A function space framework for structural total variation regularization with applications in inverse problems

MICHAEL HINTERMÜLLER[†‡], MARTIN HOLLER [§] AND KOSTAS PAPAFITSOROS [†]

ABSTRACT. In this work, we introduce a function space setting for a wide class of structural/weighted total variation (TV) regularization methods motivated by their applications in inverse problems. In particular, we consider a regularizer that is the appropriate lower semi-continuous envelope (relaxation) of a suitable total variation type functional initially defined for sufficiently smooth functions. We study examples where this relaxation can be expressed explicitly, and we also provide refinements for weighted total variation for a wide range of weights. Since an integral characterization of the relaxation in function space is, in general, not always available, we show that, for a rather general linear inverse problems setting, instead of the classical Tikhonov regularization problem, one can equivalently solve a saddle-point problem where no a priori knowledge of an explicit formulation of the structural TV functional is needed. In particular, motivated by concrete applications, we deduce corresponding results for linear inverse problems with norm and Poisson log-likelihood data discrepancy terms. Finally, we provide proof-of-concept numerical examples where we solve the saddle-point problem for weighted TV denoising as well as for MR guided PET image reconstruction.

1. INTRODUCTION

In many classical applications of inverse problems, rather than measuring data from a single channel, recently the simultaneous acquisition from multiple channels has gained importance. Besides having different sources of information available, a main advantage of such multiple measurements is due to the possibility of exploiting correlations between different data channels in the inversion process, often leading to a significant improvement for each individual channel. In particular, when the underlying quantities of interest can be visualized as image data, these correlations typically correspond to joint structures in images. In view of this background, multimodality and multicontrast imaging explore such joint structures for improved reconstruction. Successful applications of these techniques can be found for instance in biomedical imaging [9, 20, 21, 22, 40, 43, 45, 49, 52], geosciences [51], electron microscopy [27] and several more.

In this context, one distinguishes two different approaches for exploiting correlations: (i) Joint reconstruction techniques that treat all available channels equally, such as [40], and (ii) structural-prior-based regularization techniques that assume some ground truth structural information to be available. Here we focus on a particular class of type (ii), namely *structural total-variation-type regularization functionals*, i.e., functionals which integrate a spatially-dependent pointwise function of the image gradient for regularization. In this vein, given some prior information v, we consider a regularization functional of the type

$$J(u) = \int_\Omega j_v(x, \nabla u(x)) \, dx,$$

[†]Weierstrass Institute for Applied Analysis and Stochastics (WIAS), Mohrenstrasse 39, 10117, Berlin, Germany

[‡]Institute for Mathematics, Humboldt University of Berlin, Unter den Linden 6, 10099, Berlin, Germany

[§]Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstrasse 36, A-8010, Graz, Austria. The institute is a member of NAWI Graz (`www.nawigraz.at`). M. H. is a member of BioTechMed Graz (`www.biotechmed.at`).

 Emails: `Michael.Hintermueller@wias-berlin.de`
        `Martin.Holler@uni-graz.at`
        `Kostas.Papafitsoros@wias-berlin.de`

which is used in a Tikhonov-type regularization problem where u is reconstructed by solving

$$(1.1) \qquad \min_u \int_\Omega j_v(x, \nabla u(x)) \, dx + (\lambda D \circ K)(u).$$

Here, D denotes a data discrepancy term and K a bounded linear operator which reflects the forward model. A very basic, formal example of such a regularization is given when v denotes the underlying ground truth image and we incorporate information on its gradient by defining $j_v(x, z) := \frac{1}{|\nabla v(x)|}|z|$, where $|\cdot|$ is some norm. This yields

$$J(u) = \int_\Omega \frac{1}{|\nabla v(x)|} |\nabla u(x)| \, dx,$$

which corresponds to a weighted total variation (TV) functional.

Even though problems of the form (1.1) limit the exchange of information to the gradient level, they cover a large set of existing works in particular in the context of medical image processing. This can be explained on the one hand by the popularity of TV-type regularization for imaging in general, which is due to its capability of recovering jump discontinuities. On the other hand, the gradient level seems very well suited for exploiting correlations between different contrasts or modalities as it primarily encodes structural information, i.e., to some extent it is independent of the absolute magnitude of the signal. With the goal of enforcing parallel level sets, in [19], for instance, the regularizer J was chosen for a given image v as

$$(1.2) \qquad J(u) = \int_\Omega \phi\left(\psi(|\nabla v|_\beta |\nabla v|_\beta) - \psi(|(\nabla u \cdot \nabla v)|_{\beta^2})\right) \, dx,$$

where ϕ, ψ are appropriate increasing functions, $|\cdot|_\beta$, $|\cdot|_{\beta^2}$ denote smoothed norms, and $a \cdot b$ is the scalar ("dot") product of vectors $a, b \in \mathbb{R}^d$. This definition is motivated by the fact that for two vectors w, y the quantity $|w||y| - |(w, y)|$ is zero if and only if they are parallel to each other. Particular instances of (1.2) were then also used in [22] for joint MR-PET reconstruction. A similar functional, but in the spirit of [39] is

$$(1.3) \qquad J(u) = \int_\Omega |\nabla u|^2 - (\nabla u \cdot w)^2 \, dx.$$

Here w denotes some a priori vector field that contains gradient information. In the context of multicontrast MR, the authors in [20] choose

$$(1.4) \qquad J(u) = \int_\Omega \left|\left(I - \frac{\nabla v \otimes \nabla v}{|\nabla v|^2}\right) \nabla u\right| \, dx = \int_\Omega |\nabla u| \sin \theta \, dx.$$

where again a smoothed version of the absolute value was used in the denominator. Here, θ denotes the angle between ∇u and ∇v. Observe that the latter functional can also be written as

$$\int_\Omega \left(|\nabla u|^2 - \left(\nabla u \cdot \frac{\nabla v}{|\nabla v|}\right)^2\right)^{1/2},$$

bearing similarity to the functional (1.3).

We note that the regularization approaches above have been considered only in discrete settings, despite the original continuous formulations. Concerning the latter and as indicated above in connection with (1.1), it is natural to consider $u \in \mathrm{BV}(\Omega)$, i.e., $|u|$ is Lebesgue integrable and u is of bounded total variation, still allowing for jump discontinuities, i.e., sharp edges. As a consequence, the (generalized) gradient Du of u is in general a finite Radon measure, only. This fact, however, challenges the proper defininition of regularization functionals of the above types in an associated function space setting. The first part of our work aims at addressing precisely this issue. In fact,

resorting to the concept of functions of a measure [7] we propose to use the convex biconjugate as a regularizer. Indeed, starting from a general function $j : \Omega \times \mathbb{R}^d \to [0, \infty)$, $\Omega \subset \mathbb{R}^d$ with minimal assumptions, e.g., convexity, linear growth, 1-homogeneity in the second variable, we define the functional $J : L^p(\Omega) \to \mathbb{R}$ as

$$(1.5) \qquad J(u) = \begin{cases} \int_\Omega j(x, \nabla u(x)) \, dx & \text{if } u \in W^{1,1}(\Omega), \\ \infty & \text{else,} \end{cases}$$

where we omit the dependence of j on the prior v for the sake of unburdening notation. Here, $L^p(\Omega)$ and $W^{1,1}(\Omega)$ denote the usual Lebesgue and Sobolev spaces; see [1]. We note that J in (1.5) is not suitable for variational regularization as it is finite only for a class of rather regular functions (i.e., in $W^{1,1}(\Omega)$) and it is not lower semi-continuous in an appropriate topology. As a remedy, we propose to resort to the convex biconjugate J^{**} of J, which we call *structural TV functional*. We recall that J^{**} coincides with the lower semi-continuous envelope (relaxation) of J with respect to L^p convergence. In general, J^{**} may not have an explicit representation, but we show that it can always be expressed in a dual form as

$$(1.6) \qquad J^{**}(u) = \sup_{g \in Q} \int_\Omega u \operatorname{div} g \, dx, \quad \text{for } u \in L^p(\Omega),$$

where

$$(1.7) \qquad Q := \left\{ g \in W_0^q(\operatorname{div}; \Omega) \cap L^\infty(\Omega, \mathbb{R}^d) : \ j^\circ(x, g(x)) \le 1 \text{ for almost every (a.e.) } x \in \Omega \right\},$$

see (3.2) below for the precise definition of the support function j°, and we refer to [25] for details on $W_0^q(\operatorname{div}; \Omega)$. Nevertheless, based on a recent result by Amar, De Cicco and Fusco [2], by linking J^{**} to the relaxed functional of J with respect to L^1 convergence we are able to provide an integral representation, under additional assumptions. For instance, in the case where $j(x, z) = \alpha(x)|z|$, $\alpha \in \mathrm{BV}(\Omega)$, with $\alpha \ge 0$, J^{**} is equal to the weighted TV-functional

$$J^{**}(u) = \int_\Omega \alpha^- \, d|Du|, \quad u \in \mathrm{BV}(\Omega),$$

with α^- the approximate lower limit of α; see (2.3) below for its definition. The set Q then reads

$$Q = \left\{ g \in W_0^q(\operatorname{div}; \Omega) \cap L^\infty(\Omega, \mathbb{R}^d) : \ |g(x)| \le \alpha(x) \text{ for a.e. } x \in \Omega \right\}.$$

Interestingly, as a consequence of this formulation, we get certain density results of convex intersections in the spirit of [31, 32, 34] as a byproduct; compare Section 4.

Taking advantage of duality theory, in the second part of the paper we use the structural TV functional J^{**} for the regularization of linear inverse problems. In particular we study the general minimization problem

$$(1.8) \qquad \inf_{u \in L^p(\Omega)} J^{**}(u) + (\lambda D \circ K)(u).$$

As emphasized above, an explicit representation of the functional J^{**} is available only under some additional, perhaps restrictive assumptions. In order to solve (1.8) without invoking such assumptions, we employ Fenchel-Rockafellar duality and show, in the continuous setting, equivalence of (1.8) to a saddle-point problem of the form

$$(1.9) \qquad \inf_{\substack{p \in W_0^q(\operatorname{div}; \Omega) \\ p \in Q}} \sup_{u \in L^p(\Omega)} (\operatorname{div} p, u) - (\lambda D \circ K)(u),$$

3

where $(\cdot,\cdot)$ denotes an appropriate pairing. This is achieved under assumptions which are tight in the sense that we can provide a counterexample where, without these assumptions, even existence of a solution for (1.8) fails. The major advantage of the above saddle-point reformulation is that it allows to obtain a solution of the original problem without requiring an explicit form of neither J^{**} nor $(D \circ K)^*$. Note that the latter would be required for solving the predual problem. Furthermore, it has a format directly amenable to duality-based numerical optimization algorithms.

The equivalence to a saddle-point reformulation is obtained under rather general assumptions on the data discrepancy term D, which, as corollary, allows us to cover the case of any norm discrepancy term as well as the case of a log-likelihood term for Poisson noise, which is relevant for instance for PET image reconstruction and electron tomography. The latter leads to the minimization problem

$$(1.10) \qquad \min_{u \in L^p(\Omega)} J^{**}(u) + \lambda \int_\Sigma Ku - f \log(Ku + c_0)\, d\sigma + \mathcal{I}_{[0,\infty)}(u).$$

where K denotes a Radon-transform-type operator, f, c_0 some given data, and $\mathcal{I}$ is the indicator function for a set M, i.e., $\mathcal{I}_M(u) = 0$ if $u \in M$ and $\mathcal{I}_M(u) = +\infty$ otherwise.

Finally, we show the versatility of our approach with proof-of-concept numerical examples in weighted TV denoising with vanishing weight function and MR guided PET reconstruction.

We note here that there is previous work on the analysis of weighted and/or structural TV regularization in an infinite dimension setting. In [26], another instance of structural-TV type functionals is employed, but the work only considers the case of image denoising. Further, the authors simultaneously optimize over the image data and an anisotropy in the TV-term, which leads to a non-convex problem. Regarding properties of solutions of weighted TV denoising we refer to the work by Jalalzai [38] as well as to [30]. Finally, we mention that in [5] the authors analyze a weighted TV regularization model for vortex density models.

Structure of the paper. The paper is organized as follows: In Section 2, we fix our notation and we remind the reader of basic facts concerning functions of bounded variation and $W(\mathrm{div})$ spaces.

In Section 3 we describe the relaxation framework for the structural TV functional. Under some conditions, we provide an integral representation of the relaxation based on a result of [2] and we also provide a characterization of its subdifferential.

Some refinements for weighted TV functional are given in Section 4 in the case of continuous and lower semi-continuous weight functions. We show that the functional can be defined in a dual fashion, using smooth test functions and as a byproduct we obtain certain density results of convex intersections in the spirit of [32, 34].

Section 5 contains the main duality result of the paper. In particular, we show that under certain mild assumptions, the variational regularization problem with the relaxed structural TV functional as regularizer can be equivalently formulated as a saddle-point problem which requires knowledge neither of the explicit form of the relaxation–as it is the case for the primal problem –nor of the convex conjugate of the discrepancy term–as it is the case for the predual problem. As particular application, we elaborate on the case of Poisson log-likelihood discrepancy terms and show how the result can be transferred to this situation.

Finally, in Section 6 we provide proof-of-concept numerical examples, where we solve our saddle point problem for weighted TV denoising with vanishing weight function and for MR-guided PET image reconstruction.

2. Notation and Preliminaries

2.1. **Functions of bounded variation.** Throughout the paper, $\Omega \subset \mathbb{R}^d$, with $d \geq 2$, will be a bounded, open set with Lipschitz boundary, and we denote $d^* = \frac{d}{d-1}$. By p, q we always denote two real numbers such that $p, q \in [1, \infty]$ and $q = \frac{p}{p-1}$ if $p \in (1, \infty)$, $q = \infty$ if $p = 1$, and $q = 1$ if $p = \infty$.

The space of *functions of bounded variation* on Ω is denoted by $\mathrm{BV}(\Omega)$. We have that $u \in \mathrm{BV}(\Omega)$ if and only if it is in $L^1(\Omega)$ and its distributional derivative is a bounded Radon measure, denoted by Du. The total variation $\mathrm{TV}(u)$ of u is defined to be the total variation of that measure, i.e., $\mathrm{TV}(u) = |Du|(\Omega)$ and it is equal to

$$(2.1) \qquad |Du|(\Omega) = \sup\left\{ \int_\Omega u \,\mathrm{div}\phi\, dx : \phi \in C_c^\infty(\Omega, \mathbb{R}^d), \ |\phi(x)| \leq 1, \ \forall x \in \Omega \right\}.$$

The measure Du can be decomposed into

$$Du = D^a u + D^j u + D^c u,$$

where $D^a u$ is the absolutely continuous with respect to Lebesgue measure $\mathcal{L}^d$, with density function denoted by ∇u, $D^j u$ denotes the jump part which is the restriction to the jump set J_u of u, and $D^c u$ is the Cantor part of Du. We recall that D_u^j is defined over the set of points in Ω for which that $u^+(x) > u^-(x)$ where

$$(2.2) \qquad u^+(x) = \inf\left\{ t \in [-\infty, \infty] : \lim_{r \to 0} \frac{\mathcal{L}^d(\{v > t\} \cap B(x, r))}{r^d} = 0 \right\},$$

$$(2.3) \qquad u^-(x) = \sup\left\{ t \in [-\infty, \infty] : \lim_{r \to 0} \frac{\mathcal{L}^d(\{v < t\} \cap B(x, r))}{r^d} = 0 \right\},$$

are the approximate upper and lower limits of u, respectively. With these definitions, the total variation of the measure $D^j u$ can be written as

$$|D^j u|(\Omega) = \int_{J_u} |u^+(x) - u^-(x)| \, d\mathcal{H}^{d-1},$$

where $\mathcal{H}^{d-1}$ denotes the $(d-1)$-dimensional Hausdorff measure. The density functions of the measures $D^j u$ and $D^c u$ with respect to $|D^j u|$ and $|D^c u|$ are denoted by $\sigma_{D^j u}$ and $\sigma_{D^c u}$, respectively. For further details about the space $\mathrm{BV}(\Omega)$, we refer the reader to [3, 24].

2.2. **The space $W^q(\mathrm{div}; \Omega)$.** As the Banach space $W^q(\mathrm{div}; \Omega)$, with $q \in [1, \infty)$, plays a major role in our work, we recall here some basic facts.

Definition 2.1 ($W^q(\mathrm{div}; \Omega)$). *Let $1 \leq q < \infty$ and $g \in L^q(\Omega, \mathbb{R}^d)$. We have $\mathrm{div}g \in L^q(\Omega)$ if there exists $w \in L^q(\Omega)$ such that for all $\phi \in C_c^\infty(\Omega)$*

$$\int_\Omega \nabla\phi \cdot g \, dx = -\int_\Omega \phi w \, dx.$$

Furthermore we define

$$W^q(\mathrm{div}; \Omega) := \left\{ g \in L^q(\Omega, \mathbb{R}^d) : \mathrm{div}g \in L^q(\Omega) \right\},$$

with the norm $\|g\|_{W^q(\mathrm{div};\Omega)}^q := \|g\|_{L^q(\Omega)}^q + \|\mathrm{div}g\|_{L^q(\Omega)}^q$.

Remark 2.2. *By density of $C_c^\infty(\Omega)$ in $L^p(\Omega)$ we have $\mathrm{div}g = w$ as $w \in L^q(\Omega)$ is unique. By completeness of $L^q(\Omega)$ and $L^q(\Omega, \mathbb{R}^d)$ it follows that $W^q(\mathrm{div}; \Omega)$ is a Banach space when equipped with $\|\cdot\|_{W^q(\mathrm{div};\Omega)}$.*

We now state some general properties of $W^q(\mathrm{div};\Omega)$. As $W^q(\mathrm{div};\Omega)$ is just a straightforward generalization of the well-known space $H(\mathrm{div};\Omega) := W^2(\mathrm{div};\Omega)$, these results can be proven readily by generalizing from $H(\mathrm{div};\Omega)$; see [25, Chapter 1] for details on the latter.

