Acta Applicandae Mathematicae (2025) 199:11
https://doi.org/10.1007/s10440-025-00748-4

Structural Similarity in Joint Inverse Problems

Teun Schilperoort[1,2] · **Tristan van Leeuwen**[2,3]

Received: 16 June 2025 / Accepted: 21 September 2025

Abstract
Joint inverse problems occur in many practical situations, where different modalities are used to image the same object. Structural similarity is a way to regularize such joint inverse problems by imposing similarity between the images. While structural similarity has found widespread use in many practical settings, its theoretical foundations remain underexplored. This study develops an over-arching formulation for these types of problems and studies their well-posedness via the Direct Method from the calculus of variations. We focus in particular on lower semi-continuity and coerciveness as essential properties for the well-posedness of the variational problem in $W^{m,p}$ and SBV. Here quasiconvexity and growth properties of the structural similarity quantifier turns out to be essential. We find that the use of gradient-difference, cross-gradient or Schatten norms as structural similarity quantifiers is theoretically justified. A generalized form of the cross-gradient that inherently works on N coupled problems is introduced.

Keywords Inverse problems · Structural similarity · Regularization · Cross-gradient

Mathematics Subject Classification 86A22 · 49K40

1 Introduction

In an inverse problem we are given data, d, and the is aim to find a corresponding image, u, that satisfies the informal equation

$$K(u) = d,$$

with K a given operator. There are many situations where multiple measurements via similar or different data-gathering techniques are performed of the same object. This naturally leads

✉ T. Schilperoort
t.schilperoort@uva.nl

T. van Leeuwen
t.van.leeuwen@cwi.nl

1 Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, Amsterdam, 1098 XH, The Netherlands

2 Mathematical Institute, Utrecht University, Heidelberglaan 8, Utrecht, 3584 CS, The Netherlands

3 Computational Imaging, Centrum Wiskunde & Informatica, Science Park 123, Amsterdam, 1098 XG, The Netherlands

Published online: 03 October 2025

to a number of distinct inverse problems where data $d_1, \ldots, d_N$ are measured from the same object. In the applications envisioned here, $u_i : \Omega \to \mathbb{R}$ represent images defined on a given (Lipschitz) domain $\Omega \subset \mathbb{R}^n$, and the data d_i are related to them via given operators K_i. Naively, one can solve the N inverse problems separately for the best solutions $u_1, \ldots, u_N$, but this approach fails to capture the fact that $u_1, \ldots, u_N$ all describe the same underlying object.

In *joint inversion* [21], we couple the N separate inverse problems by exploiting the fact that $u_1, \ldots, u_N$ share some underlying structure. This idea has found many practical applications in geophysical reconnaissance [1, 15–18, 20, 21, 23, 30, 31, 34–36], medical imaging [4, 5, 7, 12, 13, 24, 25, 37] and image enhancement [9, 22, 27]. There are some theoretical indications for why joint inversion improves on separate inversion, but a complete theory does not exist. In [14], for example, the authors prove that for self-adjoint compact operators, the joint problem is at most as ill-posed as the worst separate problem, and often less ill-posed.

In broad terms there are two approaches to joint inverse problems; Model Fusion (MF) and Structural Similarity (SS). Mathematically, MF [4, 5, 24] (sometimes referred to as Mutual Information [37]) is based on an explicit coupling of the individual images $u_1, \ldots u_N$ and therefore relies heavily on explicit domain knowledge. The focus of this paper is on the latter approach, in which joint structure is captured by devising a regularization functional that promotes joint structures of the images (e.g., shared edges). Examples include inverting MRI anatomical information together with PET-scan functional information [6, 10, 12, 13, 28], multi-channel imaging, where we can regard RGB-valued images as three quantities d_1, d_2, d_3 corresponding to the same image [9, 22], or joint inversion of a low resolution gray-value image along with a high resolution pan-chromatic image [3, 29]

Intuitively, the added information of similar structure, when applicable, should lead to a better solution to the coupled inverse problem compared to the solutions of the individual inverse problems. This has been shown through numerical experiments for many applications (for examples, see [17, 20, 35]). The practical details of how structural similarity of parameter fields are measured can be completely different within disciplines and across applications, for example comparison of gradient fields in an analytic framework, or relative entropy (Kullback-Leibler divergence) in statistics. Moreover, the published results mainly focus on numerical results and do not present general mathematical analysis of the underlying variational problems.

The aim of this paper is to define a general variational framework which encapsulates these different joint formulations of the inverse problem, and determine criteria which leads to the well-posedness of the resulting variational problem. Additionally, we introduce new regularization functionals that take into account the assumed structural similarity present in the problem and investigate the theoretical properties (quasi-convexity, coercivity, and C_p growth conditions) of these and previously used structural regularizers. We show explicitly which regularization functions lead to well-posed problems in $W^{m,p}(\Omega)$ and $SBV(\Omega)$ and give reasons why other previously used regularization functionals might lead to ill-posed problems.

