \hat{\lambda}_1, \hat{\lambda}_2)$ , with respect to the strong topology in  $L^1(\Omega)$ , which means that  $\hat{u}$  is a minimiser of the lower level problem with parameters  $(\hat{\lambda}_1, \hat{\lambda}_2)$ .

Now, using [21, Corollary 7.20] we finally have:

$$\Phi^{\hat{\lambda}_1, \hat{\lambda}_2}(\hat{u}, f) + |D\hat{u}|(\Omega) = \lim_{n \rightarrow \infty} \Phi^{\lambda_1^{(n)}, \lambda_2^{(n)}}(u_n, f) + |Du_n|(\Omega),$$

which completes the proof. □

**Remark 4.4.** Similarly, one can prove the existence of solutions to the regularised bilevel problem (4.4). Note that in [25] similar techniques are used to prove analogous results for general regularisers and data fidelities. There, the authors proved also the outer-semicontinuity property of the solution map  $\mathcal{S}$  which guarantees that the minimisers of the regularised problem converge towards the minimisers of the one where  $\epsilon = 0$ .

The quadratic  $H^1$  and  $L^2$  regularisations in (4.4) and (4.6) are required to ensure the differentiability of the solution map, as we are going to highlight in the following. Note that in this section we make use of the formulation (4.4) where the two variables  $u_{\lambda_1, \lambda_2}, v_{\lambda_1, \lambda_2}$  are treated jointly so that the lower-level problem actually reads:

$$\begin{aligned} (u_{\lambda_1, \lambda_2}, v_{\lambda_1, \lambda_2}) \in \operatorname{argmin}_{\substack{u \in H^1(\Omega) \\ v \in L^2(\Omega)}} & \frac{\epsilon}{2} \left( \|u\|_{H^1(\Omega)}^2 + \|v\|_{L^2(\Omega)}^2 \right) + \|\nabla u\|_{\gamma, L^1(\Omega)} \\ & + \lambda_1 \|v\|_{\gamma, L^1(\Omega)} + \frac{\lambda_2}{2} \|f - u - v\|_{L^2(\Omega)}^2. \end{aligned} \quad (4.9)$$

An alternative analysis could exploit the characterisation of the IC data fidelity term given by Proposition 2.1 and consider the minimisation over  $u$  only. In such case, we believe that only an  $H^1$ -regularisation on  $u$  would be enough for the following proofs. Here, however, we stick with the joint approach to compare our results with the ones derived formally in [14].

Note that whenever  $u_{\lambda_1, \lambda_2}$  is given, one can compute the corresponding  $v_{\lambda_1, \lambda_2}$  by simply solving the optimisation problem (4.6). In particular, the following proposition makes explicit a property of  $v_{\lambda_1, \lambda_2}$  which will be needed in the following. We refer the reader to [14, Remark 2.1] for a similar characterisation in the case non-regularised case.

**Proposition 4.5.** Let  $u, f$  in  $L^2(\Omega)$ ,  $0 \leq \lambda_i \leq L_i$  with  $L_i > 0$ ,  $i = 1, 2$  and  $\epsilon > 0$ . Let  $v_{\lambda_1, \lambda_2}$  the minimiser of the functional  $\mathcal{G}_{\epsilon, \gamma}^{\lambda_1, \lambda_2}(\cdot, u, f)$  defined in (4.6). There holds:

$$v_{\lambda_1, \lambda_2} = v_{\lambda_1, \lambda_2}(u) = \operatorname{argmin}_{v \in L^2(\Omega)} \mathcal{G}_{\epsilon, \gamma}^{\lambda_1, \lambda_2}(v, u, f) = \operatorname{prox}_{\frac{\lambda_1}{\epsilon + \lambda_2} \|\cdot\|_{\gamma, L^1(\Omega)}} \left( \frac{\lambda_2(f - u)}{\epsilon + \lambda_2} \right), \quad (4.10)$$

where for any  $z \in L^2(\Omega)$ ,  $\operatorname{prox}_{\tau g}(z)$  denotes the proximal-mapping operator in  $L^2(\Omega)$  of the function  $g$  with parameter  $\tau$ . In particular,  $v_{\lambda_1, \lambda_2} : z \mapsto \operatorname{prox}_{\frac{\lambda_1}{\epsilon + \lambda_2} \|\cdot\|_{\gamma, L^1(\Omega)}}(z)$  is a firmly non-expansive operator and it is therefore 1-Lipschitz continuous, i.e.:

$$\|v_{\lambda_1, \lambda_2}(u_1) - v_{\lambda_1, \lambda_2}(u_2)\|_{L^2(\Omega)} \leq \|u_1 - u_2\|_{L^2(\Omega)}, \quad \text{for all } u_1, u_2 \in L^2(\Omega).$$

*Proof.* Straightforward calculations give:

$$\begin{aligned} v_{\lambda_1, \lambda_2} &= \operatorname{argmin}_{v \in L^2(\Omega)} \frac{\epsilon}{2} \|v\|_{L^2(\Omega)}^2 + \lambda_1 \|v\|_{\gamma, L^1(\Omega)} + \frac{\lambda_2}{2} \|f - u - v\|_{L^2(\Omega)}^2 \\ &= \operatorname{argmin}_{v \in L^2(\Omega)} \|v\|_{\gamma, L^1(\Omega)} + \frac{\epsilon + \lambda_2}{2\lambda_1} \|v\|_{L^2(\Omega)}^2 + \frac{1}{2\frac{\lambda_1}{\lambda_2} \left(1 + \frac{\epsilon}{\lambda_2}\right)} \|f - u\|_{L^2(\Omega)}^2 - \frac{\lambda_2}{\lambda_1} \int_{\Omega} (f - u)v \, dx \\ &= \operatorname{argmin}_{v \in L^2(\Omega)} \|v\|_{\gamma, L^1(\Omega)} + \frac{1}{2\frac{\lambda_1}{\epsilon + \lambda_2}} \left\| \frac{\lambda_2(f - u)}{\epsilon + \lambda_2} - v \right\|_{L^2(\Omega)}^2 \\ &= \operatorname{prox}_{\frac{\lambda_1}{\epsilon + \lambda_2} \|\cdot\|_{\gamma, L^1(\Omega)}} \left( \frac{\lambda_2(f - u)}{\epsilon + \lambda_2} \right). \end{aligned}$$

The firm non-expansiveness property (4.10) follows then directly from [6, Proposition 12.27].  $\square$



### 4.3 Optimality system

We now study in more detail the bilevel problem (4.4) with (4.9) as lower level problem and prove the existence of Lagrange multipliers by deriving the optimality system characterising its stationary points. As a by product, we find an easy formula to compute the gradient of the cost functional in terms of its adjoint state, which simplifies the design of the gradient-based algorithm which we are going to use to solve (4.4) in an efficient way.

We start defining the Hilbert space  $\mathcal{H} := H^1(\Omega) \times L^2(\Omega)$  endowed with the scalar product  $(z, w)_{\mathcal{H}} := (z_1, w_1)_{H^1(\Omega)} + (z_2, w_2)_{L^2(\Omega)}$  for all  $z, w \in \mathcal{H}$ . We further denote by  $\mathcal{S} : \mathbb{R}^2 \rightarrow \mathcal{H}$  the solution map  $\mathcal{S} : (\lambda_1, \lambda_2) \mapsto (u_{\lambda_1, \lambda_2}, v_{\lambda_1, \lambda_2})$  which assigns to the optimal parameters  $(\lambda_1, \lambda_2)$  the corresponding solution pair of (4.9). To avoid heavy notations, we will omit the explicit dependence of the pair  $(u, v)$  on the parameters  $(\lambda_1, \lambda_2)$  unless explicitly needed. Note that we take the whole space  $\mathbb{R}^2$  as differentiability set although from an imaging point of view the use of negative parameters clearly does not make sense. However, the following differentiability result holds in such case as well.