Proposition 2.3 (Density)**.** *Let $q \in [1, \infty)$. Then $C^\infty(\overline{\Omega}, \mathbb{R}^d)$ is dense in $W^q(\mathrm{div};\Omega)$ with respect to $\|\cdot\|_{W^q(\mathrm{div};\Omega)}$.*

Proposition 2.4 (Normal trace)**.** *Let $q \in [1, \infty)$ and denote by $n_\Omega(x) \in \mathbb{R}^d$ the outer normal vector to $\partial\Omega$ at $x \in \partial\Omega$. Then the mapping*

$$\tau : C^\infty(\overline{\Omega}, \mathbb{R}^d) \to \left(W^{1-\frac{1}{p},p}(\partial\Omega)\right)^*, \qquad g \mapsto \tau(g),$$

with $\tau(g)(v) := \int_{\partial\Omega}(g, n_\Omega)v\, d\mathcal{H}^{d-1}$ for $v \in W^{1-\frac{1}{p},p}(\partial\Omega)$, can be extended to a linear, continuous mapping, also denoted by $\tau : W^q(\mathrm{div};\Omega) \to \left(W^{1-\frac{1}{p},p}(\partial\Omega)\right)^*$. Further we have a Gauss-Green formula for $W^q(\mathrm{div};\Omega)$ functions:

$$\int_\Omega \nabla v \cdot g\, dx + \int_\Omega v \,\mathrm{div} g\, dx = \tau(g)(v) \quad \textit{for all } v \in W^{1,p}(\Omega),\ g \in W^q(\mathrm{div};\Omega).$$

Definition 2.5. *For $1 \le q < \infty$, we define*

$$W_0^q(\mathrm{div};\Omega) = \overline{C_c^\infty(\Omega, \mathbb{R}^d)}^{\|\cdot\|_{W^q(\mathrm{div};\Omega)}}.$$

The next proposition also uses ideas from [25] but since its proof is slightly more involved, we include it in Appendix A.

Proposition 2.6. *For $q \in [1, \infty)$ we have $W_0^q(\mathrm{div};\Omega) = \{g : \tau(g) \equiv 0\} =: \ker(\tau)$, with τ the $W^q(\mathrm{div};\Omega)$-trace operator as in Proposition 2.4 and $\tau(g) \equiv 0$ is understood in the sense of $\left(W^{1-\frac{1}{p},p}(\partial\Omega)\right)^*$.*

From this fact, another equivalent characterization of $W_0^q(\mathrm{div};\Omega)$ can be obtained readily.

Corollary 2.7. *For $q \in [1, \infty)$ and $g \in W^q(\mathrm{div};\Omega)$ we have $g \in W_0^q(\mathrm{div};\Omega)$ if and only if*

$$(2.4) \qquad \int_\Omega \nabla v \cdot g\, dx = -\int_\Omega v \,\mathrm{div} g\, dx, \quad \textit{for all } v \in W^{1,p}(\Omega).$$

3. Structural TV as a lower semi-continuous envelope

The goal of this section is to obtain a predual representation of a general TV-based functional that includes the case of weighted TV for a general choice of weights. To this aim, we will define the corresponding functional as the L^p-lower semi-continuous envelope of a restriction to $W^{1,1}$ functions, with $p \in [1, \infty)$. The approach is motivated by the paper of Bouchitte and Dal Maso [12]. We start with a few definitions.

By $j : \Omega \times \mathbb{R}^d \to [0, \infty)$ we always denote a function satisfying the following conditions:

(J1) For a.e. $x \in \Omega$, $j(x, \cdot)$ is convex and positively 1-homogeneous on $\mathbb{R}^d$.
(J2) There exists $\gamma > 0$ such that

$$0 \le j(x, z) \le \gamma(1 + |z|) \quad \text{for a.e. } x \in \Omega \text{ and every } z \in \mathbb{R}^d.$$

(J3) For every $z \in \mathbb{R}^d$, $j(x, z) = j(x, -z)$, i.e., j is an even function in the second variable.

Furthermore, for $p \in [1, \infty)$ we define $J : L^p(\Omega) \to [0, \infty]$ as

$$(3.1) \qquad J(u) := \begin{cases} \int_\Omega j(x, \nabla u(x))\, dx & \text{if } u \in W^{1,1}(\Omega), \\ \infty & \text{else.} \end{cases}$$

6

Remark 3.1. *We note that $j(x, z) = \alpha(x)|z|$ with any $\alpha : \Omega \to [0, \infty)$ bounded above satisfies the above assumptions (J1)–(J3).*

In what follows, convex conjugation of j will always be understood with respect to the second argument, and convex conjugation of J is performed in $L^p(\Omega)$. Due to positive 1-homogeneity, we get the following well known representation of j^*. For its formulation, we define the support function

$$(3.2) \qquad j^\circ(x, z^*) := \sup_{z : j(x,z) \le 1} z^* \cdot z,$$

and denote by j^* the convex conjugate of j; see [23] for more information on the latter concept.

Proposition 3.2. *Let $z^* \in \mathbb{R}^d$. Then, for any $x \in \Omega$ we have $j^*(x, z^*) = \mathcal{I}_{A_x}(z^*)$, with $\mathcal{I}_{A_x}$ the convex indicator function of the set $A_x = \{z \in \mathbb{R}^d : j^\circ(x, z) \le 1\}$.*

Proof. In case $j(x, \cdot) = 0$ the assertion holds true trivially. Hence we assume that there exists $z \in \mathbb{R}^d$ with $j(x, z) \neq 0$. For such a point z, we can write

$$j(x, z) = j\left(x, j(x, z)\frac{z}{j(x, z)}\right) = \lambda j(x, \tilde{z})$$

with $\lambda = j(x, z) \ge 0$ and for $\tilde{z}$ such that $j(x, \tilde{z}) = 1$. Hence we get

$$j^\circ(x, z^*) = \sup_{\lambda \in [0,1]} \sup_{j(x,z)=1} \lambda(z^* \cdot z) = \sup_{j(x,z)=1} z^* \cdot z \ge 0,$$

where the non-negativity follows from the fact that $j(x, \cdot)$ is even. We further have

$$j^*(x, z^*) = \sup_z \{z^* \cdot z - j(x, z)\} = \sup_{\lambda \ge 0} \sup_{j(x,z)=1} \{z^* \cdot (\lambda z) - \lambda j(x, z)\}$$

$$= \sup_{\lambda \ge 0} \sup_{j(x,z)=1} \lambda(z^* \cdot z - 1) = \sup_{\lambda \ge 0} \lambda(j^\circ(x, z^*) - 1)$$

$$= \begin{cases} \infty & \text{if } j^\circ(x, z^*) > 1, \\ 0 & \text{if } j^\circ(x, z^*) \le 1 \end{cases} = \mathcal{I}_{A_x}(z^*),$$

which completes the proof.

$\square$

Remark 3.3. *From the proposition above if follows for $j(x, z) = \alpha(x)|z|$ that*

$$j^\circ(x, z^*) = \begin{cases} \frac{|z^*|}{\alpha(x)} & \text{if } \alpha(x) \neq 0, \\ \mathcal{I}_{\{0\}}(z^*) & \text{else,} \end{cases}$$

and hence $A_x = \{z^* \in \mathbb{R}^d : |z^*| \le \alpha(x)\}$.

We note that by definition and density we have for $u^* \in L^q(\Omega)$ with $q = \frac{p}{p-1}$ that

$$J^*(u^*) = \sup_{u \in L^p(\Omega)} \{(u^*, u) - J(u)\} = \sup_{u \in BV(\Omega)} \{(u^*, u) - J(u)\} = \sup_{u \in W^{1,1}(\Omega)} \{(u^*, u) - J(u)\}.$$

Here we write $(v, w) := \int_\Omega vw \, dx$ for $v \in L^q(\Omega)$ and $w \in L^p(\Omega)$. The next proposition follows the lines of [12] and provides a characterization J^*, which also holds without the positive 1-homogeneity assumption on j.

7

Proposition 3.4. *Let $p \in (1, \infty)$. Then we have for all $u^* \in L^q(\Omega)$ that*

$$J^*(u^*) = \min_{g \in K(u^*)} \int_\Omega j^*(x, g(x)) \, dx,$$

where $K(u^*) := \{g \in W_0^q(\mathrm{div}; \Omega) \cap L^\infty(\Omega, \mathbb{R}^d) : -\mathrm{div} g = u^*\}$, and we set $\min \emptyset := +\infty$.

Proof. We define the map $F : L^1(\Omega, \mathbb{R}^d) \to \overline{\mathbb{R}}$ with $F(p) := \int_\Omega j(x, p(x)) \, dx$. Due to (J2), F is well defined. Then, by [23, Theorem X.2.1], we get for $F^* : L^\infty(\Omega, \mathbb{R}^d) \to \overline{\mathbb{R}}$ that

$$F^*(p^*) = \int_\Omega j^*(x, p^*(x)) \, dx.$$

Further, we define the unbounded operator $\Lambda : L^p(\Omega) \to L^1(\Omega, \mathbb{R}^d)$ by $\mathrm{dom}(\Lambda) = W^{1,1}(\Omega) \cap L^p(\Omega)$ and $\Lambda u := \nabla u$. Then Λ is densely defined and closed. We note that J can be re-written as

$$J(u) = \begin{cases} F(\Lambda u) & \text{if } u \in \mathrm{dom}(\Lambda), \\ +\infty & \text{else.} \end{cases}$$

Due to $|F(p)| \leq \gamma \, \mathrm{meas}(\Omega) + \gamma \|p\|_{L^1(\Omega)}$ by (J2), where $\mathrm{meas}(\cdot)$ denotes the Lebesgue measure of a set, J is bounded in a neighborhood of any Λu with $u \in \mathrm{dom}(\Lambda)$. Hence, it follows from [46, Theorem 19] that

$$J^*(u^*) = \min \{F^*(y^*) : y^* \in \mathrm{dom}(\Lambda^*), \ \Lambda^* y^* = u^*\}$$

$$= \min \left\{ \int_\Omega j^*(x, y^*(x)) \, dx : y^* \in \mathrm{dom}(\Lambda^*), \ \Lambda^* y^* = u^* \right\}.$$

To complete the proof, it is left to show that we have for any $u^* \in L^q(\Omega)$ that

$$\{y^* \in \mathrm{dom}(\Lambda^*) : \Lambda^* y^* = u^*\} = \{g \in W_0^q(\mathrm{div}; \Omega) \cap L^\infty(\Omega, \mathbb{R}^d) : -\mathrm{div} g = u^*\}.$$

We first show the inclusion "$\supset$": In case the set on the right-hand side above is empty, then the inclusion holds trivially; otherwise take $g \in W_0^q(\mathrm{div}; \Omega) \cap L^\infty(\Omega, \mathbb{R}^d)$ with $-\mathrm{div} g = u^*$ and $v \in \mathrm{dom}(\Lambda) \subset W^{1,1}(\Omega)$. By density [24, Theorem 4.3] there exists a sequence $(v_n)_{n \in \mathbb{N}}$ in $W^{1,1}(\Omega) \cap C^\infty(\overline{\Omega})$ converging to v in $W^{1,1}(\Omega)$ for which we can also assume that $v_n \to v$ in $L^p(\Omega)$ since $v \in L^p(\Omega)$. Then, by the Gauss-Green theorem for $W_0^q(\mathrm{div}; \Omega)$, see (2.4), we get

$$\int_\Omega \mathrm{div} g \, v \, dx \leftarrow \int_\Omega \mathrm{div} g \, v_n \, dx = -\int_\Omega g \cdot \Lambda v_n \, dx \to -\int_\Omega g \cdot \Lambda v \, dx.$$

Hence $|(\Lambda v, g)| \leq C_g \|v\|_{L^p(\Omega)}$ with some $C_g > 0$, and thus $g \in \mathrm{dom}(\Lambda^*)$ and $-\mathrm{div} g = \Lambda^* g$. To show the reverse inclusion "$\subset$", again assuming the set on the left-hand side to be non-empty, we take $y^* \in \mathrm{dom}(\Lambda^*) \subset L^\infty(\Omega, \mathbb{R}^d)$, for which we want to show that $y^* \in W_0^q(\mathrm{div}; \Omega)$ and $\Lambda^* y^* = -\mathrm{div} y^*$. By definition of $\mathrm{dom}(\Lambda^*)$ the mapping $v \mapsto \int_\Omega \Lambda^* y^* v \, dx$ is a continuous linear functional on $L^p(\Omega)$, hence there exists $w \in L^q(\Omega)$ such that

$$\int_\Omega w v \, dx = \int_\Omega \Lambda^* y^* v \, dx = \int_\Omega y^* \cdot \nabla v \, dx$$

for all $v \in \mathrm{dom}(\Lambda) \supset C^\infty(\overline{\Omega}) \supset C_c^\infty(\Omega)$. Hence, $y^* \in W^q(\mathrm{div}; \Omega)$. Further, we also get that $y \in \ker(\tau)$, with τ being the normal trace operator of Proposition 2.4, and hence $y^* \in W_0^q(\mathrm{div}; \Omega)$, which completes the proof. $\qquad\square$

Using positive 1-homogeneity, Proposition 3.2 immediately implies the following refinement.

Corollary 3.5. *Under the assumption of Proposition 3.4 we get that*

$$J^*(u^*) = \mathcal{I}_{\mathrm{div}(Q)}(u^*)$$

with

$$Q = \left\{ g \in W_0^q(\mathrm{div}; \Omega) \cap L^\infty(\Omega, \mathbb{R}^d) : j^\circ(x, g(x)) \leq 1 \text{ for a.e. } x \in \Omega \right\}.$$

In particular, if $j(x, z) = \alpha(x)|z|$ then we get

$$Q = \left\{ g \in W_0^q(\mathrm{div}; \Omega) \cap L^\infty(\Omega, \mathbb{R}^d) : |g(x)| \leq \alpha(x) \text{ for a.e. } x \in \Omega \right\}.$$

3.1. **Explicit representation of J^{**}.** Now we study J^{**} and aim at providing an explicit representation. Note that since J is convex, J^{**} is equal to the lower semi-continuous envelope of J with respect to both the strong and weak L^p-convergence. In other words, for every $u \in L^p(\Omega)$ we have

$$J^{**}(u) = \inf \left\{ \liminf_{n \to \infty} J(u_n) : u_n \in L^p(\Omega), \, u_n \to u \text{ in } L^p(\Omega) \right\}$$

(3.3)
$$= \inf \left\{ \liminf_{n \to \infty} J(u_n) : u_n \in L^p(\Omega), \, u_n \rightharpoonup u \text{ in } L^p(\Omega) \right\},$$

where "$\to$" refers to convergence with respect to the strong topology and "$\rightharpoonup$" with respect to the weak topology.

In order to obtain an explicit representation of J^{**}, we will employ the representation of the L^1-lower semi-continuous envelope of J, denoted here by $\overline{J}$, as derived in [2]. Note that for $u \in L^p(\Omega)$

(3.4)
$$\overline{J}(u) := \inf \left\{ \liminf_{n \to \infty} J(u_n) : u_n \in W^{1,1}(\Omega), \, u_n \to u \text{ in } L^1(\Omega) \right\}$$

(3.5)
$$= \inf \left\{ \liminf_{n \to \infty} J(u_n) : u_n \in W^{1,1}(\Omega), \, u_n \rightharpoonup u \text{ in } L^1(\Omega) \right\},$$

with the last equality again being true due to convexity of J.

We show that, under suitable coercivity assumptions on j, the functionals J^{**} and $\overline{J}$ coincide.

Lemma 3.6. *Assume $p \in [1, d^*]$ and that for some $c > 0$ we have $c|z| \leq j(x, z)$ for every $z \in \mathbb{R}^d$ and for almost every $x \in \Omega$. Then, $J^{**}(u) = \overline{J}(u)$ for all $u \in L^p(\Omega)$.*

Proof. Since we assume Ω to be bounded, L^p convergence implies L^1 convergence. Also, $J(u_n) = +\infty$ for $u \in L^p(\Omega) \setminus W^{1,1}(\Omega)$ and, consequently, $\overline{J} \leq J^{**}$. Hence we are left to show $J^{**}(u) \leq \overline{J}(u)$ for all $u \in L^p(\Omega)$. To this aim, we first note that, due to the coercivity assumption on j, $c|Du|(\Omega) \leq J(u)$ for all $u \in \mathrm{BV}(\Omega)$.

Now take $a = \lim_{n \to \infty} J(u_n)$ with $u_n \in W^{1,1}(\Omega) \subset \mathrm{BV}(\Omega)$, $u_n \to u$ in $L^1(\Omega)$. Without loss of generality, we can assume $a < \infty$. From the continuous embedding of $\mathrm{BV}(\Omega)$ into $L^{d^*}(\Omega)$, we get for a generic constant $C > 0$ and for some $K > 0$

$$\|u_n\|_{L^p(\Omega)} \leq C(\|u_n\|_{L^1(\Omega)} + |Du_n|(\Omega)) \leq C(\|u_n\|_{L^1(\Omega)} + J(u_n)) < K < \infty.$$

Hence, $(u_n)_{n \in \mathbb{N}}$ also converges weakly (up to subsequences) in $L^p(\Omega)$, and thus also weakly in $L^1(\Omega)$. By uniqueness of the weak limit we get $a = \liminf_{i \to \infty} J(u_{n_i})$ with $u_{n_i} \rightharpoonup u$ in $L^p(\Omega)$ and the proof is complete. $\square$

Remark 3.7. *Note that if the coercivity assumption of Lemma 3.6 holds, then due to the lower semi-continuity of total variation with respect to weak L^1-convergence we get $u \notin \mathrm{BV}(\Omega)$ if and only if $\overline{J}(u) = J^{**}(u) = +\infty$.*

We then get the following result, which is a direct consequence of [2, Theorem 1.1].

Proposition 3.8 (Integral representation of J^{**})**.** *Assume that $p \in [1, d^*]$ and one of the following two assertions holds true:*

(i) $j(x, z) = \alpha(x)b(z)$ with $\alpha \in BV(\Omega)$ and b being a convex function such that (J1)–(J3) hold for j.