The variational formulation of the joint inverse problem that we study consists of a data and a regularization term and is given by

$$\underset{u \in \mathcal{B}}{\arg\min} \; \|Ku - d\|_{\mathcal{H}}^2 + \alpha J(u), \tag{1}$$

where $u = (u_1, \ldots, u_N)$ denotes a multi-channel image, $d = (d_1, \ldots, d_N)$ denotes the corresponding data, $\mathcal{B} = \mathcal{B}_1 \bigoplus_1 \cdots \bigoplus_1 \mathcal{B}_N$ where $\mathcal{B}_i \in \{W^{m,p}(\Omega), SBV(\Omega)\}$ are Banach, $\mathcal{H}$

Hilbert, $K : \mathcal{B} \to \mathcal{H}$ is a bounded linear operator, $\alpha > 0$ the regularization parameter, and $J : \mathcal{B} \to \mathbb{R}_\infty$ the regularization term. We focus in particular on functionals J of the form

$$J(u) := \int_\Omega f(\nabla u(x))\mathrm{d}x + \varepsilon \sum_{i=1}^{N} \mathrm{TV}(u_i) \tag{2}$$

where $\nabla u(x) = (\nabla u_1(x), \ldots, \nabla u_N(x)) \in \mathbb{R}^{n\times N}$, $\varepsilon \geq 0$ and TV denoting the total variation functional. Many variations of (1) are possible, not limited to considering $c\|\cdot\|_{\mathcal{H}}^s$ for general $c > 0, s > 1$ or adding additional regularization terms in J. We have chosen the squared Hilbert norm as this is the natural generalization of a least-squares data-fit from Euclidean space to general Hilbert spaces and for ease of notation. Additionally, the focus of the new results is the specific form of $J(u)$ incorporating structural similarity and we implicitly assume well-posedness of the data fit. As we only concern about the well-posedness of the inverse problem, the exact form of the data fitting term is largely irrelevant as long as a $\mathcal{H}$-norm is used and the lower semi-continuity of the norm w.r.t. the topology on $\mathcal{H}$ (see Theorem 1). Algorithmic considerations such as the choice of α and optimization methods are of practical importance but fall outside the scope of this work. The integrand is defined by a proper $f : \mathbb{R}^{n\times N} \to \mathbb{R}_\infty^+ = [0, \infty]$. Note that this assumption on f implies properness of J. The exact meaning of the ∇ operator is dependent on the space $\mathcal{B}$, being either a weak derivative for $W^{m,p}(\Omega)$ or the approximate differential for $SBV(\Omega)$. $BV(\Omega)$ would be an additional natural choice as $\mathcal{B}$. However, to get weak-$*$ compactness of $BV(\Omega)$, a singular part would have to be included in J to have any hope of l.s.c. This is due to difficulties with approximations of the Cantor part of $\nabla b, b \in BV(\Omega)$ by Sobolev functions. Assuming the Cantor part to be zero leads to the $SBV(\Omega)$ formulation. The main aim of the paper is to prove the well-posedness of the variational problem given in (1) for various choices of f and the underlying function space. Progressive conditions are established regarding existence, uniqueness, and stability of a minimizer. This is accomplished using a result on the existence of minimizers in variational problems and methods from the multi-variable Direct Method of the Calculus of Variations.

The paper is organized as follows. In Sect. 2 we set the stage for the main results by introducing notation and reviewing some known results from calculus of variations that we need later on. In Sect. 3 we present the regularizers that we treat in this paper. The main results are presented in Sect. 4. Finally, Sect. 5 concludes the paper.

2 Prerequisites

Our joint (complete) spaces $\mathcal{B}, \mathcal{H}$ are constructed from the individual Banach and Hilbert spaces $\mathcal{B}_i, \mathcal{H}_i, i = 1, \ldots N$, with $N \in \mathbb{N}$ as follows.

$$\|u\|_{\mathcal{B}} := \sum_{i=1}^{N} \|u_i\|_{\mathcal{B}_i}, \tag{3}$$

$$\langle h_1, h_2\rangle_{\mathcal{H}} := \sum_{i=1}^{N} \langle h_1^i, h_2^i\rangle_{\mathcal{H}_i}. \tag{4}$$

In this paper we assume similarity of all Banach components and only look at the special cases where they are either a Sobolev space or the space of special functions of bounded variation, i.e., we let $\mathcal{B}_1 = \cdots = \mathcal{B}_N \in \{W^{m,p}(\Omega), SBV(\Omega)\}$. There are no additional assumptions on $\mathcal{H}$. There are natural choices for $\tau_{\mathcal{B}_i}$ as weak topologies and $\tau_{\mathcal{H}_i}$ as norm topologies but other options are available. Whenever a $\mathcal{H}_i$ is infinite-dimensional, one might want to fit against the weak topology on $\mathcal{H}_i$ or a more restricted class of solutions $\mathcal{B}' \subset \mathcal{B}$ is desired, whose properties can be embedded into $\tau_{\mathcal{B}_i}$. Note that this leads to a large class of possible K as there are many linear continuous operators over these spaces (see any of the previously linked papers for examples).