Recalling the definition of  $h_\gamma$  in (4.5), we have that in correspondence with an optimal pair  $y := (u, v) \in \mathcal{H}$ , we have that the following variational equality holds true for all test functions  $\Psi := (\psi_1, \psi_2) \in \mathcal{H}$ :

$$\epsilon(y, \Psi)_{\mathcal{H}} + \int_{\Omega} h_\gamma(\nabla u) \nabla \psi_1 \, dx + \lambda_1 \int_{\Omega} h_\gamma(v) \psi_2 \, dx + \lambda_2 \int_{\Omega} (u + v - f)(\psi_1 + \psi_2) \, dx = 0. \quad (4.11)$$

We now prove the main differentiability result.

**Theorem 4.6** (Fréchet differentiability of the solution map). *The solution operator  $\mathcal{S} : \mathbb{R}^2 \rightarrow \mathcal{H}$  which assigns to each parameter pair  $(\lambda_1, \lambda_2)$  the element  $y := (u, v) = \mathcal{S}(\lambda_1, \lambda_2)$ , solution of the TV–IC denoising problem (4.4)–(4.6) is Fréchet differentiable. In particular, for any  $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$  its Fréchet derivative is the unique solution  $z := \mathcal{S}'(\lambda_1, \lambda_2)[\theta_1, \theta_2] \in \mathcal{H}$  of the following linearised equation:*

$$\begin{aligned} \epsilon(z, \Psi)_{\mathcal{H}} + \int_{\Omega} h'_\gamma(\nabla u) \nabla z_1 \nabla \psi_1 \, dx + \lambda_1 \int_{\Omega} h'_\gamma(v) z_2 \psi_2 \, dx + \theta_1 \int_{\Omega} h_\gamma(v) \psi_2 \, dx \\ + \lambda_2 \int_{\Omega} (z_1 + z_2)(\psi_1 + \psi_2) \, dx + \theta_2 \int_{\Omega} (u + v - f)(\psi_1 + \psi_2) \, dx = 0, \end{aligned} \quad (4.12)$$

for all  $\Psi = (\psi_1, \psi_2) \in \mathcal{H}$ .

*Proof.* Thanks to the ellipticity of the scalar product in  $\mathcal{H}$  and the monotonicity of  $h_\gamma$ , existence and uniqueness of  $z$  of (4.12) are guaranteed by Lax-Milgram theorem.

Now, we want to show that  $z$  is the Fréchet derivative of  $\mathcal{S}$ . To do that, given  $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$  let us define  $y^+ = (u^+, v^+) := \mathcal{S}(\lambda_1 + \theta_1, \lambda_2 + \theta_2)$ . We aim to show that  $\xi = (\xi_1, \xi_2) := y^+ - y - z \in \mathcal{H}$  satisfies  $\|\xi\|_{\mathcal{H}} = o(|\theta|)$ .

Writing (4.11) for  $y$  and  $y^+$ , and (4.12) for  $z$  and combining them together, we get that for every  $\Psi \in \mathcal{H}$

there holds:

$$\begin{aligned} \epsilon(\xi, \Psi)_{\mathcal{H}} &+ \int_{\Omega} (h_{\gamma}(\nabla u^+) - h_{\gamma}(\nabla u)) \nabla \psi_1 \, dx - \int_{\Omega} h'_{\gamma}(\nabla u) \nabla z_1 \nabla \psi_1 \, dx \\ &+ \lambda_1 \int_{\Omega} (h_{\gamma}(v^+) - h_{\gamma}(v)) \psi_2 \, dx + \theta_1 \int_{\Omega} (h_{\gamma}(v^+) - h_{\gamma}(v)) \psi_2 \, dx - \lambda_1 \int_{\Omega} h'_{\gamma}(v) z_2 \psi_2 \, dx \\ &+ \lambda_2 \int_{\Omega} (\xi_1 + \xi_2)(\psi_1 + \psi_2) \, dx + \theta_2 \int_{\Omega} ((u^+ - u) + (v^+ - v)) (\psi_1 + \psi_2) \, dx = 0. \end{aligned}$$

We now add and subtract the terms:

$$\int_{\Omega} h'_{\gamma}(\nabla u) (\nabla(u^+ - u)) \nabla \psi_1 \, dx, \quad \text{and} \quad \lambda_1 \int_{\Omega} h'_{\gamma}(v) (v^+ - v) \psi_2 \, dx,$$

thus getting:

$$\begin{aligned} \epsilon(\xi, \Psi)_{\mathcal{H}} &+ \int_{\Omega} h'_{\gamma}(\nabla u) \nabla \xi_1 \nabla \psi_1 \, dx + \lambda_1 \int_{\Omega} h'_{\gamma}(v) \xi_2 \psi_2 \, dx + \lambda_2 \int_{\Omega} (\xi_1 + \xi_2)(\psi_1 + \psi_2) \, dx \\ &= - \int_{\Omega} (h_{\gamma}(\nabla u^+) - h_{\gamma}(\nabla u) - h'_{\gamma}(\nabla u)(\nabla(u^+ - u))) \nabla \psi_1 \, dx \\ &- \lambda_1 \int_{\Omega} (h_{\gamma}(v^+) - h_{\gamma}(v) - h'_{\gamma}(v)(v^+ - v)) \nabla \psi_2 \, dx - \theta_1 \int_{\Omega} (h_{\gamma}(v^+) - h_{\gamma}(v)) \psi_2 \, dx \\ &- \theta_2 \int_{\Omega} ((u^+ - u) + (v^+ - v)) (\psi_1 + \psi_2) \, dx, \quad \text{for all } \Psi \in \mathcal{H}. \end{aligned}$$

We now choose  $\Psi = \xi$ . Under this choice, by monotonicity of  $h'_{\gamma}$  we have that the last three terms on the left-hand side become non-negative. We further deduce:

$$\begin{aligned} \|\xi\|_{\mathcal{H}} &\leq C \left( \|h_{\gamma}(\nabla u^+) - h_{\gamma}(\nabla u) - h'_{\gamma}(\nabla u)(\nabla(u^+ - u))\|_{L^2(\Omega)} \right. \\ &+ |\lambda_1| \|h_{\gamma}(v^+) - h_{\gamma}(v) - h'_{\gamma}(v)(v^+ - v)\|_{L^2(\Omega)} + |\theta_1| \|h_{\gamma}(v^+) - h_{\gamma}(v)\|_{L^2(\Omega)} \\ &\left. + |\theta_2| \|u^+ - u\|_{L^2(\Omega)} + |\theta_2| \|v^+ - v\|_{L^2(\Omega)} \right). \end{aligned}$$

where  $C$  is a generic positive and finite constant which may change from line to line. By the differentiability and Lipschitz continuity of  $h_{\gamma}$  and  $h'_{\gamma}$ , we have:

$$\begin{aligned} \|\xi\|_{\mathcal{H}} &\leq C \left( o(\|\nabla u^+ - \nabla u\|_{L^2(\Omega)}) + |\lambda_1| o(\|v^+ - v\|_{L^2(\Omega)}) \right. \\ &\left. + |\theta_2| \|u^+ - u\|_{L^2(\Omega)} + (|\theta_1| + |\theta_2|) \|v^+ - v\|_{L^2(\Omega)} \right). \end{aligned} \quad (4.13)$$