(ii) There exists a constant $c > 0$ such that for every $z \in \mathbb{R}^d$ and for almost every $x \in \Omega$, $c|z| \le j(x, z)$ and also $j(\cdot, z) \in BV(\Omega)$.

Then, for any $u \in BV(\Omega)$ we have

(3.6)
$$J^{**}(u) = \int_\Omega j(x, \nabla u(x)) \, dx + \int_\Omega j^-(x, \sigma_{D^c u}) \, d|D^c u| + \int_{J_u \cap \Omega} ((u^+(x) - u^-(x))j^-(x, \sigma_{D^j u}) \, d\mathcal{H}^{d-1}.$$

Proof. Denote by $\mathcal{J}$ the right-hand side of equation (3.6). If one of the two assumptions is satisfied, then it follows from [2, Theorem 1.1] that $\overline{J}(u) = \mathcal{J}(u)$ for every $u \in BV(\Omega)$. It hence remains to show that $J^{**}(u) = \overline{J}(u)$ for any $u \in BV(\Omega)$. In case (ii) is satisfied, this is the assertion of Lemma 3.6. Assume now that (i) holds true. In that case we will show directly that $J^{**}(u) = \mathcal{J}(u)$ for every $u \in BV(\Omega)$. It follows from [2, Theorem 3.1] that $\mathcal{J}$ is lower semi-continuous with respect to L^1 convergence. Consequently it is also lower semi-continuous with respect to weak L^p convergence and hence, for all u and $(u_n)_{n \in \mathbb{N}}$ in $W^{1,1}(\Omega)$ with $u_n \rightharpoonup u$ in L^p we get

$$\mathcal{J}(u) \le \liminf_{n \to \infty} J(u_n).$$

This means that $\mathcal{J}(u)$ is a lower bound for the set

$$\left\{ \liminf_{n \to \infty} J(u_n) : u_n \in W^{1,1}(\Omega), \, u_n \rightharpoonup u \text{ in } L^p(\Omega) \right\}$$

and since, by definition of the lower semi-continuous relaxation, $J^{**}(u)$ is the infimum of this set, we get

$$\mathcal{J}(u) \le J^{**}(u).$$

Furthermore, in the proof of [2, Theorem 4.1] it is shown, with $u_h = u * \phi_h$ being a standard mollification of $u \in BV(\Omega)$ and any $A \subset\subset \Omega$ with $|Du|(\partial A) = 0$, that

$$\liminf_{h \to \infty} J(u_h, A) \le \mathcal{J}(u),$$

where, for any $A \subset \Omega$ open, $J(u, A) := \int_A j(x, \nabla u(x)) \, dx$ if $u \in W^{1,1}(\Omega)$ and $J(u, A) = +\infty$ otherwise. Denoting, for fixed $A \subset \Omega$, $\overline{J}^{d^*}(\cdot, A)$ the lower semi-continuous relaxation of $J(\cdot, A)$ in L^{d^*}, we get from convergence of u_h to u in $L^{d^*}(\Omega)$ as $h \to +\infty$ (which holds, since by embedding $u \in L^{d^*}(\Omega)$) that

$$\overline{J}^{d^*}(u, A) \le \liminf_{h \to \infty} J(u_h, A) \le \mathcal{J}(u).$$

Now by [12, Theorem 4.1], the function $A \mapsto \overline{J}^{d^*}(u, A)$ can be extended to a non-negative Borel measure. We define the uncountable family of open sets $(\Omega_\epsilon)_{\epsilon > 0}$ by $\Omega_\epsilon := \{x \in \Omega : d(x, \partial\Omega) > \epsilon\}$. Then $\Omega_\epsilon \subset\subset \Omega$ and $\partial\Omega_\epsilon \cap \partial\Omega_{\epsilon'} = \emptyset$. By sigma additivity and since $\overline{J}^{d^*}(u, \Omega) < +\infty$ we get that, for any $n \in \mathbb{N}$, the set

$$\{\epsilon > 0 : \overline{J}^{d^*}(u, \partial\Omega_\epsilon) > 1/n\}$$

is finite. Hence, the set $\{\epsilon > 0 : \overline{J}^{d^*}(u, \partial\Omega_\epsilon) > 0\}$ is at most countable and we can extract a sequence $(\Omega_{\epsilon_i})_{i \in \mathbb{N}}$ such that $\overline{J}^{d^*}(u, \partial\Omega_{\epsilon_i}) = 0$, $(\epsilon_i)_{i \in \mathbb{N}}$ is monotonically decreasing and $\bigcup_i \Omega_{\epsilon_i} = \Omega$. Hence we conclude (by $p \le d^*$ for the left-most inequality below)

$$J^{**}(u) \le \overline{J}^{d^*}(u, \Omega) = \lim_{i \to \infty} \overline{J}^{d^*}(u, \Omega_{\epsilon_i}) \le \mathcal{J}(u)$$

and the proof is complete. $\qquad\square$

Remark 3.9. *In the case $j(x, z) = \alpha(x)|z|$ with $0 \leq \alpha(x) \leq C$ and $\alpha \in \mathrm{BV}(\Omega)$, the assumptions of the above proposition are fulfilled. Using that, in this case $(j^-) = \alpha^- |\cdot|$, we get that*

$$J^{**}(u) = \int_\Omega \alpha^- |\nabla u(x)| \, dx + \int_\Omega \alpha^- |\sigma_{D^c u}| \, d|D^c u| + \int_{J_u \cap \Omega} ((u^+ - u^-)\alpha^- |\sigma_{D^j u}| \, d\mathcal{H}^{d-1}$$

$$= \int_\Omega \alpha^- \, d|Du|.$$

We provide further remarks concerning the case $j(x, z) = \alpha(x)|z|$ in Section 4 below.

3.2. Subdifferential characterization.
From Proposition 3.4 and Corollary 3.5 we directly obtain an integral characterization of the subdifferential for a wide class of structural total variation type functionals. The corresponding result is established in Proposition 3.10 below. Also, we refer to [4, 13, 16] for more elaborate results in that direction.

Proposition 3.10. *Let $p \in (1, \infty)$. Then, for $u \in L^p(\Omega)$ we have $u^* \in \partial J^{**}(u) \subset L^q(\Omega)$ if and only if*

(i) $J^{**}(u) < +\infty$, and
(ii) there exists a vector field $g \in W_0^q(\mathrm{div}; \Omega) \cap L^\infty(\Omega, \mathbb{R}^d)$ such that

$$u^* = \mathrm{div} g, \qquad J^{**}(u) = \int_\Omega u \, \mathrm{div} g \, dx, \qquad j^\circ(x, g(x)) \leq 1 \quad \text{for a.e. } x \in \Omega.$$

Proof. Let $u \in L^p(\Omega)$. We have (see, e.g., [23])

$$u^* \in \partial J^{**}(u) \iff J^{**}(u) + J^{***}(u^*) = \int_\Omega u^* u \, dx$$

$$\iff J^{**}(u) + J^*(u^*) = \int_\Omega u^* u \, dx$$

$$\iff J^{**}(u) + \mathcal{I}_{\mathrm{div}(Q)}(u^*) = \int_\Omega u^* u \, dx,$$

where Q is defined in Corollary 3.5 as

$$Q = \left\{ g \in W_0^q(\mathrm{div}; \Omega) \cap L^\infty(\Omega, \mathbb{R}^d) : j^\circ(x, g(x)) \leq 1 \text{ for a.e. } x \in \Omega \right\}.$$

Hence the result follows immediately. □

Recall that if j satisfies the coercivity assumption $c|z| \leq j(x, z)$, then Remark 3.7 implies that $J^{**}(u) < +\infty$ is equivalent to $u \in \mathrm{BV}(\Omega)$.

4. SOME REFINEMENTS FOR WEIGHTED TV

For the special case $j(x, z) = \alpha(x)|z|$ we next investigate alternative dual definitions of the structural total variation functional. This will provide some density results for pointwise bounded $W_0^d(\mathrm{div}; \Omega)$-functions, which are of interest in a variety of applications [31, 32, 33, 34].

Given $u \in L^p(\Omega)$ with $p \in (1, d^*]$, consider the following two extended-valued weighted TV-functionals:

$$(4.1) \qquad \mathrm{TV}_\alpha^W(u) := \sup \left\{ \int_\Omega u \, \mathrm{div} g \, dx : g \in W_0^q(\mathrm{div}; \Omega), \, |g(x)| \leq \alpha(x), \text{ for a.e. } x \in \Omega \right\},$$

$$(4.2) \qquad \mathrm{TV}_\alpha^C(u) := \sup \left\{ \int_\Omega u \, \mathrm{div} \phi \, dx : \phi \in C_c^\infty(\Omega, \mathbb{R}^d), \, |\phi(x)| \leq \alpha(x), \, \forall x \in \Omega \right\},$$

Recall that $\mathrm{TV}_\alpha^W = J^{**}$ by Corollary 3.5.

11

We commence by stating a preparatory result whose proof can be found in [31].

Lemma 4.1. *Let $\phi \in C_c(\Omega, \mathbb{R}^d)$. Then for every $\epsilon > 0$ there exists a function $\phi_\epsilon \in C_c^\infty(\Omega, \mathbb{R}^d)$ such that*

$$\|\phi - \phi_\epsilon\|_\infty < \epsilon \quad and \quad |\phi_\epsilon(x)| \leq |\phi(x)| \quad for\ all\ x \in \Omega.$$

Using this density result, the following result can be shown for a continuous weight function α.

Proposition 4.2. *Let $\alpha \in C(\overline{\Omega})$ with $\alpha \geq 0$ and $u \in \mathrm{BV}(\Omega)$. Then*

$$(4.3) \qquad \mathrm{TV}_\alpha^C(u) = \int_\Omega \alpha\, d|Du|.$$

Proof. From [3, Prop. 1.47] we have that

$$
\int_\Omega \alpha\, d|Du| = \sup \left\{ \sum_{i=1}^d \int_\Omega \phi_i \alpha\, dD_i u : \phi \in C_c(\Omega, \mathbb{R}^d),\ |\phi(x)| \leq 1,\ \forall x \in \Omega \right\}
$$

$$(4.4) \qquad = \sup \left\{ \sum_{i=1}^d \int_\Omega \phi_i\, dD_i u : \phi \in C_c(\Omega, \mathbb{R}^d),\ |\phi(x)| \leq \alpha(x),\ \forall x \in \Omega \right\}$$

$$
\leq \sup \left\{ \int_\Omega |\phi|\, d|Du| : \phi \in C_c(\Omega, \mathbb{R}^d),\ |\phi(x)| \leq \alpha(x),\ \forall x \in \Omega \right\}
$$

$$
\leq \int_\Omega \alpha\, d|Du|.
$$

The proof will be completed by showing that (4.4) is equal to

$$(4.5) \qquad \sup \left\{ \sum_{i=1}^d \int_\Omega \phi_i\, dD_i u : \phi \in C_c^\infty(\Omega, \mathbb{R}^d),\ |\phi(x)| \leq \alpha(x),\ \forall x \in \Omega \right\}.$$

Indeed, if $\phi \in C_c^\infty(\Omega, \mathbb{R}^d)$, then $\sum_{i=1}^d \int_\Omega \phi_i\, dD_i u = -\int_\Omega u \operatorname{div}\phi\, dx$ and thus (4.5) is equal to the right-hand side of (4.3). Lemma 4.1 now allows to approximate any term in (4.4) by a term in (4.5) satisfying the pointwise constraint, which completes the proof. $\qquad\square$

As we show next, (4.1) and (4.2) are equivalent for $u \in \mathrm{BV}(\Omega)$ and a continuous weight function.

Proposition 4.3. *Let $\alpha \in C(\overline{\Omega})$ with $\alpha \geq 0$ and $u \in \mathrm{BV}(\Omega)$. Then*

$$(4.6) \qquad \mathrm{TV}_\alpha^C(u) = \mathrm{TV}_\alpha^W(u) = \int_\Omega \alpha\, d|Du|.$$

Proof. We have that for every $u \in \mathrm{BV}(\Omega)$ there exists a sequence $(u_n)_{n \in \mathbb{N}} \subset C^\infty(\Omega) \cap W^{1,1}(\Omega) \cap L^{d^*}(\Omega)$ that converges to u in $L^{d^*}(\Omega)$ and also $\int_\Omega \alpha\, d|Du_n| \to \int_\Omega \alpha\, d|Du|$; see for instance [3, Thm. 3.9 & Prop. 3.15]. This, together with the fact that $\int_\Omega \alpha\, d|D\cdot|$ is lower semi-continuous with respect to L^{d^*}-convergence, and also that J^{**} is the L^{d^*}-lower semi-continuous envelope of J implies that

$$\mathrm{TV}_\alpha^C(u) = \int_\Omega \alpha\, d|Du| = J^{**}(u) = \mathrm{TV}_\alpha^W(u),$$

where the first equality is due to Proposition 4.2 and the last equality is due to Corollary 3.5. $\qquad\square$

Note that in the proof above, we are not able to directly use the result of [2] or its corollaries proven in the previous section, as $\alpha \in C(\overline{\Omega})$ does not imply $\alpha \in \mathrm{BV}(\Omega)$. For smooth (but not necessarily with integrable gradient) u the following analogous result holds true; see Appendix A for its proof.

Proposition 4.4. *Let $\alpha \in C(\overline{\Omega})$ with $\alpha \geq 0$ and $u \in C^1(\Omega)$. Then $\mathrm{TV}_\alpha^C(u) = \int_\Omega \alpha |\nabla u| \, dx$.*

We now reduce the regularity of α by assuming the following property for the function $\alpha^- \geq 0$:

(P_{lsc}) There exists $(\alpha_n)_{n \in \mathbb{N}} \subset C(\overline{\Omega})$, $0 \leq \alpha_n(x) \leq \alpha^-(x)$ such that $\alpha_n(x) \to \alpha^-(x)$, $\forall x \in \Omega$.

Observe that the above requirement slightly generalizes lower semi-continuity.

Proposition 4.5. *Suppose that $\alpha \in \mathrm{BV}(\Omega) \cap L^\infty(\Omega)$, with $\alpha \geq 0$, and α^- satisfies (P_{lsc}). Then for all $u \in \mathrm{BV}(\Omega)$ we have*

$$\mathrm{TV}_{\alpha^-}^C(u) = \mathrm{TV}_\alpha^W(u) = \int_\Omega \alpha^- \, d|Du|. \tag{4.7}$$

Proof. Since α^- satisfies (P_{lsc}), we have $\alpha^-(x) = \sup\{\tilde{\alpha}(x) : \tilde{\alpha} \in C(\Omega), \ \tilde{\alpha} \leq \alpha\}$, which yields

$$\sup_{\substack{\tilde{\alpha} \leq \alpha^- \\ \tilde{\alpha} \in C(\overline{\Omega})}} \int_\Omega \tilde{\alpha} \, d|Du| = \int_\Omega \alpha^- \, d|Du|. \tag{4.8}$$

Moreover, we find

$$\sup_{\substack{\tilde{\alpha} \leq \alpha^- \\ \tilde{\alpha} \in C(\overline{\Omega})}} \int_\Omega \tilde{\alpha} \, d|Du| = \sup_{\substack{\tilde{\alpha} \leq \alpha^- \\ \tilde{\alpha} \in C(\overline{\Omega})}} \sup\left\{\int_\Omega u \, \mathrm{div}\psi \, dx : \psi \in C_c^\infty(\Omega, \mathbb{R}^d), \ |\psi(x)| \leq \tilde{\alpha}(x), \ \forall x \in \Omega\right\}$$

$$\leq \sup\left\{\int_\Omega u \, \mathrm{div}\phi \, dx : \phi \in C_c^\infty(\Omega, \mathbb{R}^d), \ |\phi(x)| \leq \alpha^-(x), \ \forall x \in \Omega\right\}$$

$$= \mathrm{TV}_{\alpha^-}^C(u).$$

But, the above inequality is, in fact, an equality as for every element $\int_\Omega u \, \mathrm{div}\phi \, dx$ with $\phi \in C_c^\infty(\Omega, \mathbb{R}^d)$ and $|\phi| \leq \alpha^-$, we have

$$\int_\Omega u \, \mathrm{div}\phi \, dx \leq \sup\left\{\int_\Omega u \, \mathrm{div}\psi \, dx : \psi \in C_c^\infty(\Omega, \mathbb{R}^d), \ |\psi(x)| \leq |\phi(x)|, \ \forall x \in \Omega\right\}.$$

From this, (4.8), Corollary 3.5 and Remark 3.9, we get that

$$\mathrm{TV}_{\alpha^-}^C(u) = \mathrm{TV}_\alpha^W(u) = \int_\Omega \alpha^- \, d|Du|,$$

which ends the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Under a uniform positivity assumption on the weight, we obtain a density result, which is of use, e.g., when predualizing the renowned Rudin-Osher-Fatemi model; see [47], and [29, 33, 34] for its dualization.

Proposition 4.6. *Let the assumptions of Proposition 4.5 hold true, and assume in addition that $\alpha > c > 0$ a.e. in Ω. Then we have for $q \in [d, +\infty)$ that*

$$\overline{\{\mathrm{div}\phi : \phi \in C_c^\infty(\Omega, \mathbb{R}^d), \ |\phi| \leq \alpha^-\}}^{L^q(\Omega)} = \left\{\mathrm{div}g : g \in W_0^q(\mathrm{div}; \Omega), \ |g| \leq \alpha^-, \ a.e.\right\}.$$

Proof. For every $u \in \mathrm{BV}(\Omega)$ we have

$$\mathrm{TV}_{\alpha^-}^C(u) = \int_\Omega \alpha^- d|Du| = \overline{J}(u) = J^{**}(u) = \mathrm{TV}_\alpha^W(u),$$

where the first equality stems from Proposition 4.5, the second one is due to [2], the third equality comes from Proposition (3.6), and Corollary 3.5 yields the final relation.