We now focus on the terms depending on the difference between  $v^+ = v_{\lambda_1 + \theta_1, \lambda_2 + \theta_2}(u^+)$  and  $v = v_{\lambda_1, \lambda_2}(u)$ . By triangle inequality we have:

$$\begin{aligned} \|v^+ - v\|_{L^2(\Omega)} &\leq \|v_{\lambda_1 + \theta_1, \lambda_2 + \theta_2}(u^+) - v_{\lambda_1 + \theta_1, \lambda_2 + \theta_2}(u)\|_{L^2(\Omega)} + \|v_{\lambda_1 + \theta_1, \lambda_2 + \theta_2}(u) - v_{\lambda_1, \lambda_2}(u)\|_{L^2(\Omega)} \\ &\leq \|u^+ - u\|_{L^2(\Omega)} + |\theta|, \end{aligned}$$

where the last inequality follows from Proposition 4.5 and from the continuity property of the proximal mapping with respect to its parameter which can be easily checked in our case recalling that  $|h_{\gamma}(\cdot)| \leq 1$ . We then deduce in (4.13) that:

$$\|\xi\|_{\mathcal{H}} \leq C \left( o(|\theta|) + o(\|u^+ - u\|_{H^1(\Omega)}) \right) \quad (4.14)$$

Thus, to conclude it only remains to show that  $o(\|u^+ - u\|_{H^1(\Omega)}) = o(|\theta|)$ . To do that, we use standard Sobolev embeddings and the regularity result of Gröger for second-order systems [29, Theorem 1] and get that for  $p > 2$ :

$$\begin{aligned} \|u^+ - u\|_{H^1(\Omega)} &\leq \|u^+ - u\|_{W^{1,p}(\Omega)} \leq C|\theta| \left( \|\operatorname{div}(h_\gamma(\nabla u))\|_{W^{-1,p}(\Omega)} + \|h_\gamma(\nabla u)\|_{W^{-1,p}(\Omega)} \right) \\ &\leq C|\theta| \|h_\gamma(\nabla u)\|_{L^\infty(\Omega)} \leq C|\theta|, \end{aligned}$$

since  $|h_\gamma(\cdot)| \leq 1$ . Combining with (4.14) this finishes the proof.  $\square$

**Remark 4.7.** *The regularity result by Gröger is a classical argument for the proof of Fréchet differentiability in similar bilevel problems (see, e.g., [13]). Note, however, that the original result in [29] was proved for  $C^2$ -regular domains, while its extension to convex Lipschitz domains (such as image domains) has been proved by Dauge in [22].*

We now prove the existence and uniqueness of the adjoint state of the problem (4.4).

**Theorem 4.8** (Adjoint equation). *Let  $(u, v) \in \mathcal{H}$ . There exists a unique solution  $\Pi := (p_1, p_2) \in \mathcal{H}$  to the adjoint PDE:*

$$\begin{aligned} \epsilon(\Pi, W)_{\mathcal{H}} + \int_{\Omega} h'_\gamma(\nabla u) \nabla w_1 \nabla p_1 \, dx + \lambda_1 \int_{\Omega} h'_\gamma(v) w_2 p_2 \, dx \\ + \lambda_2 \int_{\Omega} (p_1 + p_2) w_1 \, dx + \lambda_2 \int_{\Omega} (p_1 + p_2) w_2 \, dx = - \int_{\Omega} F'(u) w_1 \, dx, \end{aligned} \quad (4.15)$$

for any  $W := (w_1, w_2) \in \mathcal{H}$ .

*Proof.* For  $w \in \mathcal{H}$ , let us consider the following bilinear form on  $\mathcal{H} \times \mathcal{H}$ :

$$\begin{aligned} a(\Pi, W) := \epsilon(\Pi, W)_{\mathcal{H}} + \int_{\Omega} h'_\gamma(\nabla u) \nabla w_1 \nabla p_1 \, dx \\ + \lambda_1 \int_{\Omega} h'_\gamma(v) w_2 p_2 \, dx + \lambda_2 \int_{\Omega} (p_1 + p_2) w_1 \, dx + \lambda_2 \int_{\Omega} (p_1 + p_2) w_2 \, dx. \end{aligned}$$

The form  $a(\cdot, \cdot)$  is trivially symmetric and coercive, since by taking  $W = \Pi$ , we get:

$$a(\Pi, \Pi) = \epsilon \|\Pi\|_{\mathcal{H}}^2 + \int_{\Omega} h'_\gamma(\nabla u) \nabla p_1 \nabla p_1 + \lambda_1 \int_{\Omega} h'_\gamma(v) p_2 p_2 \, dx + \lambda \int_{\Omega} (p_1 + p_2)^2 \, dx \geq \epsilon \|\Pi\|_{\mathcal{H}}^2.$$

by monotonicity of  $h'_\gamma$ . By Lax-Milgram theorem, we infer that there exists a unique solution of (4.15).  $\square$

Note that by taking  $w_2 = 0$  in (4.15), we get the optimality condition for  $p_1$ , i.e.:

$$\epsilon(p_1, w_1)_{H^1(\Omega)} + \int_{\Omega} h'_\gamma(\nabla u) \nabla w_1 \nabla p_1 \, dx + \lambda_2 \int_{\Omega} (p_1 + p_2) w_1 \, dx = - \int_{\Omega} F'(u) w_1 \, dx,$$

for any  $w_1 \in H^1(\Omega)$ . Similarly, for  $w_1 = 0$ , we get the optimality condition for  $p_2$ :

$$\epsilon \int_{\Omega} w_2 p_2 \, dx + \lambda_1 \int_{\Omega} h'_\gamma(v) w_2 p_2 \, dx + \lambda_2 \int_{\Omega} (p_1 + p_2) w_2 \, dx = 0,$$

for any  $w_2 \in L^2(\Omega)$ .

Finally, we now combine the results above to derive the optimality system of the bilevel problem (4.4). We recall that by  $y = (u, v) \in \mathcal{H}$  we denote the solution pair.

**Theorem 4.9** (Optimality system). *Let  $(\bar{\lambda}_1, \bar{\lambda}_2) \in \mathbb{R}^+ \times \mathbb{R}^+$  an optimal solution of the problem (4.4). Then, there exists a Lagrange multiplier  $\Pi := (p_1, p_2) \in \mathcal{H}$  and  $\mu_1, \mu_2 \in \mathbb{R}^+$  such that the following system holds:*

$$\begin{aligned}
& \epsilon(y, \Psi)_{\mathcal{H}} + \int_{\Omega} h_{\gamma}(\nabla u) \nabla \psi_1 dx + \lambda_1 \int_{\Omega} h_{\gamma}(v) \psi_2 dx \\
& + \lambda_2 \int_{\Omega} (u + v - f)(\psi_1 + \psi_2) dx = 0, \quad \text{for all } \Psi = (\psi_1, \psi_2) \in \mathcal{H}, \\
& \epsilon(\Pi, W)_{\mathcal{H}} + \int_{\Omega} h'_{\gamma}(\nabla u) \nabla w_1 \nabla p_1 dx + \lambda_1 \int_{\Omega} h'_{\gamma}(v) w_2 p_2 dx \\
& + \lambda_2 \int_{\Omega} (p_1 + p_2)(w_1 + w_2) dx = - \int_{\Omega} F'(u) w_1 dx, \quad \text{for all } W = (w_1, w_2) \in \mathcal{H}, \\
& \mu_1 := \int_{\Omega} h'_{\gamma}(v) p_2 dx, \quad \mu_2 := \int_{\Omega} (f - v - u)(p_1 + p_2) dx, \\
& \mu_1 \geq 0, \quad \mu_2 \geq 0, \quad \mu_1 \cdot \bar{\lambda}_1 = \mu_2 \cdot \bar{\lambda}_2 = 0.
\end{aligned} \tag{4.16}$$

*Proof.* We can write the bilevel problem (4.4) in a reduced form as:

$$\min_{\lambda_1, \lambda_2 \geq 0} \mathcal{F}(\lambda_1, \lambda_2) := F(u_{\lambda_1, \lambda_2}).$$

Using [57, Theorem 3.1], we deduce the existence of multipliers  $\mu_1, \mu_2 \in \mathbb{R}^+$  such that:

$$\begin{aligned}
\mu_1 &= \nabla_{\lambda_1} \mathcal{F}(\bar{\lambda}_1, \bar{\lambda}_2), \\
\mu_2 &= \nabla_{\lambda_2} \mathcal{F}(\bar{\lambda}_1, \bar{\lambda}_2), \\
\mu_1 &\geq 0, \quad \mu_2 \geq 0, \quad \mu_1 \cdot \bar{\lambda}_1 = \mu_2 \cdot \bar{\lambda}_2 = 0.
\end{aligned}$$

Computing the gradient of  $\mathcal{F}$  by using the chain rule we get:

$$\nabla \mathcal{F}(\lambda_1, \lambda_2)[\theta_1, \theta_2] = (F'(u_{\lambda_1, \lambda_2}), \mathcal{S}'(\lambda_1, \lambda_2)[\theta_1, \theta_2])_{L^2(\Omega)} = \int_{\Omega} F'(u_{\lambda_1, \lambda_2}) z dx$$

where  $z \in \mathcal{H}$  is the linearised state provided by Theorem 4.6. Theorem 4.8 ensures that there exist a Lagrange multiplier  $\Pi := (p_1, p_2)$  satisfying the adjoint equation (4.15), which entails:

$$\begin{aligned}
\int_{\Omega} F'(u_{\lambda_1, \lambda_2}) z dx &= - \epsilon(\Pi, z)_{\mathcal{H}} - \int_{\Omega} h'_{\gamma}(\nabla u) \nabla z_1 \nabla p_1 dx \\
&\quad - \lambda_1 \int_{\Omega} h'_{\gamma}(v) z_2 p_2 dx - \lambda_2 \int_{\Omega} (p_1 + p_2)(z_1 + z_2) dx \\
&= \theta_1 \int_{\Omega} h_{\gamma}(v) p_2 dx + \theta_2 \int_{\Omega} (u + v - f)(p_1 + p_2) dx,
\end{aligned}$$

which completes the proof.  $\square$

**Remark 4.10.** *Theorem 4.9 provides also the following handy formula for the computation of the gradient of the bilevel problem (4.4) with data fidelity (4.6) in a reduced form:*

$$\nabla \mathcal{F}(\lambda_1, \lambda_2)[\theta_1, \theta_2] = \theta_1 \int_{\Omega} h_{\gamma}(v) p_2 dx + \theta_2 \int_{\Omega} (u + v - f)(p_1 + p_2) dx. \tag{4.17}$$

**Remark 4.11.** *Compared to the optimality system derived in [14, Section 7] via Lagrangian formalism, we note that the optimality system (4.16) presents an additional quadratic  $\epsilon$ -regularisation on  $v$ . This is needed in the proof of Theorem 4.6 to get a uniform, finite estimate for  $\xi_2$  not depending on  $\lambda_2$  which may be zero.*

## 5 Numerical experiments

In this section we report some numerical results on the computation of the optimal parameters  $(\bar{\lambda}_1, \bar{\lambda}_2)$  of the TV–IC model by means of the bilevel optimisation strategy described in the previous section. For given training images with a mixture of Gaussian and Salt & Pepper noise with various intensities, we describe in Algorithm 1 the main steps to compute  $(\bar{\lambda}_1, \bar{\lambda}_2)$  by solving the optimality system (4.16) via a second-order BFGS optimisation approach. A general review of second-order numerical methods for PDE-constrained optimisation models can be found in the book [23], while more details on the numerical realisation of similar bilevel models can be found in [13, Section 8.3] and in [26, Section 4.1].

In the following numerical computations:

- We consider test images of size  $N \times N \equiv 256 \times 256$  pixels. The differential operators are discretised using finite difference schemes with mesh step size  $h = 1/N$ . Standard forward/backward differences are used for the discretisation of the divergence/gradient operator, respectively.
- For illustrative purposes, we report the results obtained using a training set consisting of one training pair  $(\tilde{u}, \tilde{f})$  only. We recall that  $\tilde{f}$  represents the noisy version of  $\tilde{u}$  corrupted by a mixture of Gaussian and Salt & Pepper noise of various intensities which we will specify in each case. The efficient extension to multiple constraints can be done similarly as in [13, 15].
- The lower-level regularised TV–IC problem (4.4) is solved by means of the SemiSmooth Newton (SSN) algorithm described in [14] with a warm start. The regularisation parameters are chosen as  $\epsilon = 10^{-10}$  and  $\gamma = 10^3$ . The algorithm is stopped if either the difference between two consecutive iterates is below  $\text{tol} = 10^{-6}$  or if the maximum number of iterations  $\text{maxiter} = 35$  is reached.
- For the outer BFGS iterations, an Armijo line-search with parameter  $\eta = 10^{-4}$  is employed together with a curvature verification. The Armijo rule:

$$\mathcal{F}(\boldsymbol{\lambda}_k + \alpha_k \mathbf{d}_k) - \mathcal{F}(\boldsymbol{\lambda}_k) \leq \eta \alpha_k \nabla \mathcal{F}(\boldsymbol{\lambda}_k)^\top \mathbf{d}_k$$

is checked at any iteration  $k \geq 2$ . Here  $\boldsymbol{\lambda}_k = (\lambda_k^1, \lambda_k^2)$  stands for the parameter pair updated along the iterations,  $\mathbf{d}_k$  for the quasi-Newton descent direction and  $\alpha_k$  for the line step. The expression of  $\nabla \mathcal{F}$  is given in Remark 4.10. The outer algorithm is stopped when the maximum between the norm of the gradient of the cost functional and the difference of two subsequent iterates is smaller than  $\text{tol}_1 = 10^{-6}$ .

- The adjoint equation is solved by means of standard sparse linear solvers.
- Our validation images are taken from the public domain.

In Figure 5 we report a numerical experiment confirming the effectiveness of the bilevel optimisation approach on images corrupted by a mixture of Gaussian and Salt & Pepper noise with Gaussian variance  $\sigma^2 = 0.01$  and percentage of missing pixels  $d = 10\%$ . We report the result obtained with respect to both the  $L^2$  cost functional (4.7) and the Huberised  $L^1$  gradient cost (4.8). As observed in [26], we remark that minimising with respect to the  $L^2$  cost is indeed equivalent to PSNR optimisation, while the optimisation with respect to the Huberised  $L^1$  gradient cost produces better visual results which is similar to optimising the SSIM.