As $\alpha > c > 0$ a.e. in Ω, we have $\mathrm{TV}^C_{\alpha^-}(u) = \mathrm{TV}^W_\alpha(u) = +\infty$ for every $u \in L^p(\Omega) \setminus \mathrm{BV}(\Omega)$. Hence, we have $\mathrm{TV}^C_{\alpha^-} = \mathrm{TV}^W_\alpha$ in all of $L^p(\Omega)$ which is equivalent to

$$\mathcal{I}^*_{\left\{\mathrm{div}\phi : \phi \in C^\infty_c(\Omega, \mathbb{R}^d),\, |\phi| \leq \alpha^-\right\}}(u) = \mathcal{I}^*_{\left\{\mathrm{div}g : g \in W^q_0(\mathrm{div};\Omega),\, |g| \leq \alpha^-,\, a.e.\right\}}(u), \quad \text{for all } u \in L^p(\Omega).$$

After dualization and using the fact that the second set in the equation above is closed in $L^q(\Omega)$ (compare [13] for a proof for scalar α which readily carries over to the present setting), we obtain for all $u^* \in L^q(\Omega)$ that

$$\mathcal{I}_{\overline{\left\{\mathrm{div}\phi : \phi \in C^\infty_c(\Omega, \mathbb{R}^d),\, |\phi| \leq \alpha^-\right\}}^{L^q(\Omega)}}(u^*) = \mathcal{I}_{\left\{\mathrm{div}g : g \in W^q_0(\mathrm{div};\Omega),\, |g| \leq \alpha^-,\, a.e.\right\}}(u^*),$$

which proves the claimed density. $\qquad\square$

Density results of the above type have recently gained attention in the literature [31, 32, 33, 34] and enjoy a variety of applications. In the context of variational regularization in image reconstruction, an analogous density result for continuous weights α was used in [33] in order to show equivalence of a weighted TV-regularization problem and a corresponding predual problem; see [29] for a scalar weight. We emphasize that the result of Proposition 4.6 allows for the dualization for a larger class of weights rather than continuous functions. In the following section, we discuss an even more general duality result.

5. A GENERAL DUALITY RESULT

In this section we consider the variational regularization of linear inverse problems using structural total variation regularization. Our goal is to show existence of a solution as well as equivalence to a saddle-point formulation in the continuous setting under mild conditions that are naturally satisfied by the applications of our interest. The saddle-point problem will be formulated in a way that it only requires an explicit form of J^*, but not of J^{**}, and such that its numerical solution by duality-based optimization algorithms is direct.

Since we aim to capture diverse applications, such as structural-TV-regularized MR and PET reconstruction, our main duality result will be rather general with technical assumptions. In order to better demonstrate the essence of our result, we first consider the particular case where data fidelity will be guaranteed by a norm discrepancy. For this purpose, consider

$$(5.1) \qquad \inf_{u \in L^p(\Omega)} J^{**}(u) + \lambda \|Ku - f\|_S,$$

where $p \in (1, d^*]$, $(S, \|\cdot\|_S)$ is a Banach space with $f \in S$, $K : L^p(\Omega) \to S$ is a bounded linear operator (i.e., $K \in \mathcal{L}(L^p(\Omega), S)$), J^{**} corresponds to the structural TV functional as defined above, and $\lambda > 0$ is a regularization parameter.

Note that, without further assumptions (see Propositions 3.4 and 3.8), we only have $J^* = J^{***}$, but J^{**} is not available explicitly. Hence we are interested in showing equivalence of (5.1) to an appropriate predual problem which only requires J^*. We will see that, for general J, this is possible if either $c|z| \leq j(x, z)$ for every $z \in \mathbb{R}^d$ and for almost every $x \in \Omega$, or the inversion of K is essentially well-posed, i.e., K has a closed range and a finite dimensional kernel. Regarding the latter, we note that this is in particular true if we assume K^*K to be invertible, with K^* being the adjoint of K. A first result for the particular setting of (5.1) is stated next.

Proposition 5.1. *Let $p \in (1, d^*]$, $(S, \|\cdot\|_S)$ a Banach space with $f \in S$, $K \in \mathcal{L}(L^p(\Omega), S)$ and $\lambda > 0$. Assume that at least one of the following two conditions holds:*

  (i) There exists $c > 0$ such that $c|z| \leq j(x, z)$ for every $z \in \mathbb{R}^d$ and for almost every $x \in \Omega$.
  (ii) $\mathrm{Rg}(K^*)$ is closed and $\ker(K)$ is finite dimensional.

Then there exists a solution to the primal problem

$$(5.2) \qquad \inf_{u \in L^p(\Omega)} J^{**}(u) + \lambda \|Ku - f\|_S,$$

to a corresponding predual problem, and to the saddle-point problem

$$(5.3) \qquad \inf_{\substack{p \in W_0^q(\mathrm{div};\Omega) \\ p \in Q}} \sup_{u \in L^p(\Omega)} (\mathrm{div}p, u) - \lambda \|Ku - f\|_S,$$

where $Q = \{g \in W_0^q(\mathrm{div};\Omega) \cap L^\infty(\Omega, \mathbb{R}^d) : j^\circ(x, g(x)) \leq 1 \text{ for a.e. } x \in \Omega\}$. They all coincide and are equivalent in the sense that the pair (p, u) is a solution to the saddle-point problem if and only if u is a solution to (5.2) and p is a solution to the predual problem.

Proof. This is a special case of Theorem 5.4 below. We, hence, refer to the corresponding proof. $\square$

We mention that the above result readily carries over to data fidelities of the type $\|Ku - f\|_S^r$, with $r \in [1, \infty)$. Here, often $r = 2$ is of interest when S is a Hilbert space, or $r = p$ when $S = W^{k,p}(\Omega)$, with $k \in \mathbb{N}_0$ and $p \in [1, \infty)$.

Note that the coercivity assumption (i) of Proposition 5.1 excludes the case where $J^{**}(u) = \int_\Omega \alpha \, d|Du|$ with vanishing weight α. The following example shows that if the weight is not bounded uniformly away from zero, existence for (5.2) is not guaranteed, in general (not in $\mathrm{BV}(\Omega)$ but also not even in $L^2(\Omega)$). This observation justifies assumption (i) very well.

Proposition 5.2. *There exists $\Omega \subset \mathbb{R}^d$, $\alpha \in C(\overline{\Omega})$ with α vanishing only at one point, a Banach space S, data $f \in S$ and an injective, bounded linear operator $K : L^2(\Omega) \to S$ such that the minimization problem*

$$(5.4) \qquad \inf_{u \in L^2(\Omega)} \int_\Omega \alpha \, d|Du| + \|Ku - f\|_S,$$

does not have a solution.

Proof. Let $d = 2$ and define $\Omega = [-L, L] \times [-L, L]$. Moreover, let α be a positive continuous function on Ω that vanishes only at the origin and which satisfies

$$(5.5) \qquad \alpha(x) \leq \frac{1}{n^2}, \quad \text{for all } x \in M_n := \left\{ y \in \Omega : |y| = \frac{1}{n} \right\}.$$

Define $S := \left(\overline{C_c^\infty(\Omega)}^{\|\cdot\|} \right)^*$, where $\|\cdot\| := \|\cdot\|_\infty + \|\nabla \cdot\|_\infty$, and $K : L^2(\Omega) \to S$ with

$$Ku(\phi) = \int_\Omega u\phi \, dx, \quad \text{for all } \phi \in \overline{C_c^\infty(\Omega)}^{\|\cdot\|}.$$

Note that K is injective, linear and bounded. In fact, concerning the latter observe

$$\|Ku\|_S = \sup_{\|\phi\| \leq 1} Ku(\phi) = \sup_{\|\phi\| \leq 1} \int_\Omega u\phi \, dx \leq \sup_{\|\phi\| \leq 1} C\|\phi\|_\infty \|u\|_{L^2(\Omega)} \leq \sup_{\|\phi\| \leq 1} C\|\phi\| \|u\|_{L^2(\Omega)} \leq C\|u\|_{L^2(\Omega)},$$

for an appropriate constant $C > 0$. Finally choose $f = \delta_0$, i.e., the Dirac measure at zero. We claim that with this set-up the infimum in (5.4) is zero. Indeed, define $u_n := \frac{1}{\mathrm{meas}(N_n)} \mathcal{X}_{N_n}$ where $N_n := \left\{ x \in \mathbb{R}^2 : |x| \leq \frac{1}{n} \right\}$. Given that $\mathrm{meas}(N_n) = \frac{\pi}{n^2}$, we have

$$\int_\Omega \alpha \, d|Du_n| \leq \frac{1}{n^2} |Du_n|(M_n) \leq \frac{1}{n^2} \frac{n^2}{\pi} \mathcal{H}(M_n) = \frac{1}{n^2} \frac{n^2}{\pi} 2\pi \frac{1}{n} = \frac{2}{n} \to 0.$$

Now for the fidelity term we have

$$\|Ku_n - \delta_0\|_S \le \sup_{\|\phi\| \le 1} \left| \int_\Omega u_n \phi \, dx - \phi(0) \right| \le \sup_{\|\phi\| \le 1} \frac{1}{\operatorname{meas}(N_n)} \int_{N_n} |\phi(x) - \phi(0)| \, dx$$

$$\le \sup_{\|\phi\| \le 1} \frac{\|\nabla \phi\|_\infty}{\operatorname{meas}(N_n)} \int_{N_n} |x|_2 \, dx \le \frac{1}{\operatorname{meas}(N_n)} \int_{N_n} \frac{1}{n} \, dx \to 0.$$

Hence, if there was a minimizer $\tilde{u} \in L^2(\Omega)$, then $K\tilde{u}$ should be equal to δ_0 in the sense that $\int_\Omega \tilde{u}\phi \, dx = \phi(0)$ for every $\phi \in C_c^\infty(\Omega)$. But this is impossible and hence there is no minimizer for the problem (5.4). □

Observe that the operator K in the proof of Proposition 5.2 has no closed range. This can be seen easily as for the sequence $(u_n)_{n \in \mathbb{N}}$, in the proof above, we have $\|Ku_n - \delta_0\|_S \to 0$ and $\delta_0 \notin \operatorname{Rg}(K)$. From the closed range theorem we also get that the range of K^* is not closed. As K is injective, and, in particular, it then has finite dimensional kernel, all other assumptions of Proposition 5.1 are satisfied. Hence the closed range assumption is tight in the sense that, for non-closed range operators we cannot expect a similar result without further assumptions on the integrand j.

Motivated by the particular case of a norm discrepancy as data fidelity, we now consider a more general setting. In fact, let $p \in (1, d^*]$, $\lambda > 0$, $K \in \mathcal{L}(L^p(\Omega), S)$ with S being a Banach space, and assume $D : S \to \overline{\mathbb{R}}$ to be convex and lower semi-continuous. We aim to solve

$$(5.6) \qquad \inf_{u \in L^p(\Omega)} J^{**}(u) + (\lambda D \circ K)(u),$$

where J^{**} again corresponds to the structural TV functional as defined above. We recall first the following result which follows from [46, Theorem 19].

Lemma 5.3. *Let S be a Banach space and $D : S \to \overline{\mathbb{R}}$ convex, lower semi-continuous, and continuous and finite at zero. Further, let $K : L^p(\Omega) \to S$ be a bounded linear operator. Then for all $x^* \in L^q(\Omega)$ we have*

$$(D \circ K)^*(x^*) = \min_{\substack{s^* \in S^* \\ K^* s^* = x^*}} D^*(s^*),$$

where the minimum is attained. Consequently, it holds that $\operatorname{dom}((D \circ K)^*) = K^* \operatorname{dom}(D^*)$.

We will also need the following generalization of an orthonormal decomposition of $L^q(\Omega)$ (see for instance [37, Corollary 6.1]): For $q \in (1, \infty)$, we denote the space of constant functions in $L^q(\Omega)$ by $\ker(\nabla)$. Then, with

$$\ker(\nabla)^\perp := \{u \in L^q(\Omega) : (v, u) = 0 \text{ for all } v \in L^p(\Omega), \text{ with } v \text{ a constant function}\},$$

we get that $\ker(\nabla)^\perp$ is closed, $L^q(\Omega) = \ker(\nabla) + \ker(\nabla)^\perp$ and $\ker(\nabla) \cap \ker(\nabla)^\perp = \{0\}$. In particular, we denote by $P_{\ker(\nabla)} : L^q(\Omega) \to \ker(\nabla)$ the continuous linear projection such that $\operatorname{Rg}(P_{\ker(\nabla)}) = \ker(\nabla)$ and $\ker(P_{\ker(\nabla)}) = \ker(\nabla)^\perp$.

Having these prerequisites at hand, we now arrive at the main duality result of the paper.

Theorem 5.4. *Let $p \in (1, d^*]$, $\lambda > 0$, S a Banach space, and $K \in \mathcal{L}(L^p(\Omega), S)$. Further let $D : S \to \overline{\mathbb{R}}$ be convex, lower semi-continuous, and continuous and finite at zero. Also, assume that at least one of the following two conditions hold:*

(i) $\begin{cases} \text{There exists } c > 0 \text{ such that } c|z| \le j(x, z) \text{ for every } z \in \mathbb{R}^d \text{ and for almost every } x \in \Omega, \\ P_{\ker(\nabla)}\left(K^*\left[\bigcup_{\mu \ge 0} \mu \operatorname{dom}((\lambda D)^*) \right] \right) \text{ is a vector space and } 0 \in \operatorname{dom}((\lambda D)^*). \end{cases}$

16

$(ii) \begin{cases} \mathrm{Rg}(K^*) \text{ is closed, } \ker(K) \text{ is finite dimensional and } D \text{ is coercive.} \end{cases}$

Then there exists a solution to the primal problem

(P) $$\inf_{u \in L^p(\Omega)} J^{**}(u) + (\lambda D \circ K)(u)$$

as well as to the predual problem

(pD) $$\inf_{\substack{p \in W_0^q(\mathrm{div};\Omega) \\ p \in Q}} (\lambda D \circ K)^*(\mathrm{div} p),$$

where $Q = \{g \in W_0^q(\mathrm{div};\Omega) \cap L^\infty(\Omega, \mathbb{R}^d) : j^\circ(x, g(x)) \leq 1 \text{ for a.e. } x \in \Omega\}$, and to the saddle-point problem

(sp) $$\inf_{\substack{p \in W_0^q(\mathrm{div};\Omega) \\ p \in Q}} \sup_{u \in L^p(\Omega)} (\mathrm{div} p, u) - (\lambda D \circ K)(u).$$

Further, these problems all coincide at their optimal values and are equivalent in the sense that the pair (p, u) is a solution to the saddle-point problem (sp) *if and only if u is a solution to* (P) *and p is a solution to* (pD).

Before we provide the proof, let us motivate the rather technical assumptions by showing that they exactly reduce to the setting of Proposition 5.1 if the discrepancy term D is a norm: In this setting, we have $D(v) = \|v - f\|_S$, which is obviously finite and continuous at zero, and it is coercive in S. Furthermore, from [11, Theorem 4.4.10], coercivity (i.e., the first ingredient of assumption (i)) implies that $B_\epsilon(0) \subset \mathrm{dom}((\lambda D)^*)$ for some $\epsilon > 0$. Hence the union over all positive factors times $\mathrm{dom}((\lambda D)^*)$ is all of S^* and since $P_{\ker(\nabla)}$ is linear, the second part of (i) also holds. Thus, the remaining assumptions correspond exactly to the assumptions of Proposition 5.1. We remark that the rather technical assumptions above allow to also cover settings like the one for structural TV-regularized PET reconstruction with positivity constraints.

Proof of Theorem 5.4. We first show strict duality. Note that convex conjugation is always carried out in the space where the functional is defined. Define $X = W_0^q(\mathrm{div};\Omega)$, $Y = L^q(\Omega)$, the bounded linear operator $T : X \to Y$, $Tp = \mathrm{div} p$, and the functionals $F : X \to [0, \infty]$, $F(p) = \mathcal{I}_Q(p)$, and $G : Y \to [0, \infty]$, $G(u^*) = (\lambda D \circ K)^*(u^*)$. Our goal is to show the following duality relation, which also asserts existence of a minimum for the right-hand side:

$$\inf_{x \in X} F(x) + G(Tx) = - \min_{y \in Y^*} F^*(-T^*y) + G^*(y).$$

For the right-hand side we get

$$- \min_{y \in Y^*} F^*(-T^*y) + G^*(y) = - \min_{u \in L^p(\Omega)} \mathcal{I}_Q^*(-(\mathrm{div})^* u) + (\lambda D \circ K)(u)$$

$$= - \min_{u \in L^p(\Omega)} \sup_{z \in W_0^q(\mathrm{div};\Omega)} \{(-\mathrm{div} z, u) - \mathcal{I}_Q(z)\} + (\lambda D \circ K)(u)$$

$$= - \min_{u \in L^p(\Omega)} J^{**}(u) + (\lambda D \circ K)(u).$$

To show the duality relation, according to [6], it suffices to show that

(5.7) $$U := \bigcup_{\mu \geq 0} \mu \left[\mathrm{dom}((\lambda D \circ K)^*) - T \mathrm{dom}(F) \right] = \bigcup_{\mu \geq 0} \mu \left[K^* \mathrm{dom}((\lambda D)^*) - \mathrm{div}(Q) \right] \subset L^q(\Omega)$$

17

is a closed vector space. Note that the second equality holds true due to Lemma 5.3. Now first consider the case that assumption (i) holds. We claim that in this case

$$U = P_{\ker(\nabla)}\left(K^*\left[\bigcup_{\mu \geq 0} \mu \operatorname{dom}((\lambda D)^*)\right]\right) + \ker(\nabla)^\perp$$

and hence, by assumption (i), U is a closed vector space for being the sum of a finite dimensional and a closed vector space. It is clear that U is a subset of the right-hand side since

$$\bigcup_{\mu \geq 0} \mu\left[K^* \operatorname{dom}((\lambda D)^*) - \operatorname{div}(Q)\right] \subset \bigcup_{\mu \geq 0} \mu K^* \operatorname{dom}((\lambda D)^*) - \bigcup_{\mu \geq 0} \mu \operatorname{div}(Q)$$

$$\subset K^*\left[\bigcup_{\mu \geq 0} \mu \operatorname{dom}((\lambda D)^*)\right] + \ker(\nabla)^\perp$$

$$\subset P_{\ker(\nabla)}\left(K^*\left[\bigcup_{\mu \geq 0} \mu \operatorname{dom}((\lambda D)^*)\right]\right)$$

$$+ P_{\ker(\nabla)^\perp}\left(K^*\left[\bigcup_{\mu \geq 0} \mu \operatorname{dom}((\lambda D)^*)\right]\right) + \ker(\nabla)^\perp$$

$$\subset P_{\ker(\nabla)}\left(K^*\left[\bigcup_{\mu \geq 0} \mu \operatorname{dom}((\lambda D)^*)\right]\right) + \ker(\nabla)^\perp.$$

For the reverse inclusion, we employ a result of [37] which states that for any $\delta > 0$ there is an $\epsilon > 0$ such that $B_\epsilon(0) \cap \ker(\nabla)^\perp \subset \{\operatorname{div} g : g \in W_0^q(\operatorname{div}; \Omega), |g(x)| \leq \delta$ for a.e. $x \in \Omega\}$. In fact, this was shown in [37] for $p = d^*$, but the extension to $1 < p < d^*$ is direct.