**Algorithm 1** Bilevel optimisation algorithm for computing optimal  $\lambda_1$  and  $\lambda_2$  in (4.4)

---

**Input:** Training pair  $(\tilde{u}, \tilde{f})$ . Regularisation parameters:  $\gamma \gg 1, \epsilon \ll 1$ .  
**Output:** Optimal parameters  $\bar{\lambda}_1$  and  $\bar{\lambda}_2$ .  
**Initialise:**  $\lambda_1^0, \lambda_2^0, n = 1$ .  
**while** not converging **do**  
 SSN algorithm to compute  $(u^n, v^n)$  by solving (4.11) with parameters  $(\lambda_1^{(n)}, \lambda_2^{(n)})$ ;  
 compute adjoint states  $(p_1^n, p_2^n)$ ;  
 compute  $F'(u^n)$  using (4.17);  
 BFGS update to compute new  $(\lambda_1^{n+1}, \lambda_2^{n+1})$ ;  
 Armijo line-search with parameter  $\eta \ll 1$ ;  
 $n = n + 1$ ;  
**end while**

---

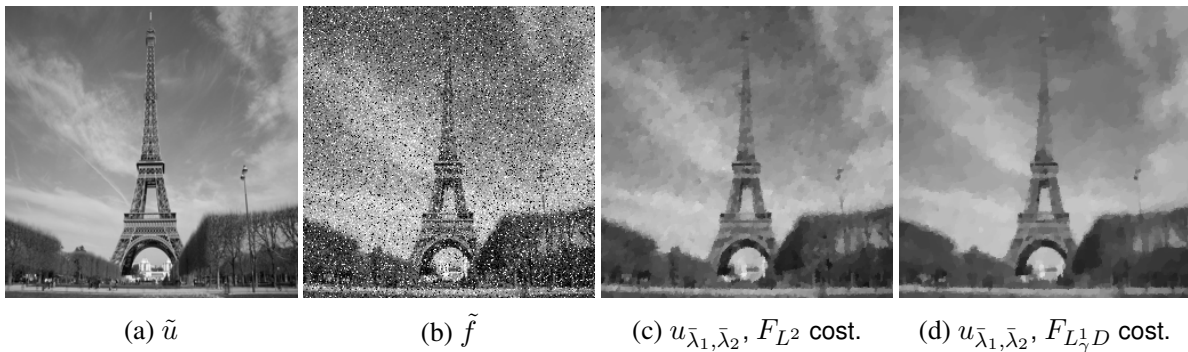


Figure 5: Optimal TV-IC denoising results for initial guess  $(\lambda_1^0, \lambda_2^0) = (1, 10)$  w.r.t.  $F_{L^2}$  and  $F_{L^1_D}$  costs (4.7)–(4.8). Noisy image corrupted with Gaussian noise with variance  $\sigma^2 = 0.01$  and percentage of missing pixels  $d = 10\%$ . 5b Noisy image  $\tilde{f}$ : PSNR=14.40 dB, SSIM=0.17. 5c Optimal  $u_{\bar{\lambda}_1, \bar{\lambda}_2}$  w.r.t.  $F_{L^2}$  cost: PSNR=28.35 dB, SSIM=0.81. 5d Optimal  $u_{\bar{\lambda}_1, \bar{\lambda}_2}$  w.r.t.  $F_{L^1_D}$  cost: PSNR=27.91 dB, SSIM=0.83.

In order to validate numerically the theoretical insights given by the analysis performed in Section 2, we report in the following Figure 6 a plot of the optimal parameters  $(\bar{\lambda}_1, \bar{\lambda}_2)$  computed by solving the bilevel system (4.16) with  $L^2$  cost (4.7) and in correspondence of a training pair  $(\tilde{u}, \tilde{f}_\theta)$  where the noisy image is corrupted by a mixture of Gaussian and Salt & Pepper noise with varying intensity. Namely, for  $\theta \in [0, 1]$ , we corrupt the training image  $\tilde{u}$  in 7a with Gaussian noise with distribution  $\mathcal{N}(0, \theta\sigma^2)$ ,  $\sigma^2 = 0.005$  and Salt & Pepper noise with a percentage of corrupted pixels equal to  $d = (1 - \theta)10\%$ . Consequently, the noisy image  $\tilde{f}_\theta$  is corrupted by pure impulsive noise for  $\theta = 0$  (Figure 7b), by pure Gaussian noise for  $\theta = 1$  (Figure 7f) and by a mixture of the two for  $\theta \in (0, 1)$  (Figures 7c-7e). As suggested by the theory (see, in particular, Proposition 2.3 and Corollary 2.4), we observe that when  $\theta = 0$  the bilevel strategy selects a large optimal parameter  $\bar{\lambda}_2$ , enforcing a TV- $L^1$  denoising model which is well known to be optimal for this type of noise (see, e.g., [28, 44]). Furthermore, by Corollary 2.4 we also have that the lower-level solution of the mixed noise problem approximates (in the sense of  $\Gamma$ -convergence) a solution of the corresponding TV- $L^1$  model. It is then interesting to notice that in the case  $\theta = 1$  the bilevel optimisation strategy does not enforce a TV- $L^2$  model (i.e., a large  $\bar{\lambda}_1$ ), but rather a combination of TV- $L^1$  and TV- $L^2$ . This might be an indication that in practice the TV- $L^1$  model works well also in the case of pure Gaussian noise removal. This is reflected in our experiment where the estimated optimal data model turns out to be indeed a combination of the two discrepancies. Note that the two parameters  $\lambda_1$  and  $\lambda_2$  scale

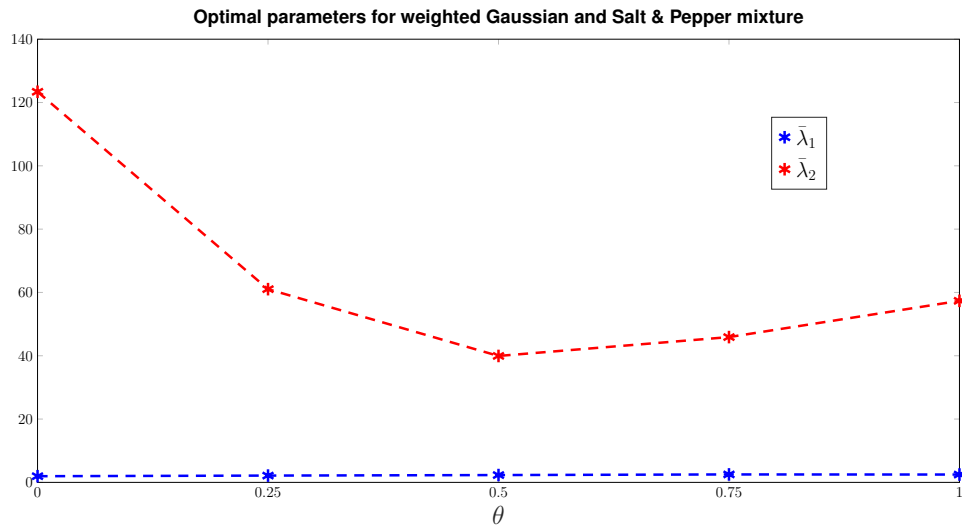


Figure 6: Optimal parameters  $(\bar{\lambda}_1, \bar{\lambda}_2)$  computed by solving the bilevel system (4.16) with  $L^2$  cost and training pair  $(\tilde{u}, \tilde{f}_\theta)$  depending on a parameter  $\theta \in [0, 1]$  which controls the amount of Gaussian and Salt & Pepper noise in the data, see Figure 7. For  $\theta = 0$  only Salt & Pepper noise is present and the model computes optimal parameters enforcing TV- $L^1$  denoising model. For  $0 < \theta < 1$  the data discrepancies are weighted by smaller parameters  $\bar{\lambda}_1$  and  $\bar{\lambda}_2$ . For each  $\theta$  the bilevel Algorithm 1 is initialised with  $(\lambda_1^0, \lambda_2^0) = (1, 1)$ .

differently, with  $\bar{\lambda}_1$  only slightly varying across the different simulations.