Take any $u = P_{\ker(\nabla)}(\mu K^* v) + w$ with $\mu \geq 0$, $v \in \operatorname{dom}((\lambda D)^*)$ and $w \in \ker(\nabla)^\perp$. We re-write $u = \mu K^* v - (P_{\ker(\nabla)^\perp}(\mu K^* v) - w) = \mu K^* v - \tilde{w}$ with $\tilde{w} \in \ker(\nabla)^\perp$. By assumption (i), we get that $J^*(u^*) \leq c\,\mathrm{TV}^*(c^{-1} u^*)$ for all $u^* \in L^q(\Omega)$, which means that if $c^{-1} u^* \in \operatorname{dom}(\mathrm{TV}^*)$, i.e., $u^* = \operatorname{div} g_1$ with $\|g_1\|_{L^\infty(\Omega)} \leq c$ then $u^* \in \operatorname{dom}(J^*) = \operatorname{div}(Q)$, i.e., $u^* = \operatorname{div} g_2$ with $j^\circ(x, g_2(x)) \leq 1$. Taking hence ϵ sufficiently small such that $\epsilon \tilde{w} = \operatorname{div} g$ with $\|g\|_{L^\infty(\Omega)} \leq c$, we get that $\epsilon \tilde{w} \in \operatorname{div}(Q)$. Now define $\eta = \min\{1/\mu, \epsilon\}$. Then $\eta \mu v \in \operatorname{dom}((\lambda D)^*)$ by convexity of $(\lambda D)^*$ and assumption (i). Consequently we get

$$u = \frac{1}{\eta}(\eta \mu K^* v - \eta \tilde{w}) \in U,$$

as claimed.

Before considering condition (ii), we show that $V := \bigcup_{\mu \geq 0} -\mu \operatorname{div}(Q)$ is a vector space. To see this, take $\lambda \in \mathbb{R}$ and $y_1 := -\mu_1 T(z_1)$ and $y_2 := -\mu_2 T(z_2)$ with $z_1, z_2 \in Q$ and $\mu_1, \mu_2 \geq 0$. We show that $\lambda y_1 + y_2 \in V$. We note that, by definition, Q is convex and also $\lambda z \in Q$ for $z \in Q$ and $\lambda \in [0, 1]$. Further, from assumption (J3) on j, one can check that $-z \in Q$ for $z \in Q$. Define $\eta = \max\{\mu_1, \mu_2\} \max\{|\lambda|, 1\}$. Then $\frac{\lambda \mu_1 z_1}{\eta} \in Q$ and $\frac{\mu_2 z_2}{\eta} \in Q$. Hence by convexity of Q, we have

$$\lambda y_1 + y_2 = -(\lambda \mu_1 T z_1 + \mu_2 T z_2) = 2\eta\left(-T\left(\frac{\lambda \mu_1 z_1 + \mu_2 z_2}{2\eta}\right)\right) \in V.$$

Suppose now condition (ii) holds. We will show that in this case U can be written as a sum of $\ker(K)^\perp$ and a finite dimensional space. Indeed, first note that, following [14, Section 2.4, Example 2], $\ker(K)^\perp$ is of finite codimension and, consequently, there exists a finite dimensional space $\tilde{U} \subset L^q(\Omega)$ such that $L^q(\Omega)$ is the direct sum of $\ker(K)^\perp$ and $\tilde{U}$. Thus we can define the

18

canonical continuous linear projection onto $\tilde{U}$, $P_{\tilde{U}} : L^q(\Omega) \to \tilde{U}$. We claim now that

$$U = \ker(K)^\perp + P_{\tilde{U}}(V),$$

which is again a closed vector space for being the sum of a closed and a finite dimensional vector space. First we note that by the close range theorem we have $\ker(K)^\perp = \mathrm{Rg}(K^*)$, and hence U is included in the right-hand side since

$$U \subset \mathrm{Rg}(K^*) + V \subset \mathrm{Rg}(K^*) + P_{\ker(K)^\perp}(V) + P_{\tilde{U}}(V) = \ker(K)^\perp + P_{\tilde{U}}(V).$$

To show the other subset inclusion, take $u = K^*v + P_{\tilde{U}}(-\mu w)$ with $w \in \mathrm{div}(Q)$. Again we re-write $u = K^*v - P_{\ker(K)^\perp}(-\mu w) - \mu w = K^*\tilde{v} - \mu w$ with $\tilde{v} \in S^*$. Again by [11, Theorem 4.4.10], coercivity of λD implies continuity of $(\lambda D)^*$ at 0. Hence there exists $\epsilon > 0$ such that $\overline{B_\epsilon(0)} \subset \mathrm{dom}((\lambda D)^*)$. Setting $\lambda = \min\{\epsilon/\|\tilde{v}\|, 1/\mu\}$ we can write

$$u = \frac{1}{\lambda}(K^*(\lambda \tilde{v}) - \lambda \mu w) \in U$$

since $\eta w \in \mathrm{div}(Q)$ for $\eta \in [0,1]$. This shows strict duality and existence for (P) under assumption (i) or (ii).

Now we show existence for the predual problem. From the continuity of D at 0 and since K is bounded, $\lambda D \circ K$ is continuous at 0 and again by [11, Theorem 4.4.10] $(\lambda D \circ K)^*$ is coercive in $L^q(\Omega)$. We show that the set Q is bounded with respect to the L^q-norm. For this purpose, we define $C = 2\gamma > 0$, where γ is the constant from assumption (J2), and observe that, for any $g \in Q$ and almost every $x \in \Omega$,

$$C^{-1}|g(x)| = \sup_{\substack{\tilde{z}^* \in \mathbb{R}^d \\ |\tilde{z}^*| \leq (C/\gamma)-1}} C^{-1}g(x) \cdot \tilde{z}^* \overset{(*)}{\leq} \sup_{\substack{\tilde{z}^* \in \mathbb{R}^d \\ j(x,\tilde{z}^*) \leq C}} g(x) \cdot C^{-1}\tilde{z}^* = \sup_{\substack{z^* \in \mathbb{R}^d \\ j(x,z^*) \leq 1}} g(x) \cdot z^* = j^\circ(x, g(x)) \leq 1.$$

Indeed, since by (J2), $j(x,z) \leq \gamma(1 + |z|)$, $(*)$ holds true for $C > \gamma$ since then, for $|z^*| \leq (C/\gamma) - 1$, $j(x,z^*) \leq \gamma(1 + |z^*|) \leq C$. This implies that $\|g\|_{L^\infty(\Omega)} \leq C$ and hence also Q is bounded in $L^q(\Omega)$.

Now taking $(p_n)_{n \in \mathbb{N}}$ an infimizing sequence for the predual problem, we get by boundedness of Q and coercivity of $(\lambda D \circ K)^*$ that there exist $\hat{p} \in L^q(\Omega, \mathbb{R}^d)$ and $w \in L^q(\Omega)$ such that, up to subsequences, $p_n \rightharpoonup \hat{p}$ and $\mathrm{div}(p_n) \rightharpoonup w$. This implies that $p \in W^q(\mathrm{div}; \Omega)$ and $\mathrm{div}p = w$. Further we note that $\{(g, \mathrm{div}g) : g \in W_0^q(\mathrm{div}; \Omega) \cap Q\}$ is a convex and closed subset of $L^q(\Omega, \mathbb{R}^{d+1})$, hence it is also weakly closed. By weak convergence of $(p_n, \mathrm{div}p_n)$ to $(\hat{p}, \mathrm{div}\,\hat{p})$ and lower semi-continuity of $(\lambda D \circ K)$ with respect to weak convergence in $L^q(\Omega, \mathbb{R}^d)$ it follows that $\hat{p}$ is a solution to (pD).

Finally, [23, Proposition III.3.1] guarantees equivalence to the saddle-point problem (sp) as claimed and the proof is complete. $\qquad\square$

This concludes the main existence and duality result of the paper, from which the application to linear inverse problems with norm discrepancy follows as a special case. A second situation we want to consider in more detail is an inverse problem where the measured data describes the physical density of some quantity and is corrupted by Poisson noise. The main application we have in mind for this setting is PET image reconstruction, where the Poisson log-likelihood and positivity constraints are used for data fidelity.

5.1. **A Poisson noise model.** We are now interested in the problem

$$(5.8) \qquad \min_{u \in L^p(\Omega)} J^{**}(u) + \lambda \int_\Sigma (Ku)(\sigma) - f(\sigma)\log((Ku)(\sigma) + c_0(\sigma))\,d\sigma + \mathcal{I}_{[0,\infty)}(u),$$

where K is a linear operator (a slightly modified Radon transform), $f \geq 0$ is the given data, $c_0 > 0$ is an estimate for measurements due to scattering and random events and Σ is a subset of $\mathbb{R}^{n_d}$ with

some $n_d \geq 1$. The function $\mathcal{I}_{[0,\infty)}$ constrains the unknown to the non-negative reals. Note that in PET imaging with real data, the estimate c_0 of scatter and random events is an integral part of the (so-called) reconstruction pipeline, as it describes a non-negligible part of the data. Such an estimate is typically delivered by the scanner software as preprocessing step.

More specifically, in what follows we invoke the following data assumptions:

$$\text{(AP)} \quad \begin{cases} (i) \quad \text{meas}(\Sigma) < +\infty, \quad f, c_0 \in L^1(\Sigma), \ f \geq 0, \ c_0 > 0, \\[4pt] (ii) \ \sigma \mapsto f(\sigma)\log(c_0(\sigma)) \in L^1(\Sigma), \ \sigma \mapsto \frac{f(\sigma)}{c_0(\sigma)} \in L^\infty(\Sigma), \\[4pt] (iii) \ K \in \mathcal{L}(L^p(\Omega), L^1(\Sigma)), \quad Ku \geq 0 \text{ whenever } u \geq 0, \\[4pt] (iv) \ \text{the constant functions are contained in } \text{Rg}(K^*). \end{cases}$$

When reading the above integrability assumptions on the data and the scatter and random events, one has to keep in mind that typically f describes the Radon transform of some density which is supported in the interior of Ω and c_0 some defects that increase with the density, but are present on every measurement line. We believe that in such a context, the posed assumptions are realistic and not very restrictive. Clearly, they are satisfied if we assume f and c_0 to be bounded and c_0 to be uniformly bounded away from zero, which is often assumed in the literature for the measured data in order to obtain stability results [48]. In this respect one has to keep in mind, however, that the Radon transform of any function which is bounded above will converge to zero for measurement lines whose length converges to zero. Hence we believe that there is a benefit in using the assumption (AP) rather than uniform boundedness away from zero.

Regarding the assumptions on the forward operator K, we recall in the following some basic properties of the Radon transform, which in particular show that for such a K assumption (AP) is fulfilled. In fact, the classical Radon transform R is a bounded linear operator from $L^1(\mathbb{R}^d)$ to $L^1(\mathcal{S}^{d-1} \times \mathbb{R})$; see for instance [42]. Here $\mathcal{S}^{d-1} \times \mathbb{R}$ is equipped with the measure $\mu := \mathcal{H}^{d-1} \lfloor \mathcal{S}^{d-1} \times \mathcal{L}$. In this context, $d\sigma$ denotes integration with respect the measure μ. Define now the bounded linear operator $E : L^p(\Omega) \to L^1(\mathbb{R}^d)$ as the extension-by-zero outside Ω. Then, if $K := R \circ E$ we have that K is linear, it maps $L^p(\Omega)$ to $L^1(\Sigma)$, with Σ a bounded subset of $\mathcal{S}^{d-1} \times \mathbb{R}$, and it is bounded, since for every $u \in L^p(\Omega)$

$$\|Ku\|_{L^1(\Sigma)} = \|(R \circ E)(u)\|_{L^1(\Sigma)} \leq C\|Eu\|_{L^1(\mathbb{R}^d)} = C\|u\|_{L^1(\Omega)} \leq \tilde{C}\|u\|_{L^p(\Omega)}.$$

According to [28], the adjoint $R^* : L^\infty(\Sigma) \to L^\infty(\mathbb{R}^d)$ of the Radon transform is given by

$$R^*v(x) = \int_{\mathcal{S}^{d-1}} v(\theta, \theta \cdot x)\, d\mathcal{H}^{d-1}(\theta), \quad x \in \mathbb{R}^d.$$

One can easily see now that $K^* : L^\infty(\Sigma) \to L^q(\Omega)$ is simply the restriction of R^* in Ω, since for every $u \in L^p(\Omega)$ and every $v \in L^\infty(\Sigma)$ we have

$$\int_\Omega uR^*v\, dx = \int_{\mathbb{R}^d} EuR^*v\, dx = \int_\Sigma R(Eu)v\, d\sigma = \int_\Sigma (Ku)v\, d\sigma.$$

In this case, the constant functions belong to $\text{Rg}(K^*)$. Moreover, from the definition of the Radon transform it follows immediately that $Ku \geq 0$ for $u \geq 0$ with both inequalities to be understood in the almost everywhere sense. Hence, assumption (AP) holds true when K is the Radon transform. Note also that the assumption on Σ is valid since due to Ω being bounded, for every $u \in L^1(\Omega)$, we have that Ku is supported in a fixed compact subset of Σ.

Now considering the formal minimization problem for Poisson-corrupted data (5.8) in view of our general duality result, some issues arise. First of all we need to rigorously define the data discrepancy term as a function from $L^p(\Omega)$ to the extended reals, and secondly neither the data discrepancy nor the positivity constraint will be continuous in $L^p(\Omega)$ around 0.

The following modification of the data term resolves some of these issues without changing the original problem. For $f \in [0, \infty)$ and $c_0 \in (0, \infty)$ we define the integrand

$$l_{f,c_0}(t) := \begin{cases} t - f \log(t + c_0) & \text{if } t \geq 0, \\ -f \log(c_0) + \left(1 - \frac{f}{c_0}\right)t & \text{if } t < 0, \end{cases}$$

and for $f : \Sigma \to [0, \infty)$ and $c_0 : \Sigma \to (0, \infty)$ the functional

$$D_{\mathrm{KL}} : L^1(\Sigma) \to \overline{\mathbb{R}}, \quad v \mapsto \int_\Sigma l_{f(\sigma), c_0(\sigma)}(v(\sigma)) \, d\sigma + L\|v_-\|_{L^1(\Sigma)},$$

where we set $L := \|1 - \frac{f}{c_0}\|_\infty + 1$ and $v_- := \min\{v, 0\}$ in a pointwise almost everywhere sense. The point in the definition of D_{KL} is that we change the original data fidelity only in points where Ku is negative, which, however, can never occur due to the positivity constraint in u and $Ku \geq 0$ for $u \geq 0$. Hence, when considering the minimization problem

$$(5.9) \qquad \inf_{u \in L^p(\Omega)} J^{**}(u) + \lambda D_{\mathrm{KL}}(Ku) + \mathcal{I}_{[0, \infty)}(u),$$

it is immediate that the set of optimal solutions is exactly the same as for (5.8). The modified fidelity D_{KL} enjoys the following properties.

Lemma 5.5. *Assume that* (AP) *holds. Then D_{KL} is convex and continuous in $L^1(\Sigma)$. Further there exist constants $M, N > 0$ such that*

$$\|v\|_{L^1(\Sigma)} \leq M D_{KL}(v) + N \quad \text{for all } v \in L^1(\Sigma).$$

Proof. Regarding convexity, we note that $v \mapsto \|v_-\|_{L^1(\Sigma)}$ is convex and hence it suffices to show convexity of the integrand $z \mapsto l_{f(\sigma), c_0(\sigma)}(z)$ for $\sigma \in \Sigma$ fixed. To this aim, we note that convexity is equivalent to the derivative of $l_{f(\sigma), c_0(\sigma)}$ being monotonously increasing. The latter is indeed true since for $z > 0$ and $z < 0$ the integrand is convex and since, as can be readily checked, the left and right derivative of $l_{f(\sigma), c_0(\sigma)}(\cdot)$ at $z = 0$ coincide.

Regarding continuity, for $v \in L^1(\Sigma)$ we denote $\Sigma_1 = \{\sigma \in \Sigma : v(\sigma) \geq 0\}$, $\Sigma_2 = \Sigma \setminus \Sigma_1$ and estimate

$$D_{\mathrm{KL}}(v) \leq \int_{\Sigma_1} v(\sigma) - f(\sigma) \log(c_0(\sigma)) \, d\sigma + \int_{\Sigma_2} -f(\sigma) \log(c_0(\sigma)) + \left(1 - \frac{f(\sigma)}{c_0(\sigma)}\right) v(\sigma) \, d\sigma + L\|v_-\|_{L^1(\Sigma)}$$

$$\leq L\|v\|_{L^1(\Sigma)} + \|f \log(c_0)\|_{L^1(\Sigma)} + \left(1 + \left\|\frac{f}{c_0}\right\|_\infty\right) \|v_-\|_{L^1(\Sigma)}$$

$$\leq \left(1 + \left\|\frac{f}{c_0}\right\|_\infty + L\right) \|v\|_{L^1(\Sigma)} + \|f \log(c_0)\|_{L^1(\Sigma)}.$$

Hence D_{KL} is bounded above in a neighborhood of any point and, for being a convex function [11, Prop. 4.1.4] it is also continuous at any point.