We report in Table 2 the numerical values of  $(\bar{\lambda}_1, \bar{\lambda}_2)$  and the corresponding PSNR values of the noisy images  $\tilde{f}_\theta$  and of the optimal reconstructions  $u_{\bar{\lambda}_1, \bar{\lambda}_2}$ .

$\theta$	PSNR $\tilde{f}_\theta$	$\bar{\lambda}_1$	$\bar{\lambda}_2$	PSNR $u_{\bar{\lambda}_1, \bar{\lambda}_2}$
0	15.41 dB	1.95	123.39	29.71 dB
0.25	16.19 dB	2.15	61.04	24.66 dB
0.5	17.49 dB	2.31	39.94	25.35 dB
0.75	19.28 dB	2.51	45.90	24.01 dB
1	22.74 dB	2.47	57.35	24.25 dB

Table 2

## 6 Conclusions and outlook

In this paper we have presented a fine analysis of the TV–IC denoising model originally proposed in [14] for mixed Gaussian and Salt & Pepper noise removal. Our study started with a characterisation of the IC data-discrepancy as a Huber-regularised  $L^1$  discrepancy. We then studied in detail the asymptotic behaviour of the solutions of the TV–IC model using  $\Gamma$ -convergence arguments, showing that the solutions of the single TV- $L^1$  and TV- $L^2$  denoising models can be retrieved in the limit as the parameters tend to infinity. We gained more insights on the model by calculating some exact solutions for simple one-dimensional data functions. Using these theoretical results we then formulated and rigorously analysed a bilevel optimisation approach in function spaces for the estimation of the optimal parameters of the model. With the use of a counterexample motivated by our theoretical work, we showed that box constraints on the parameters are indeed necessary to obtain existence.

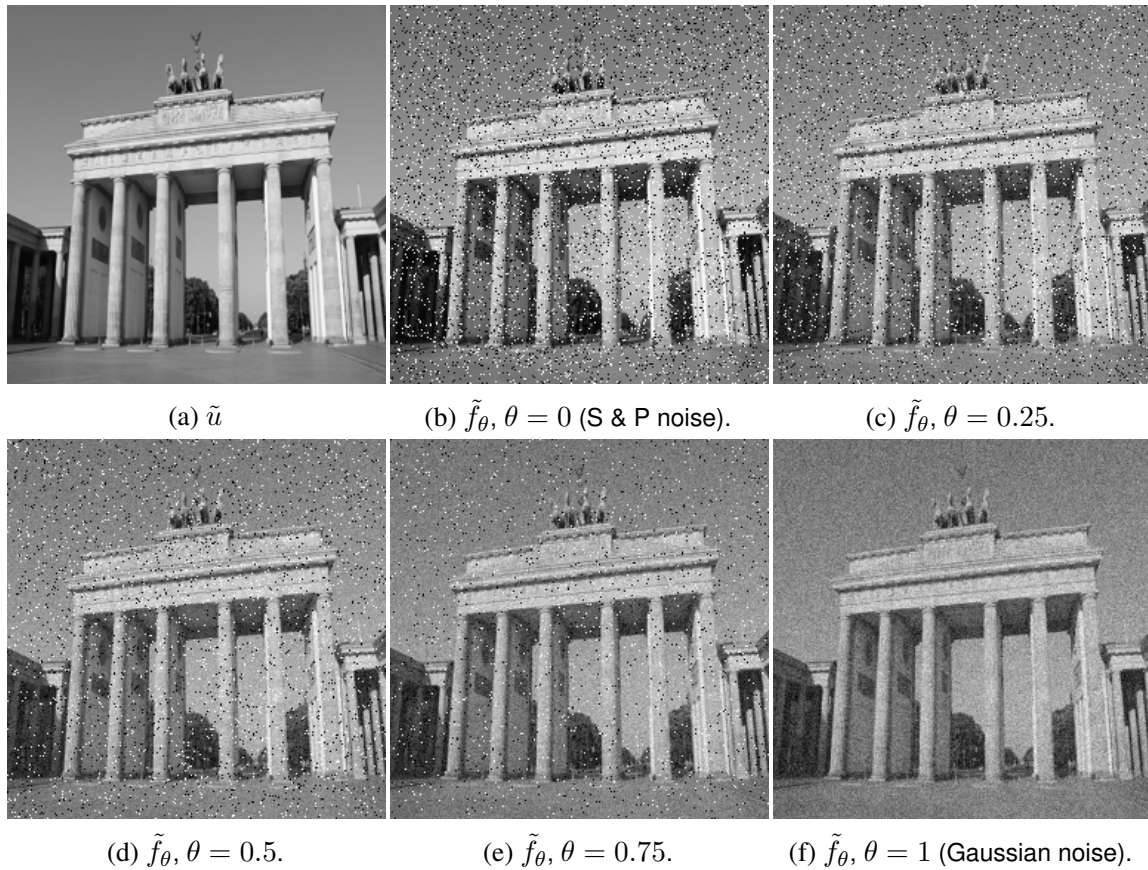


Figure 7: Training images used for the computation of the optimal parameters  $(\bar{\lambda}_1, \bar{\lambda}_2)$  whose values are reported in Figure 6; The noise-free image  $\tilde{u}$  is corrupted with a mixture of Gaussian and Salt & Pepper noise of different intensities, ranging from pure impulsive noise when  $\theta = 0$  to pure Gaussian when  $\theta = 1$ .

In the spirit of [13, 26], after a suitable regularisation of the non-smooth problem, we then proved the differentiability of the solution map via standard arguments as well as the existence of the adjoint states by which we can derive the corresponding optimality system in a compact form. Thanks to the handy characterisation of the gradient of the bilevel cost functional in terms of its adjoint states, efficient numerical schemes can be easily implemented. In particular, in Section 5 we considered the second-order BFGS Algorithm 1 to numerically compute the optimal parameters of the TV-IC model for noise mixtures with different noise levels. The numerical results show good agreement with the analytical study of the first sections, making this strategy appealing for blind image denoising applications, where the intensity of each noise component is unknown.

Further research could address the validation of the TV-IC model over a set of images in order to estimate the preferred noise model with respect to the denoising application at hand (i.e., the noise intensity, the structure of the image etc.), similarly as done in [26]. The IC discrepancy could further be combined with higher-order regularisers (such as TGV [9]) for more visually pleasing reconstructions. Finally, a similar study could be done for more general noise mixtures such as Gaussian & Poisson noise, for which the IC discrepancy has been shown in [14] to be a statistically consistent model.