To show the coercivity estimate, first pick any $v \in L^1(\Sigma)$ with $v \leq 0$. Choosing $c = \|1 - \frac{f}{c_0}\|_\infty$ we get

$$(5.10) \quad \int_\Sigma -f \log(c_0) + \left(1 - \frac{f}{c_0}\right) v \, d\sigma + L\|v\|_{L^1(\Sigma)} \geq -\|f \log(c_0)\|_{L^1(\Sigma)} - c\|v\|_{L^1(\Sigma)} + L\|v\|_{L^1(\Sigma)}$$

$$= -\|f \log(c_0)\|_{L^1(\Sigma)} + \|v\|_{L^1(\Sigma)}.$$

Further, for $v \in L^1(\Sigma)$ with $v \geq 0$ the Poisson log-likelihood can be estimated below as follows [10, 44]: It is easy to check by differentiating that the function

$$t \mapsto \left(\frac{4}{3} + \frac{2t}{3}\right)(t\log(t) - t + 1) - (t-1)^2, \quad t \geq 0,$$

is convex and attains its minimum value of 0 at $t = 1$. Setting $t = \frac{f(\sigma)}{v(\sigma)+c_0(\sigma)}$ we get that

$$(v(\sigma) + c_0(\sigma) - f(\sigma))^2 \leq \left[\frac{2}{3}f(\sigma) + \frac{4}{3}(v(\sigma) + c_0(\sigma))\right]\left(f(\sigma)\log(\frac{f(\sigma)}{v(\sigma) + c_0(\sigma)}) - f(\sigma) + v(\sigma) + c_0(\sigma)\right)$$

$$= \left[\frac{2}{3}f(\sigma) + \frac{4}{3}(v(\sigma) + c_0(\sigma))\right]\left(v(\sigma) - f(\sigma)\log(v(\sigma) + c_0(\sigma))\right)$$

$$+ c_0(\sigma) - f(\sigma) + f(\sigma)\log(f(\sigma))\bigg).$$

Now taking the square root on both sides, integrating and using the Cauchy-Schwarz inequality we get that

$$\|v + c_0 - f\|_{L^1(\Sigma)}^2 \leq \left[\frac{2}{3}\|f\|_{L^1(\Sigma)} + \frac{4}{3}\|v + c_0\|_{L^1(\Sigma)}\right]\left(\int_\Sigma v - f\log(v + c_0)\,d\sigma + \int_\Sigma c_0 - f + f\log(f)\,d\sigma\right).$$

Further, using that $\|v + c_0\|_{L^1(\Sigma)}^2 \leq 2\|v + c_0 - f\|_{L^1(\Sigma)}^2 + 2\|f\|_{L^1(\Sigma)}^2$ and denoting by M, N generic constants with $M > 0$ we get

$$\|v + c_0\|_{L^1(\Sigma)}\left(\|v + c_0\|_{L^1(\Sigma)} - \frac{8}{3}\left(\int_\Sigma v - f\log(v + c_0)\,d\sigma\right) - N\right) \leq M\left(\int_\Sigma v - f\log(v + c_0)\,d\sigma\right) + N.$$

Now in case $\left(\|v + c_0\|_{L^1(\Sigma)} - \frac{8}{3}\left(\int_\Sigma v - f\log(v + c_0)\,d\sigma\right) - N\right) \leq 1$ we get

$$\|v + c_0\|_{L^1(\Sigma)} \leq \frac{8}{3}\left(\int_\Sigma u - f\log(u + c_0)\,d\sigma\right) + N + 1.$$

In the other case we get

$$\|v + c_0\|_{L^1(\Sigma)} \leq M\left(\int_\Sigma v - f\log(v + c_0)\,d\sigma\right) + N.$$

Hence in any case there exists constants $M > 0$, $N \in \mathbb{R}$ such that

(5.11) $$\|v\|_{L^1(\Sigma)} \leq M\left(\int_\Sigma v - f\log(v + c_0)\,d\sigma\right) + N.$$

Now splitting an arbitrary $v \in L^1(\Sigma)$ into its positive and negative parts, and using the estimates (5.10), (5.11) accordingly, we get that there exists $M > 0$, $N \in \mathbb{R}$ such that

$$\|v\|_{L^1(\Sigma)} \leq M D_{\mathrm{KL}}(v) + N,$$

which completes the proof. $\qquad\square$

Using these properties of the data discrepancy, we obtain the following existence result by standard arguments.

Proposition 5.6. *Assume that* (AP) *is fulfilled and there exists $c > 0$ such that $c|z| \leq j(x, z)$ for every $z \in \mathbb{R}^d$ and for almost every $x \in \Omega$. Then there exists a solution to*

(5.12) $$\inf_{u \in L^p(\Omega)} J^{**}(u) + \lambda D_{KL}(Ku) + \mathcal{I}_{[0,\infty)}(u).$$

22

Proof. We only provide a sketch: Let $(u_n)_{n\in\mathbb{N}}$ be an infimizing sequence for (5.12). Since $\ker(K) = \mathrm{Rg}(K^*)^\perp$ [14, Corollary 2.18], K does not vanish on constant functions. Hence, we get from the equivalence of J^{**} to TV and the coercivity of D_{KL} (Lemma 5.5) as well as norm equivalence in finite dimensional spaces that $(u_n)_{n\in\mathbb{N}}$ is bounded in $L^p(\Omega)$. Hence there exists a weakly convergent subsequence and from weak lower semi-continuity of all terms appearing in the objective functional, existence follows readily.

For applying the general duality result of Theorem 5.4 to the situation of Poisson noise, the only issue that remains is the fact that functional realizing the positivity constraint is not continuous at zero. To overcome this, we will introduce a slight modification of this functional as follows. We define the penalty term

$$H(u) = M\|u_-\|_{L^1(\Omega)}.$$

The following properties of H are immediate.

Lemma 5.7. *The functional $H : L^p(\Omega) \to \overline{\mathbb{R}}$ is convex, lower semi-continuous, bounded above in a neighborhood of zero and $H^*(u^*) = \mathcal{I}_{[-M,0]}(u^*)$.*

Replacing $\mathcal{I}_{[0,\infty)}$ by H obviously is modification of the original problem for which, in general, we cannot expect that the set of solution coincides with the one of the original problem. However, H is chosen in the spirit of exact penalty functions. For the latter it can be shown that for sufficiently large penalty parameter, the solution of the original problem also solves the penalized one. We also note that our goal is to show equivalence of (5.9) to a saddle-point problem, which is then numerically solved with a primal-dual algorithm. For the original problem, the positivity constraint would then result in a projection of the unknown to the positive reals in every iteration. The modified objective results in a soft-shrinkage operation for the negative values at each iteration, i.e., u is unchanged at positive values and replaced by $u(x) = \min\{u(x) + M, 0\}$ for negative values.

We finally get the following result for the modified PET problem:

Proposition 5.8. *Assume that (AP) is satisfied, and that there exists $c > 0$ such that $c|z| \leq j(x, z)$ for every $z \in \mathbb{R}^d$ and for almost every $x \in \Omega$. Then there exists a solution to the primal problem*

(P-PET)
$$\inf_{u\in L^p(\Omega)} J^{**}(u) + \lambda D_{KL}(Ku) + H(u),$$

*to the corresponding predual problem (*pD-PET*), and to the saddle-point problem*

$$\inf_{\substack{p\in W_0^q(\mathrm{div};\Omega)\\ p\in Q}} \sup_{u\in L^p(\Omega)} (\mathrm{div}\,p, u) - \lambda D_{KL}(Ku) - H(u).$$

*The optimal values of all of these problems coincide, and the problems are equivalent in the sense that a pair (p, u) is a solution to the saddle-point problem if and only if u is a solution to (*P-PET*) and p is a solution to (*pD-PET*).*

Proof. We apply the results of Theorem 5.4 with the data term being $u \mapsto (\lambda\tilde{D} \circ \tilde{K})(u)$, where $\tilde{K} : L^p(\Omega) \to L^1(\Sigma) \times L^p(\Omega)$ with $\tilde{K}u = (Ku, u)$, and $\tilde{D} : L^1(\Sigma) \times L^p(\Omega) \to \overline{\mathbb{R}}$, $(v, u) \mapsto \tilde{D}(v, u) = D_{\mathrm{KL}}(v) + \lambda^{-1}H(u)$.

Since both D_{KL} and H are continuous and finite at 0, also $\tilde{D}$ is continuous and finite at 0. It is hence only left to show that the condition (i) of Proposition 5.4 holds true. For this purpose, we first note that $\mathrm{dom}((\lambda\tilde{D})^*) = \mathrm{dom}((\lambda D_{\mathrm{KL}})^*) \times \mathrm{dom}(H^*)$. Hence

$$\tilde{K}^*\Big[\bigcup_{\mu\geq 0} \mu\,\mathrm{dom}((\lambda\tilde{D})^*)\Big] = \tilde{K}^*\Big[\bigcup_{\mu\geq 0} \mu(\mathrm{dom}((\lambda D_{\mathrm{KL}})^*) \times \mathrm{dom}(H^*))\Big]$$

23

$$\supset \tilde{K}^* \big[\bigcup_{\mu \geq 0} \mu(B_\epsilon(0) \times \{0\}) \big]$$

$$= \mathrm{Rg}(K^*).$$

Now, since the constant functions are contained in $\mathrm{Rg}(K^*)$, we get that

$$P_{\ker(\nabla)} \left(\tilde{K}^* \big[\bigcup_{\mu \geq 0} \mu \, \mathrm{dom}((\lambda \tilde{D})^*) \big] \right) = \ker(\nabla).$$

Further it is easy to see that $(0,0) \in \mathrm{dom}((\lambda \tilde{D})^*)$. Hence, all assumptions of Theorem 5.4 hold and the result follows. $\qquad\square$

Remark 5.9. *Summarizing, both for norm and Poison-log likelihood discrepancies, in the situation where the underlying inverse problem is indeed ill-posed, coercivity of the integrand as it appears in the definition of the structural TV prior is necessary for showing existence and for establishing our main duality result. As in this situation the structural TV prior is topologically equivalent to isotropic TV, we note that also classical stability and convergence results for vanishing noise levels can be shown by standard techniques. We refer, for instance, to [36] for the norm-discrepancy case and to [48] for the case of a Poison-log likelihood discrepancy.*

6. Numerical examples

The goal of this section is to exemplary show the numerical realization of two applications of the above general setting. The first one considers a pure denoising problem. This example is included to provide a setting where the inversion of the forward operator, in this case the identity, is well-posed and hence, using the second assumption of Proposition 5.1, the duality result also holds for integrands which vanish on non-trivial sets. The second problem considers the practically more relevant application of MR-guided PET reconstruction and shows how recently proposed regularization approaches [49, 21] can be realized within our framework.

In an abstract setting, the variational problem setting of the previous section can be written as follows:

$$\min_{u \in L^p(\Omega)} J^{**}(u) + \mathcal{G}(u) + \mathcal{H}(u),$$

where $\mathcal{G}$ and $\mathcal{H}$ are smooth and "simple" functions, respectively, that are to be specified for the concrete problem setting, e.g., one chooses $\mathcal{G}(u) = \|u - f\|_2^2$, $\mathcal{H} \equiv 0$ for denoising. We have shown that, under suitable assumptions, this problem is equivalent to the saddle-point problem

$$\min_{\substack{p \in W_0^q(\mathrm{div};\Omega) \\ p \in Q}} \max_{u \in L^p(\Omega)} (\mathrm{div} p, u) - \mathcal{G}(u) - \mathcal{H}(u).$$

The main point in this reformulation is that it allows to solve the variational problem without requiring explicit knowledge of J^{**}. We recall that our main results contains a saddle-point formulation whose numerical realization is directly amenable to first-order primal-dual algorithms; see, e.g., [17, 18]. In fact, since for $p = 2$ the saddle-point problem is defined in a Hilbert space, a corresponding formulation of algorithm can even be shown to be convergent in the infinite dimensional setting [18]. In practice, however, a proper infinite-dimensional version of the algorithm would require the computation of some proximal mappings in $H(\mathrm{div}; \Omega)$ space, which in turn requires the solution of a non-trivial problem (essentially TV denoising) on its own.

For the derivation of an implementable algorithm, we first discretize and change to the Euclidean norm for both underlying spaces. The iteration steps of the algorithm presented in [18] for solving

the saddle-point problem above can then be specified in a general form as follows: Given some triple $(u, p, \bar{p})$, the updates $u_+, p_+, \bar{p}_+$ are computed as

(6.1)
$$\begin{cases} u_+ = \operatorname{prox}_{\sigma, \mathcal{H}} \left(u + \sigma \operatorname{div} \bar{p} + \sigma \nabla \mathcal{G}(u) \right), \\ p_+ = \operatorname{proj}_Q \left(p + \tau \nabla u_+ \right), \\ \bar{p}_+ = 2p_+ - p. \end{cases}$$

Here, $\operatorname{prox}_{\sigma, \mathcal{H}}$ denotes the proximal mapping defined as $\operatorname{prox}_{\sigma, \mathcal{H}}(u) = \arg\min_y \frac{\|u - y\|_2^2}{2\sigma} + \mathcal{H}(y)$, which is well-defined by convexity of $\mathcal{H}$. The term proj_Q is a projection onto the set Q and results from the proximal mapping of $\mathcal{I}_Q$, and $\sigma, \tau > 0$ need to be suitably chosen. This iteration is then repeated with $(u, p, \bar{p}) := (u_+, p_+, \bar{p}_+)$ until some stopping rule is satisfied.

6.1. **Weighted TV with vanishing weight.** For the case of weighted TV denoising one can simply choose $\mathcal{H} \equiv 0$ and $\mathcal{G}(u) := \frac{\lambda}{2}\|u - f\|_{L^2(\Omega)}^2$, with $f \in L^2(\Omega)$ some given data. This results in the saddle-point problem

(6.2)
$$\inf_{\substack{p \in W_0^2(\operatorname{div};\Omega) \\ p \in Q}} \sup_{u \in L^2(\Omega)} (\operatorname{div} p, u) - \frac{\lambda}{2}\|u - f\|_{L^2(\Omega)}^2,$$

where $Q = \{g \in W_0^2(\operatorname{div};\Omega) \cap L^\infty(\Omega, \mathbb{R}^d) : |g(x)| \leq \alpha(x) \text{ for a.e. } x \in \Omega\}$. For a proof of concept, we compute a weighted TV-based denoising example, with vanishing continuous weight function α, i.e., $j(x, z) = \alpha(x)|z|$, with $\alpha \in C(\overline{\Omega})$, $\alpha \geq 0$. This fits into the regime of condition (ii) of Theorem 5.4, i.e., $K = id$, hence $\operatorname{Rg}(K^*)$ is closed and $\ker(K)$ is finite dimensional and the fact that α vanishes implies that the coercivity assumption (i) of the same theorem does not hold. Notice that because of this lack of coercivity we cannot expect the solution u to be in $\operatorname{BV}(\Omega)$. Thus, the characterization $J^{**}(u) = \int_\Omega \alpha\, d|Du|$ is not valid here. This deficiency, however, can be overcome by solving the corresponding saddle-point problem.

After a standard discretization procedure the saddle-point problem (6.2) can be solved straight-forwardly with the algorithm in (6.1), where the involved proximal mappings $\operatorname{prox}_{\sigma, \mathcal{H}}$ and proj_Q can be computed explicitly and pointwise. We use standard forward differences for the discrete gradient with pixel replication at the boundary. The discrete divergence operator is defined by the adjoint relation $\nabla = -\operatorname{div}^T$. We use an accelerated version of the primal-dual algorithm of [17], as described in [18], where for the step-sizes τ and σ, the update rule $\tau_n = 1/(n+1)$, $\sigma_n = 8/\tau$ is employed, with $\tau_0 = 10^{-4}$ and $\sigma_0 = 8 \times 10^4$. Here, n is the iteration number and 8 is an upper bound for the squared norm of the discrete gradient operator, i.e., $\|\nabla\|^2 \leq 8$. The primal and dual variables were initialized as $u_0 = f$ and $p_0 = 0$ respectively. As a stopping rule we used a maximum number of iterations ($n = 2000$), after which no considerable change in the iterates was observed.

Regarding the weight function α, we note that our purpose here is not to determine a procedure for automatically selecting it – for that we refer the reader to [33, 35]. Nevertheless, in order to produce the weight function α needed for the examples in Figure 1 we enable the following procedure: Initially we detect the edges of the noisy image in Figure 1b (Gaussian noise of zero mean and standard deviation $\sigma = 0.5$) by using an edge detection algorithm such as Canny algorithm [15]; compare Figure 1c. For that we used the MATLAB's built-in Canny function using the parameter values $Thresh = [0.275, 0.55]$ and $Sigma = 1.8$.

Then we construct a positive continuous weight function α which takes the value zero exactly on the edges of the image. This can be done by computing the distance function to the edge set, which can be achieved for instance by solving the Eikonal equation [50]; see Figure 1d. The fact that α vanishes exactly at the edge set leads to a piecewise constant reconstruction with less loss of contrast

and better edge preservation when compared to the standard scalar TV-regularization; compare Figures 1e and 1f. See also [30] for a theoretical justification of this experimental observation.
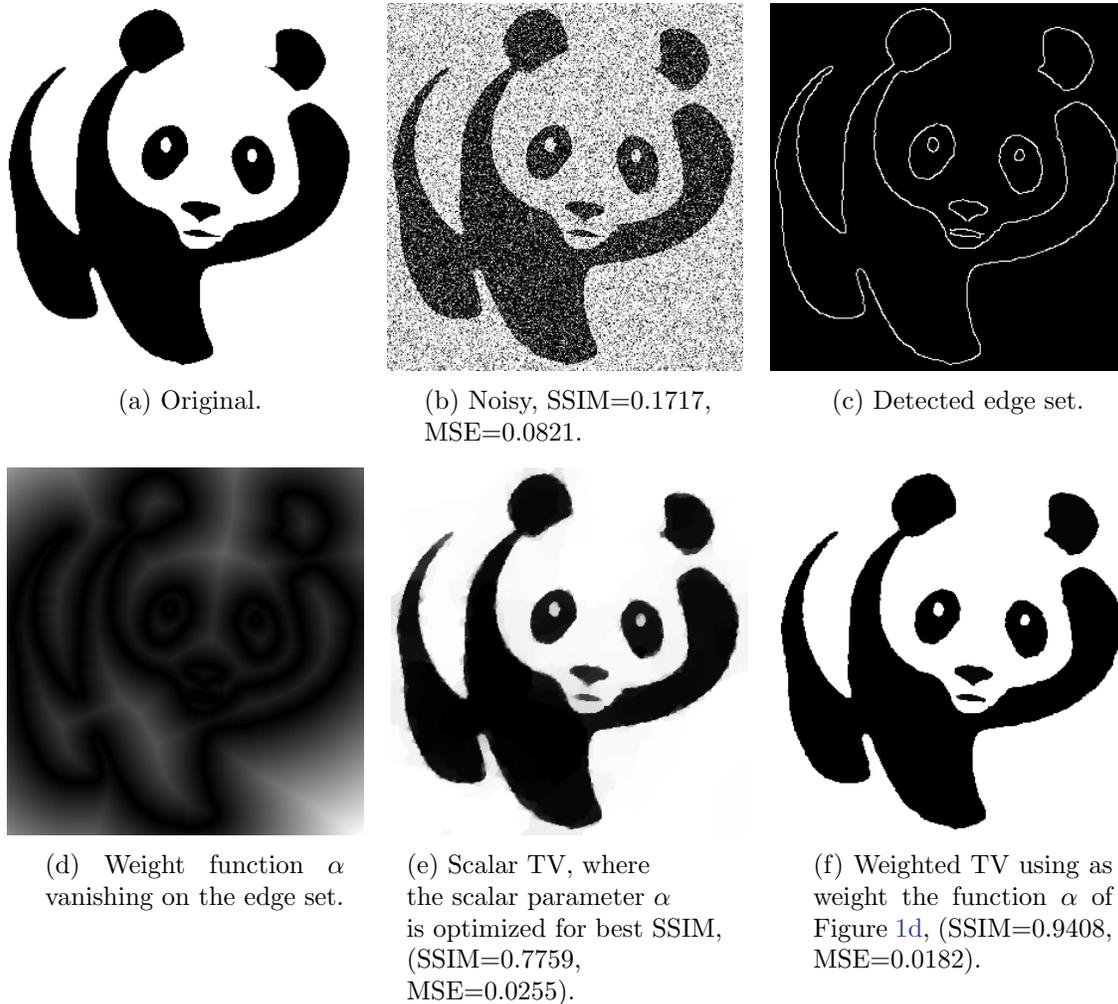

(a) Original.

(b) Noisy, SSIM=0.1717, MSE=0.0821.

(c) Detected edge set.

(d) Weight function α vanishing on the edge set.

(e) Scalar TV, where the scalar parameter α is optimized for best SSIM, (SSIM=0.7759, MSE=0.0255).

(f) Weighted TV using as weight the function α of Figure 1d, (SSIM=0.9408, MSE=0.0182).

FIGURE 1. Weighted TV-based denoising via solution of the saddle-point problem (sp) for $K = id$ and $j(x, z) = \alpha(x)|z|$, with $\alpha \in C(\overline{\Omega})$, $\alpha \geq 0$.
Image source: https://pixabay.com, (licence Creative Commons CC0).

6.2. **MR-PET reconstruction.** Magnetic resonance imaging (MRI) and positron emission tomography (PET) are two complementary imaging techniques that are both intensively used in clinical applications. While MRI allows to resolve soft tissue contrast with comparably high spatial resolution, standard MRI techniques deliver qualitative information only. In contrast, PET imaging suffers from a poor spatial resolution, but it is able to deliver quantitative information (the distribution of a radioactively marked tracer). In the past years, and recently particularly motivated by the availability of joint MR-PET scanners, there has been a substantial research effort towards combining MR and PET data in order to obtain benefits for the ill-posed image reconstruction process; see for instance the references provided in [21]. While some works aim at a simultaneous reconstruction of MR and PET [22, 40, 43], aiming to obtain mutual benefits for both modalities,

most approaches employ an MR-based structural prior to improve upon the poor spatial resolution of PET [21, 49, 52]. Our work here provides an analytical basis for the latter. We note that our theory covers different priors that were proposed for structure-guided PET reconstruction in [21, 49]. In particular, it allows to conclude that well-posedness of the PET reconstruction problem can only be expected if the prior satisfies the coercivity condition stated in Proposition 5.8.

We now exemplary work out the application of our framework to one particular prior, that is motivated by [20, 49]. Assume that we are given an already reconstructed MR image v defined on a joint image domain Ω. We suppppose that $\nabla v \in L^1_{\mathrm{loc}}(\Omega, \mathbb{R}^2)$. We then define the structural TV-based prior for PET reconstruction as follows. For parameters $\eta, \nu > 0$ with $\eta \in (0,1)$ and ν small we define

$$w(x) = \frac{\nabla v(x)}{\sqrt{|\nabla v(x)|^2 + \nu}},$$

the matrix-field

$$A(x) = \sqrt{I - \eta^2 w(x) \otimes w(x)},$$

where $\sqrt{\cdot}$ refers to the matrix square root, and the integrand

$$j(x, z) = |A(x)z|.$$

It is easy to see that j fulfills the assumptions (J1)–(J3). Hence, we can define the structural prior

$$J(u) = \begin{cases} \int_\Omega |A(x)\nabla u(x)| \, dx & \text{if } u \in W^{1,1}(\Omega), \\ +\infty & \text{else,} \end{cases}$$

and consider its lower semi-continuous relaxation for regularization. The effect of this prior can be understood by re-writing

$$j(x, z) = \sqrt{z^T A^T(x) A(x) z} = \sqrt{|z|^2 - \eta^2 (z, w(x))^2} = |z|\sqrt{1 - \eta\left(\frac{z}{|z|}, w(x)\right)^2}.$$

This shows that, at any point $x \in \Omega$, the cost of $j(x, z)$ depends on the extent to which z is aligned with $w(x)$. In particular, the smallest cost will appear if z and $w(x)$ are parallel. Regarding the parameters η and ν, we note that ν is used to carry out a regularized division in the normalization of w, as suggested for instance in [21]. The parameter $\eta \in (0,1)$ does not appear in previous works and we use it to ensure coercivity of j, and hence that J^{**} indeed has a regularizing effect. In fact, noting that

$$j(x, z) = \sqrt{(|z|^2 - (z, w(x))^2)\eta^2 + (1 - \eta^2)|z|^2} \geq \sqrt{1 - \eta^2}|z|,$$

we obtain coercivity of j and hence, by Proposition 5.8, well-posedness of the structural-prior based PET reconstruction problem given as

(6.3) $$\min_{u \in L^p(\Omega)} J^{**}(u) + \lambda D_{\mathrm{KL}}(Ku) + H(u).$$

Further, equivalence and well-posedness of a corresponding saddle-point problem can be concluded in function space and, due to topological equivalence of J^{**} with TV, standard stability results can be expected; see Remark 5.9.

Algorithmic realization. In order to derive a numerical algorithm for solving the structural-prior-based PET reconstruction problem in two dimensions, we now replace the image domain Ω and the data domain Σ by a discretized pixel grid, that is, we set $\Omega = \mathbb{R}^{N_\Omega \times M_\Omega}$ and $\Sigma = \mathbb{R}^{N_\Sigma \times M_\Sigma}$ with $N_\Omega, M_\Omega, N_\Sigma, M_\Sigma \in \mathbb{N}$. For simplicity, we keep the notation of the continuous setting, where now all involved operators, integrals and norms are replaced by their discrete counterparts, using standard discretizations. In particular, for the forward operator for PET, we use the implementation of [41],

which is designed to work with real scanner data and includes resolution modeling by a convolution with a Gaussian kernel.

In order to carry out the iteration steps as described in (6.1) and obtain a convergent algorithm in the discretized setting, we need that the derivative of the data term, on which we perform an explicit descent step, is Lipschitz continuous. To achieve this, we carry out a C^2 instead of a C^1 interpolation of the Poisson-log-likelihood in regions with negative data. In fact, for $f \in [0, \infty)$ and $c_0 \in (0, \infty)$, we define the modified integrand

$$\tilde{l}_{f,c_0}(t) := \begin{cases} t - f \log(t + c_0) & \text{if } t \geq 0, \\ -f \log(c_0) + \left(1 - \frac{f}{c_0}\right)t + \frac{f}{2c_0^2}t^2 & \text{if } t < 0, \end{cases}$$

and a modified data term

$$\widetilde{D_{\mathrm{KL}}}(v) := \int_\Sigma \tilde{l}_{f(\sigma),c_0(\sigma)}(v(\sigma))\, d\sigma + L \int_\Sigma g(v(\sigma))\, d\sigma,$$

where, for a small parameter $\epsilon > 0$, g is a smoothed 1-norm for the negative values given by

$$g(x) = \begin{cases} 0 & \text{if } x \geq 0, \\ -\frac{x^3}{\epsilon^2} - \frac{x^4}{2\epsilon^3} & \text{if } x \in [-\epsilon, 0], \\ |x| + \frac{\epsilon}{2} & \text{if } x \leq -\epsilon. \end{cases}$$

In this context, it is important to remember that any change of D_{KL} that only affects negative values of v does not change the optimal solutions as long as Au is constrained to be positive at every point.

In order to simplify the computation of proj_Q in the iteration (6.1), we further carry out a change of variables in the saddle-point reformulation of (6.3) as $p = A^T q$, where $A^T q$ is defined as pointwise matrix-vector multiplication $A^T q(x) = A^T(x)q(x)$. The resulting saddle-point problem then reads

$$\min_{q: \|q\|_\infty \leq 1} \max_u (\mathrm{div}\, A^T q, u) - \lambda \widetilde{D_{\mathrm{KL}}}(Ku) - H(u),$$

and the iteration steps in (6.1) can be stated in explicit form as

(6.4)
$$\begin{cases} u_+ = \mathrm{prox}_{\sigma,H}\left(u + \sigma \mathrm{div} A^T \overline{q} - \sigma \lambda \nabla \widetilde{D_{\mathrm{KL}}}(u)\right), \\ q_+ = \mathrm{proj}_{\{\|\cdot\| \leq 1\}}\left(q + \tau A \nabla u_+\right), \\ \overline{q}_+ = 2q_+ - q, \end{cases}$$

where $(\mathrm{prox}_{\sigma,H}(u))(x) = \max\{u(x), \min\{u(x) + \sigma M, 0\}\}$ is a soft-thresholding operator and $(\mathrm{proj}_{\{|\cdot|_2 \leq 1\}}(q))(x) = q(x)/\max\{1, |q(x)|_2\}$. Following [18], convergence of the algorithm can be guaranteed under the step-size constraints $\frac{1}{\tau}(\frac{1}{\sigma} - L) \geq \|\mathrm{div}\, K^T\|$. Hence, we analytically estimate both $\|\mathrm{div}\, A^T\|$ and L, with the latter being the Lipschitz constant of the derivative of $\lambda \widetilde{D_{\mathrm{KL}}}$. However, in order to accelerate convergence in practice, we multiply the analytical estimate of L by 10^{-3}, which increases the admissible step-size. Even tough convergence naturally cannot be guaranteed for this increase step-size choice, we did not experience any convergence issue in practice, which might indicate that our analytical estimate of L are too conservative.

To evaluate the effect of structural coupling compared to standard TV-regularization, we simulated a 2D PET and MR scan using a slightly modified version of the brain phantom of [8]; see Figure 2. In fact, we used the software provided in [8] to simulate an MR image with MPRAGE contrast and a FDG-PET image. Both images share similar structures, but they also contain separate features. In addition, a separate lesion was added to each image, in the top right area for the PET image and in the top left area for the MR image, see corresponding red squares, (see [40] for
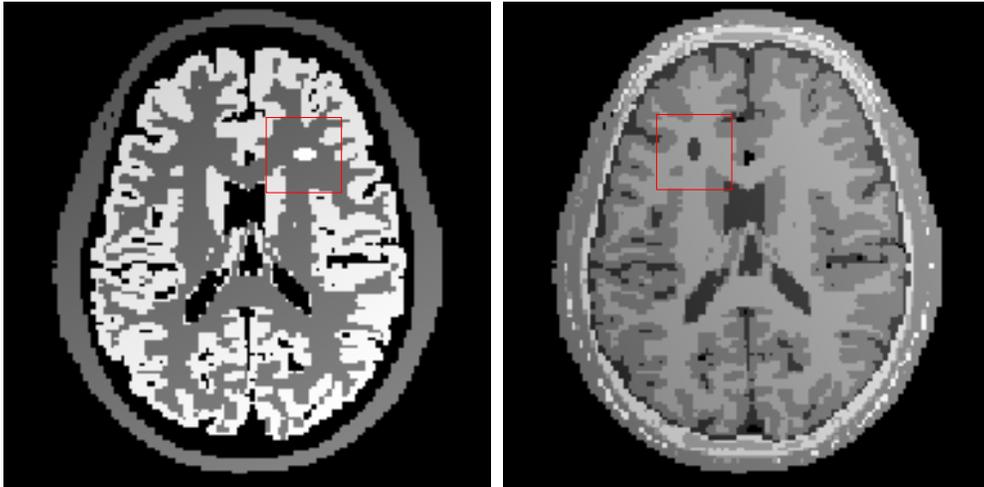
FIGURE 2. Left: Ground truth PET image. Right: MRI MPRAGE contrast image used as structural prior. Note that, in addition to the phantom from [8], we have added two separate lesions (PET: Top right, MRI: Top left) and a linear gray-value gradient in non-background regions (PET: Increasing from top left to bottom right, MRI: Increasing from bottom left to top right).

more details on quantitative values for the PET phantom). Further, to avoid a bias resulting from piecewise constancy, we added two different linear grayscale gradients in the non-background region of both images.

The MR image was then used as structural prior to define the regularization term J as above. PET data, denoted above by f, was generated by forward-projecting the PET phantom, adding simulated random and scatter events (denoted above by c_0) and adding multiplicative Poisson-noise. Two different noise levels (strong and medium) were simulated. For comparison, we also carried out a standard TV-regularized reconstruction. The reconstructions for both methods were obtained by iterating the steps provided in (6.4), where for standard TV-regularization A was replaced by the identity. In order to ensure optimality, we carried out 10^4 iterations for each method, while in practice about 1000 iterations seem sufficient to obtain close-to-optimal results. For both TV- and structural TV-regularization we tested a range of different regularization parameters (λ) and provide here the results with the lowest mean-squared error.

The resulting images can be seen in Figure 3. As one would expect, for both noise levels, the structural-prior based reconstruction is able to recover sharper edges whenever they are aligned with the prior image. This results also in improved MSE values. In the PET only lesion, it can be observed that both TV and structural TV obtain similar results. In the region of the PET image where the MR prior has an additional feature, no particular feature transfer can be observed. This indicates a certain robustness of the approach with respect to non-aligned features between the prior and the reconstructed image. We emphasize again that the results presented here should be understood as a proof-of-concept only, and we refer to [21, 49] for a detailed evaluation of this kind of prior for PET reconstruction, also on 3D PET patient data.

7. CONCLUSION

In this work, we have analyzed a particular class of structural-prior-based regularization approaches for linear inverse problems in function space. While these approaches have recently been
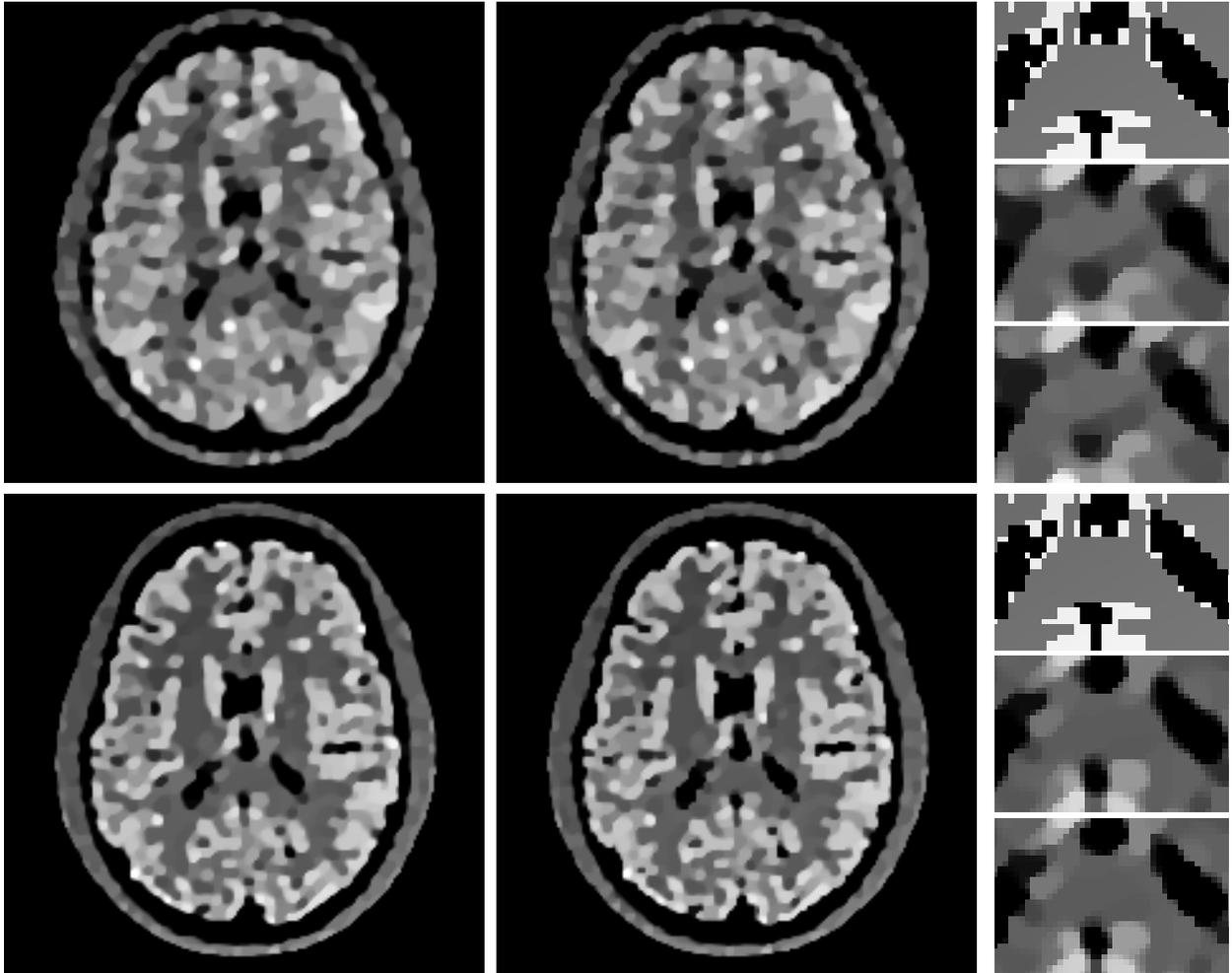
FIGURE 3. Standard TV versus structural TV-reconstruction. Top: Strong noise, Bottom: Medium noise. Left: TV regularized reconstruction, Middle: Structural-TV regularized reconstruction, Right: Close up views: Ground truth, standard TV and structural TV (from top to bottom). Mean-squared errors are as follows. Top: 0.3992 (standard TV), 0.3813 (structural TV). Bottom: 0.3445 (standard TV), 0.3209 (structural TV).

successfully applied to several practically relevant inverse problem settings, an analysis in function space was still missing. Our results show that, in function space, a certain regularity of the prior information in space is necessary in order to obtain an explicit representation of the regularizer. To account for this fact, we have shown how the original minimization problem can alternatively be solved using a saddle-point reformulation that does not require explicit knowledge of the prior. In that context, our main message is that coercivity of the prior is necessary in order to obtain well-posedness (and, consequently, stability by standard results) in non-trivial inverse problem settings. Ultimately, we have shown how to numerically solve the proposed saddle-point reformulation for two proof-of-concept applications including the practically relevant application of structure-guided PET reconstruction.