## References

- [1] Allard, W.: Total Variation Regularization for Image Denoising, I. Geometric Theory. *SIAM Journal on Mathematical Analysis* **39**(4), 1150–1190 (2008). <http://dx.doi.org/10.1137/060662617>
- [2] Allard, W.: Total Variation Regularization for Image Denoising, II. Examples. *SIAM Journal on Imaging Sciences* **1**(4), 400–417 (2008). <http://dx.doi.org/10.1137/070698749>
- [3] Allard, W.: Total Variation Regularization for Image Denoising, III. Examples. *SIAM Journal on Imaging Sciences* **2**(2), 532–568 (2009). <http://dx.doi.org/10.1137/070711128>
- [4] Ambrosio, L., Fusco, N., Pallara, D.: *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford University Press, USA (2000)
- [5] Baus, F., Nikolova, M., Steidl, G.: Fully smoothed  $\ell_1$ -TV models: Bounds for the minimizers and parameter choice. *Journal of Mathematical Imaging and Vision* **48**(2), 295–307 (2014). <https://doi.org/10.1007/s10851-013-0420-0>
- [6] Bauschke, H., Combettes, P.: *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York (2011). <https://doi.org/10.1007/978-3-319-48311-5>
- [7] Benning, M., Burger, M.: Error estimates for general fidelities. *Electronic Transactions on Numerical Analysis* **38**, 44–68 (2011)
- [8] Benvenuto, F., La Camera, A., Theys, C., Ferrari, A., Lantéri, H., Bertero, M.: The study of an iterative method for the reconstruction of images corrupted by Poisson and Gaussian noise. *Inverse Problems* **24**(3), 035,016 (2008). <http://stacks.iop.org/0266-5611/24/i=3/a=035016>
- [9] Bredies, K., Kunisch, K., Pock, T.: Total Generalized Variation. *SIAM Journal on Imaging Sciences* **3**(3), 492–526 (2010). <http://dx.doi.org/10.1137/090769521>
- [10] Bredies, K., Kunisch, K., Valkonen, T.: Properties of  $L^1$ -TGV<sup>2</sup>: The one-dimensional case. *Journal of Mathematical Analysis and Applications* **398**(1), 438 – 454 (2013). <http://dx.doi.org/10.1016/j.jmaa.2012.08.053>
- [11] Burger, M., Papafitsoros, K., Papoutsellis, E., Schönlieb, C.B.: Infimal convolution regularization functionals of BV and  $L^p$  spaces. Part I: The finite  $p$  case. *Journal of Mathematical Imaging and Vision* **55**(3), 343–369 (2016). <http://dx.doi.org/10.1007/s10851-015-0624-6>
- [12] Cai, J., Chan, R., Nikolova, M.: Two-phase approach for deblurring images corrupted by impulse plus Gaussian noise. *Inverse Problems and Imaging* **2**(2), 187–204 (2008). <http://dx.doi.org/10.3934/ipi.2008.2.187>
- [13] Calatroni, L., Chung, C., De Los Reyes, J.C., Schönlieb, C.B., Valkonen, T.: Bilevel approaches for learning of variational imaging models. In: *RADON book Series on Computational and Applied Mathematics*, vol. 18. Berlin, Boston: De Gruyter (2017). <https://www.degruyter.com/view/product/458544>

- [14] Calatroni, L., De Los Reyes, J., Schönlieb, C.B.: Infimal convolution of data discrepancies for mixed noise removal. *SIAM Journal on Imaging Sciences* **10**(3), 1196–1233 (2017). <http://dx.doi.org/10.1137/16M1101684>
- [15] Calatroni, L., De Los Reyes, J.C., Schönlieb, C.B.: Dynamic sampling schemes for optimal noise learning under multiple nonsmooth constraints. In: C. Pötzsche, C. Heuberger, B. Kaltenbacher, F. Rendl (eds.) *System Modeling and Optimization, IFIP Advances in Information and Communication Technology*, vol. 443, pp. 85–95. Springer Berlin Heidelberg (2014). [http://dx.doi.org/10.1007/978-3-662-45504-3\\_8](http://dx.doi.org/10.1007/978-3-662-45504-3_8)
- [16] Caselles, V., Chambolle, A., Novaga, M.: The discontinuity set of solutions of the TV denoising problem and some extensions. *Multiscale Modeling & Simulation* **6**(3), 879–894 (2007). <http://dx.doi.org/10.1137/070683003>
- [17] Chambolle, A., Duval, V., Peyré, G., Poon, C.: Geometric properties of solutions to the total variation denoising problem. *Inverse Problems* **33**(1), 015,002 (2017). <http://stacks.iop.org/0266-5611/33/i=1/a=015002>
- [18] Chan, T., Esedoglu, S.: Aspects of Total Variation regularized  $L^1$  function approximation. *SIAM Journal on Applied Mathematics* pp. 1817–1837 (2005). <http://dx.doi.org/10.1137/040604297>
- [19] Choksi, R., Fonseca, I., Zwicknagl, B.: A few remarks on variational models for denoising. *Communications in Mathematical Sciences* **12**(5), 843–857 (2014). <http://dx.doi.org/10.4310/CMS.2014.v12.n5.a3>
- [20] Cristoferi, R.: Exact solutions for the denoising problem of piecewise constant images in dimension one. arXiv preprint arXiv:1612.05508v2 (2017). <https://arxiv.org/abs/1612.05508v2>
- [21] Dal Maso, G.: Introduction to  $\Gamma$ -convergence. Birkhäuser (1993)
- [22] Dauge, M.: Neumann and mixed problems on curvilinear polyhedra. *Integral Equations and Operator Theory* **15**(2), 227–261 (1992). <https://doi.org/10.1007/BF01204238>
- [23] De Los Reyes, J.C.: Numerical PDE-constrained optimization. SpringerBriefs in Optimization. Springer (2015). <https://dx.doi.org/10.1007/978-3-319-13395-9>
- [24] De Los Reyes, J.C., Schönlieb, C.B.: Image denoising: learning the noise model via nonsmooth PDE-constrained optimization. *Inverse Problems and Imaging* **7**(4), 1183–1214 (2013). <http://dx.doi.org/10.3934/ipi.2013.7.1183>
- [25] De Los Reyes, J.C., Schönlieb, C.B., Valkonen, T.: The structure of optimal parameters for image restoration problems. *Journal of Mathematical Analysis and Applications* **434**, 464–500 (2016). <https://doi.org/10.1016/j.jmaa.2015.09.023>
- [26] De Los Reyes, J.C., Schönlieb, C.B., Valkonen, T.: Bilevel parameter learning for higher-order Total Variation regularisation models. *Journal of Mathematical Imaging and Vision* **57**(1), 1–25 (2017). <https://doi.org/10.1007/s10851-016-0662-8>
- [27] Dong, B., Ji, H., Li, J., Shen, Z., Xu, Y.: Wavelet-based blind image inpainting. *Applied and Computational Harmonic Analysis* **32**(2), 268–279 (2012). <http://dx.doi.org/10.1016/j.acha.2011.06.001>