Appendix A.

Proof of Proposition 2.6. It follows by definition of $W^{1,p}(\Omega)$ and $W_0^q(\mathrm{div};\Omega)$ that $W_0^q(\mathrm{div};\Omega) \subset \ker(\tau)$. In order to prove $\ker(\tau) \subset W_0^q(\mathrm{div};\Omega)$ we show that $C_c^\infty(\Omega, \mathbb{R}^d)$ is dense in $\ker(\tau)$. Based on an idea according to [25, Theorem I.2.6], this is done by showing that any functional $L \in \ker(\tau)^*$ vanishing on $C_c^\infty(\Omega, \mathbb{R}^d)$ is zero on $\ker(\tau)$. So let $L \in \ker(\tau)^*$ with $L(\phi) = 0$ for all $\phi \in C_c^\infty(\Omega, \mathbb{R}^d)$. Since $\ker(\tau)$ is a linear subspace of $W^q(\mathrm{div};\Omega)$, there exists by the Hahn-Banach extension theorem a continuous extension $\overline{L} \in W^q(\mathrm{div};\Omega)^*$. Similarly to the space $H(\mathrm{div};\Omega)$ [25], we can show that there thus exist $l = (l_1, ..., l_{d+1}) \in L^p(\Omega, \mathbb{R}^{d+1})$ such that

$$\overline{L}(g) = \sum_{i=1}^d \int_\Omega l_i g_i \, dx + \int_\Omega l_{d+1} \mathrm{div} g \, dx$$

for $g = (g_1, ..., g_d) \in W^q(\mathrm{div};\Omega)$. Now, for $\phi \in C_c^\infty(\Omega, \mathbb{R}^d)$ we have

$$0 = \overline{L}(\phi) = \sum_{i=1}^d \int_\Omega l_i \phi_i \, dx + \int_\Omega l_{d+1} \mathrm{div}\phi \, dx$$

from which we conclude that $l_{d+1} \in W^{1,q}(\Omega)$ and $\nabla l_{d+1} = (l_1, ..., l_d)$. Thus we can apply the Gauss-Green formula for $W^q(\mathrm{div};\Omega)$ to l_{d+1} and $h \in \ker(\tau)$ arbitrary and get:

$$0 = \int_\Omega (\nabla l_{d+1}, h) \, dx + \int_\Omega l_{d+1} \mathrm{div} h \, dx = L(h)$$

which concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Proposition 4.4. Note first that since $u \in C^1(\Omega)$, a simple integration by parts yields

$$\mathrm{TV}_\alpha^C(u) = \sup\left\{ \int_\Omega \nabla u \cdot \phi \, dx : \phi \in C_c^1(\Omega, \mathbb{R}^d),\ |\phi(x)| \le \alpha(x) \text{ for all } x \in \Omega \right\}.$$

This shows that $\mathrm{TV}_\alpha^C(u) \le \int_\Omega \alpha |\nabla u| \, dx$ and we are left to show $\int_\Omega \alpha |\nabla u| \, dx \le \mathrm{TV}_\alpha^C(u)$. In order to do so, it suffices to prove that for every $\delta > 0$

(A.1)
$$\int_{\Omega_\delta} \alpha |\nabla u| \, dx \le \mathrm{TV}_\alpha^C(u).$$

where $\Omega_\delta := \{x \in \Omega : \mathrm{dist}(x, \partial\Omega) > \delta\}$. Note that since $u \in C^1(\Omega)$, for every $\delta > 0$, $\int_{\Omega_\delta} \alpha |\nabla u| \, dx < +\infty$ and hence $u \in \mathrm{BV}(\Omega_\delta)$ with $Du = \nabla u \mathcal{L}^d \lfloor \Omega_\delta$. Now if

$$V_\alpha(u, \Omega_\delta) = \sup\left\{ \int_{\Omega_\delta} u \, \mathrm{div}\phi \, dx : \phi \in C_c^\infty(\Omega_\delta, \mathbb{R}^d),\ |\phi(x)| \le \alpha(x),\ \text{for all } x \in \Omega_\delta \right\},$$

31

we have that $V_\alpha(u, \Omega_\delta) \leq \mathrm{TV}_\alpha^C(u)$, and in order to establish (A.1), it suffices to show that

$$(\text{A.2}) \qquad \int_{\Omega_\delta} \alpha |\nabla u|\, dx = V_\alpha(u, \Omega_\delta).$$

This follows directly from Proposition 4.2 and the fact that $u \in \mathrm{BV}(\Omega_\delta)$.

$\square$

References

1. R.A. Adams, *Sobolev Spaces*, Academic Press, 1975.
2. M. Amar, V. De Cicco, and N. Fusco, *Lower semicontinuity and relaxation results in BV for integral functionals with BV integrants*, ESAIM: Control, Optimization and Calculus of Variations **14** (2008), no. 3, 456–477, https://doi.org/10.1051/cocv:2007061.
3. L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford University Press, USA, 2000.
4. G. Anzellotti, *On the minima of functionals with linear growth*, Rendiconti del Seminario Matematico della Università di Padova **75** (1986), 91–110, http://www.numdam.org/item?id=RSMUP_1986__75__91_0.
5. P. Athavale, R.L. Jerrard, M. Novaga, and G. Orlandi, *Weighted TV minimization and applications to vortex density models*, Journal of Convex Analysis **24** (2017), no. 4, http://cvgmt.sns.it/media/doc/paper/2790/AJNO.pdf.
6. H. Attouch and H. Brezis, *Duality for the sum of convex functions in general Banach spaces*, Aspects of Mathematics and its Applications **34** (1986), 125–133, https://doi.org/10.1016/S0924-6509(09)70252-1.
7. G. Aubert and P. Kornprobst, *Mathematical problems in image processing: partial differential equations and the calculus of variations*, vol. 147, Springer-Verlag New York Inc, 2006.
8. B. Aubert-Broche, A.C. Evans, and L. Collins, *A new improved version of the realistic digital brain phantom*, Neuroimage **32** (2006), no. 1, 138–145, http://dx.doi.org/10.1016/j.neuroimage.2006.03.052.
9. C. Bathke, T. Kluth, C. Brandt, and P. Maaß, *Improved image reconstruction in magnetic particle imaging using structural a priori information*, International Journal on Magnetic Particle Imaging **3** (2017), no. 1, https://journal.iwmpi.org/index.php/iwmpi/article/view/64.
10. J.M. Borwein and A.S. Lewis, *Convergence of best entropy estimates*, SIAM Journal on Optimization **1** (1991), no. 2, 191–205, https://doi.org/10.1137/0801014.
11. J.M. Borwein and J.D. Vanderwerff, *Convex Functions*, Cambridge University Press, 2010.
12. G. Bouchitté and G. Dal Maso, *Integral representation and relaxation of convex local functionals on $BV(\Omega)$*, Annali della Scuola Normale Superiore di Pisa-Classe di Scienze **20** (1993), no. 4, 483–533, http://www.numdam.org/item?id=ASNSP_1993_4_20_4_483_0.
13. K. Bredies and M. Holler, *A pointwise characterization of the subdifferential of the total variation functional*, arXiv:1609.08918 (2012), https://arxiv.org/abs/1609.08918.
14. H. Brezis, *Functional analysis, Sobolev spaces and partial differential equations*, Springer Science & Business Media, 2010.
15. J. Canny, *A computational approach to edge detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-8** (1986), no. 6, 679–698, http://dx.doi.org/10.1109/TPAMI.1986.4767851.
16. A. Chambolle, M. Goldman, and M. Novaga, *Fine properties of the subdifferential for a class of one-homogeneous functionals*, Advances in Calculus of Variations **8** (2015), no. 1, 31–42, https://doi.org/10.1515/acv-2012-0025.
17. A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision **40** (2011), no. 1, 120–145, http://dx.doi.org/10.1007/s10851-010-0251-1.
18. _____, *On the ergodic convergence rates of a first-order primal–dual algorithm*, Mathematical Programming (2015), 1–35, http://dx.doi.org/10.1007/s10107-015-0957-3.
19. M.J. Ehrhardt and S.R. Arridge, *Vector-valued image processing by parallel level sets*, IEEE Transactions on Image Processing **23** (2014), no. 1, 9–18, http://dx.doi.org/10.1109/TIP.2013.2277775.
20. M.J. Ehrhardt and M.M. Betcke, *Multicontrast MRI reconstruction with structure-guided total variation*, SIAM Journal on Imaging Sciences **9** (2016), no. 3, 1084–1106, https://doi.org/10.1137/15M1047325.
21. M.J. Ehrhardt, P.J. Markiewicz, M. Liljeroth, A. Barnes, V. Kolehmainen, J.S. Duncan, L. Pizarro, D. Atkinson, S. Ourselin, B.F. Hutton, K. Thielemans, and S.R. Arridge, *PET reconstruction with an anatomical MRI prior*

using parallel level sets, IEEE Transaction on Medical Imaging **35** (2016), no. 9, 21892199, `https://doi.org/10.1109/TMI.2016.2549601`.

22. M.J. Ehrhardt, K. Thielemans, L. Pizarro, D. Atkinson, S. Ourselin, B.F. Hutton, and S.R. Arridge, *Joint reconstruction of PET-MRI by exploiting structural similarity*, Inverse Problems **31** (2015), no. 1, 015001, `http://stacks.iop.org/0266-5611/31/i=1/a=015001`.

23. I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, SIAM, 1999.

24. L.C. Evans and R.F. Gariepy, *Measure theory and fine properties of functions*, CRC Press, Boca Raton, FL, 1992.

25. V. Girault and P.-A. Raviart, *Finite Element Method for Navier-Stokes Equation*, Springer, 1986.

26. M. Grasmair and F. Lenzen, *Anisotropic total variation filtering*, Applied Mathematics & Optimization **62** (2010), no. 3, 323–339, `http://dx.doi.org/10.1007/s00245-010-9105-x`.

27. G. Haberfehlner, A. Orthacker, M. Albu, J. Li, and G. Kothleitner, *Nanoscale voxel spectroscopy by simultaneous EELS and EDS tomography*, Nanoscale **6** (2014), no. 23, 14563–14569, `http://dx.doi.org/10.1039/C4NR04553J`.

28. A. Hertle, *Continuity of the Radon transform and its inverse on Euclidean space*, Mathematische Zeitschrift **184** (1983), no. 2, 165–192, `http://eudml.org/doc/173356`.

29. M. Hintermüller and K. Kunisch, *Total bounded variation as bilaterally constrained optimization problem*, SIAM Journal on Applied Mathematics **64** (2004), 1311–1333, `https://doi.org/10.1137/S0036139903422784`.

30. M. Hintermüller, K. Papafitsoros, and C.N. Rautenberg, *Analytical aspects of spatially adapted total variation regularisation*, Journal of Mathematical Analysis and Applications **454** (2017), no. 2, 891 – 935, `https://doi.org/10.1016/j.jmaa.2017.05.025`.

31. _____, *Variable step mollifiers and applications*, in preparation (2017).

32. M. Hintermüller and C.N Rautenberg, *On the density of classes of closed convex sets with pointwise constraints in Sobolev spaces*, Journal of Mathematical Analysis and Applications **426** (2015), no. 1, 585 – 593, `http://dx.doi.org/10.1016/j.jmaa.2015.01.060`.

33. M. Hintermüller and C.N. Rautenberg, *Optimal selection of the regularization function in a generalized total variation model. Part I: Modelling and theory*, Journal of Mathematical Imaging and Vision (2017), `https://doi.org/10.1007/s10851-017-0744-2`.

34. M. Hintermüller, C.N. Rautenberg, and S. Rösel, *Density of convex intersections and applications*, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **473** (2017), no. 2205, `http://dx.doi.org/10.1098/rspa.2016.0919`.

35. M. Hintermüller, C.N. Rautenberg, T. Wu, and A. Langer, *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 (2017), `http://dx.doi.org/10.1007/s10851-017-0736-2`.

36. B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer, *A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*, Inverse Problems **23** (2007), no. 3, 987–1010, `http://stacks.iop.org/0266-5611/23/i=3/a=009`.

37. M. Holler and K. Kunisch, *On infimal convolution of TV-type functionals and applications to video and image reconstruction*, SIAM Journal on Imaging Sciences **7** (2014), no. 4, 2258–2300, `http://dx.doi.org/10.1137/130948793`.

38. K. Jalalzai, *Discontinuities of the minimizers of the weighted or anisotropic total variation for image reconstruction*, arXiv preprint (2014), `http://arxiv.org/abs/1402.0026`.

39. J.P. Kaipio, V. Kolehmainen, M. Vauhkonen, and E. Somersalo, *Inverse problems with structural prior information*, Inverse Problems **15** (1999), no. 3, 713, `http://stacks.iop.org/0266-5611/15/i=3/a=306`.

40. F. Knoll, M. Holler, T. Kösters, R. Otazo, K. Bredies, and D.K. Sodickson, *Joint MR-PET reconstruction using a multi-channel image regularizer*, IEEE Transactions on Medical Imaging **36** (2017), no. 1, 1–16, `http://dx.doi.org/10.1109/TMI.2016.2564989`.

41. T. Kösters, K. Schäfers, and F. Wübbeling, *EMRECON: An expectation maximization based image reconstruction framework for emission tomography data*, IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011, `http://dx.doi.org/10.1109/NSSMIC.2011.6153840`, pp. 4365–4368.

42. D.M. Oberlin and E.M. Stein, *Mapping properties of the Radon transform*, Indiana University Mathematics Journal **31** (1982), no. 5, 641–650, `http://www.jstor.org/stable/24893225`.

43. J. Rasch, E.-M. Brinkmann, and M. Burger, *Joint reconstruction via coupled Bregman iterations with applications to PET-MR imaging*, arXiv:1704.06073 (2017), `https://arxiv.org/abs/1704.06073`.

44. E. Resmerita and R.S. Anderssen, *Joint additive Kullback–Leibler residual minimization and regularization for linear inverse problems*, Mathematical Methods in the Applied Sciences **30** (2007), no. 13, 1527–1544, `https://doi.org/10.1002/mma.855`.

45. D.S. Rigie and P.J La Rivière, *Joint reconstruction of multi-channel, spectral CT data via constrained total nuclear variation minimization.*, Physics in Medicine & Biology **60** (2015), no. 5, 1741–1762 (eng), http://dx.doi.org/10.1088/0031-9155/60/5/1741.

46. R.T. Rockafellar, *Conjugate duality and optimization*, vol. 16, SIAM, 1974.

47. L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D **60** (1992), 259–268, http://dx.doi.org/10.1016/0167-2789(92)90242-F.

48. A. Sawatzky, C. Brune, T. Kösters, F. Wübbeling, and M. Burger, *EM-TV methods for inverse problems with poisson noise*, Level set and PDE based reconstruction methods in imaging, Springer, 2013, http://dx.doi.org/10.1007/978-3-319-01712-9_2, pp. 71–142.

49. G. Schramm, M. Holler, A. Rezaei, K. Vunckx, F. Knoll, K. Bredies, F. Boada, and J. Nuyts, *Evaluation of parallel level sets and Bowsher's method as segmentation-free anatomical priors for time-of-flight PET reconstruction*, Submitted (2017).

50. J.A. Sethian, *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*, vol. 3, Cambridge university press, 1999.

51. K. Steklova and E. Haber, *Joint hydrogeophysical inversion: state estimation for seawater intrusion models in 3D*, Computational Geosciences **21** (2017), no. 1, 75–94, http://dx.doi.org/10.1007/s10596-016-9595-y.

52. K. Vunckx, A. Atre, K. Baete, A. Reilhac, C.M. Deroose, K. Van Laere, and J. Nuyts, *Evaluation of three MRI-based anatomical priors for quantitative PET brain imaging.*, IEEE Transactions Medical Imaging **31** (2012), no. 3, 599–612 (eng), http://dx.doi.org/10.1109/TMI.2011.2173766.