- [28] Duval, V., Aujol, J., Gousseau, Y.: The TV- $L^1$  model: a geometric point of view. *SIAM Journal on Multiscale Modeling & Simulation* **8**(1), 154–189 (2009). <http://dx.doi.org/10.1137/090757083>
- [29] Gröger, K.: A  $W^{1,p}$ -estimate for solutions to mixed boundary value problems for second order elliptic differential equations. *Mathematische Annalen* **283**(4), 679–688 (1989). <http://eudml.org/doc/164533>
- [30] Hintermüller, M., Stadler, G.: An Infeasible Primal-Dual Algorithm for Total Bounded Variation-Based Inf-Convolution-Type Image Restoration. *SIAM Journal on Scientific Computing* **28**(1), 1–23 (2006). <http://dx.doi.org/10.1137/040613263>
- [31] Hintermüller, M., Holler, M., Papafitsoros, K.: A function space framework for structural total variation regularization with applications in inverse problems. *Inverse Problems* **34**(6), 064,002 (2018). <http://stacks.iop.org/0266-5611/34/i=6/a=064002>
- [32] Hintermüller, M., Langer, A.: Subspace correction methods for a class of nonsmooth and non-additive convex variational problems with mixed  $L^1/L^2$  data-fidelity in image processing. *SIAM Journal on Imaging Sciences* **6**(4), 2134–2173 (2013). <http://dx.doi.org/10.1137/120894130>
- [33] Hintermüller, M., Monserrat Rincon-Camacho, M.: Expected absolute value estimators for a spatially adapted regularization parameter choice rule in  $L^1$ -TV-based image restoration. *Inverse Problems* **26**(8), 085,005 (2010). <http://stacks.iop.org/0266-5611/26/i=8/a=085005>
- [34] Hintermüller, M., Rautenberg, C.N.: Optimal selection of the regularization function in a weighted total variation model. part I: Modelling and theory. *Journal of Mathematical Imaging and Vision* **59**(3), 498–514 (2017). <https://doi.org/10.1007/s10851-017-0744-2>
- [35] Hintermüller, M., Rautenberg, C.N., Wu, T., Langer, A.: Optimal selection of the regularization function in a weighted total variation model. part II: Algorithm, its analysis and numerical tests. *Journal of Mathematical Imaging and Vision* **59**(3), 515–533 (2017). <https://doi.org/10.1007/s10851-017-0736-2>
- [36] Jalalzai, K.: Some Remarks on the Staircasing Phenomenon in Total Variation-Based Image Denoising. *Journal of Mathematical Imaging and Vision* **54**(2), 256–268 (2015). <http://dx.doi.org/10.1007/s10851-015-0600-1>
- [37] Jezierska, A., Chouzenoux, E., Pesquet, J.C., Talbot, H.: A Convex Approach for Image Restoration with Exact Poisson-Gaussian Likelihood. *SIAM Journal on Imaging Science* **62**(1), 17–30 (2015). <http://dx.doi.org/10.1137/15M1014395>
- [38] Klatzer, T., Pock, T.: Continuous hyper-parameter learning for support vector machines. In: 20<sup>th</sup> Computer Vision Winter Workshop Paul Wohlhart, Vincent Lepetit (eds.) Seggau, Austria, February 9-11, 2015 (2015). <http://dx.doi.org/10.3217/978-3-85125-388-7>
- [39] Kunisch, K., Pock, T.: A bilevel optimization approach for parameter learning in variational models. *SIAM Journal on Imaging Sciences* **6**(2), 938–983 (2013). <http://dx.doi.org/10.1137/120882706>

- [40] Langer, A.: Automated parameter selection for total variation minimization in image restoration. *Journal of Mathematical Imaging and Vision* **57**(2), 239–268 (2017). <http://dx.doi.org/10.1007/s10851-016-0676-2>
- [41] Langer, A.: Automated parameter selection in the  $L^1$ - $L^2$ -TV model for removing gaussian plus impulse noise. *Inverse Problems* **33**(7), 074,002 (2017). <http://stacks.iop.org/0266-5611/33/i=7/a=074002>
- [42] Lanza, A., Morigi, S., Sgallari, F., Wen, Y.W.: Image restoration with Poisson-Gaussian mixed noise. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* **2**, 12–24 (2014). <http://dx.doi.org/10.1080/21681163.2013.811039>
- [43] Meyer, Y.: Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures, vol. 22. American Mathematical Society (2001)
- [44] Nikolova, M.: A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision* **20**(1), 99–120 (2004). <http://dx.doi.org/10.1023/B:JMIV.0000011326.88682.e5>
- [45] Nikolova, M., Wen, Y., Chan, R.: Exact histogram specification for digital images using a variational approach. *Journal of Mathematical Imaging and Vision* **46**(3), 309–325 (2013). <http://dx.doi.org/10.1007/s10851-012-0401-8>
- [46] Ochs, P., Ranftl, R., Brox, T., Pock, T.: Bilevel optimization with nonsmooth lower level problems. In: J.F. Aujol, M. Nikolova, N. Papadakis (eds.) *Scale Space and Variational Methods in Computer Vision*, pp. 654–665. Springer International Publishing (2015). [https://doi.org/10.1007/978-3-319-18461-6\\_52](https://doi.org/10.1007/978-3-319-18461-6_52)
- [47] Papafitsoros, K.: Novel higher order regularisation methods for image reconstruction. Ph.D. thesis, University of Cambridge (2014). <https://www.repository.cam.ac.uk/handle/1810/246692>
- [48] Papafitsoros, K., Bredies, K.: A study of the one dimensional total generalised variation regularisation problem. *Inverse Problems and Imaging* **9**(2), 511–550 (2015). <http://dx.doi.org/10.3934/ipi.2015.9.511>
- [49] Papafitsoros, K., Valkonen, T.: Asymptotic behaviour of total generalised variation. In: J.F. Aujol, M. Nikolova, N. Papadakis (eds.) *Scale Space and Variational Methods in Computer Vision: 5th International Conference, SSVM 2015, Proceedings*, pp. 702–714. Springer International Publishing (2015). [http://dx.doi.org/10.1007/978-3-319-18461-6\\_56](http://dx.doi.org/10.1007/978-3-319-18461-6_56)
- [50] Pöschl, C., Scherzer, O.: Exact solutions of one-dimensional total generalized variation. *Communications in Mathematical Sciences* **13**(1), 171–202 (2015). <http://dx.doi.org/10.4310/CMS.2015.v13.n1.a9>
- [51] Ring, W.: Structural properties of solutions to total variation regularization problems. *ESAIM: Mathematical Modelling and Numerical Analysis* **34**(4), 799–810 (2000). <http://dx.doi.org/10.1051/m2an:2000104>
- [52] Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* **60**(1-4), 259–268 (1992). [http://dx.doi.org/10.1016/0167-2789\(92\)90242-F](http://dx.doi.org/10.1016/0167-2789(92)90242-F)

- [53] Valkonen, T.: The jump set under geometric regularization. Part 1: Basic technique and first-order denoising. *SIAM Journal on Mathematical Analysis* **47**(4), 2587–2629 (2015). <https://doi.org/10.1137/140976248>
- [54] Valkonen, T.: The jump set under geometric regularisation. Part 2: Higher-order approaches. *Journal of Mathematical Analysis and Applications* **453**(2), 1044–1085 (2017). <https://doi.org/10.1016/j.jmaa.2017.04.037>
- [55] Yan, M.: Restoration of images corrupted by impulse noise and mixed gaussian impulse noise using blind inpainting. *SIAM Journal on Imaging Sciences* **6**(3), 1227–1245 (2013). <http://dx.doi.org/10.1137/12087178X>
- [56] Yu, W., Heber, S., Pock, T.: Learning Reaction-Diffusion Models for Image Inpainting, vol. 9358, pp. 356–367. Springer International Publishing AG, Switzerland (2015). [https://doi.org/10.1007/978-3-319-24947-6\\_29](https://doi.org/10.1007/978-3-319-24947-6_29)
- [57] Zowe, J., Kurcyusz, S.: Regularity and stability for the mathematical programming problem in Banach spaces. *Applied Mathematics and Optimization* **5**(1), 49–62 (1979). <https://doi.org/10.1007/BF01442543>