To prove the existence of minimizers of the variational problem given in (1) the standard result below will be applied.

Theorem 1 (Existence of a minimizer [11]) *Let $\mathcal{B}$ a Banach space and $\mathcal{H}$ a Hilbert space associated with the topologies $\tau_{\mathcal{B}}, \tau_{\mathcal{H}}$. Assume the pair $(\mathcal{B}, \tau_{\mathcal{B}})$ has the property that bounded sequences have $\tau_{\mathcal{B}}$-convergent subsequences. Moreover, assume the norm on $\mathcal{H}$ is $\tau_{\mathcal{H}}$-l.s.c. and that the operator $K : \mathcal{B} \to \mathcal{H}$ is linear and sequentially continuous with respect to the topologies $\tau_{\mathcal{B}}$ and $\tau_{\mathcal{H}}$. Let $d \in \mathcal{H}, \alpha > 0$. Furthermore, let $J : \mathcal{B} \to \mathbb{R}^+_\infty$ satisfy the following properties*

- *J is lower-bounded and proper,*
- *J is coercive or (K, J) is (mean) coercive in the sense of Lemma 2.*
- *J is $\tau_{\mathcal{B}}$-l.s.c.,*

If all these conditions are met, then the functional defined in (1) has a minimizer.

Lemma 2 (Mean coercivity [11]) *Let $K : \mathcal{B} \to \mathcal{H}$ linear, $J : \mathcal{B} \to \mathbb{R}^+_\infty$ and $d \in \mathcal{H}$. Let $p_0 \in \mathcal{B}^*$ (the topological dual), and $b_0 \in \mathcal{B}$, such that $\langle p_0, b_0 \rangle = 1$, and $b_0 \notin \mathcal{N}(K)$ (the null space of K), chosen such that J is coercive on*

$$\mathcal{B}_0 := \{b \in \mathcal{B} \mid \langle p_0, b \rangle = 0\},$$

in the sense that for $b \in \mathcal{B}$

$$\|b - \langle p_0, b \rangle b_0\|_{\mathcal{B}} \to \infty \implies J(b) \to \infty.$$

If these choices can be made we say that the pair (K, J) is mean coercive.

We define a set of assumptions defining our optimization setting and which we refer throughout.

Assumption 3 (Minimal Setting) We make the following blanket assumptions

- $\Omega \subset \mathbb{R}^n$ open, bounded and Lipschitz.
- $\mathcal{H} := \oplus_{i=1}^N \mathcal{H}_i$, with $\mathcal{H}_i, i = 1, \ldots N$ Hilbert with $\tau_{\mathcal{H}_i}$ l.s.c. norms.
- $K : \mathcal{B} \to \mathcal{H}$ a (K, J) mean coercive bounded linear operator that is sequentially continuous w.r.t. $\tau_{\mathcal{B}}, \tau_{\mathcal{H}}$.

To apply Theorem 1 we shift the necessary requirements from the joint versions $K, \mathcal{B}, \mathcal{H}$ to properties on the components $K_i, \mathcal{B}_i, \mathcal{H}_i, i = 1, \ldots, N$.

Corollary 4 (Joint construction) *Let $N \in \mathbb{N}$. Take $\mathcal{B}, \mathcal{H}$ constructed via (3) and (4). Let $K_1, \ldots, K_N$ be bounded linear operators that are sequentially continuous with respect to topologies on $(\mathcal{B}_1, \mathcal{H}_1), \ldots, (\mathcal{B}_N, \mathcal{H}_N)$ respectively. Let the norms on $\mathcal{H}_1, \ldots, \mathcal{H}_N$ be*

$\tau_{\mathcal{H}_1}, \ldots, \tau_{\mathcal{H}_N}$ lower semi-continuous. Let $(\mathcal{B}_1, \tau_{\mathcal{B}_1}), \ldots, (\mathcal{B}_N, \tau_{\mathcal{B}_N})$ have the property that bounded sequences have convergent sub-sequences. Let $K : \mathcal{B} \mapsto \mathcal{H}, (u_1, \ldots, u_N) \mapsto (K_1 u_1, \ldots, K_N u_N)$ be an operator that is linear and sequentially continuous w.r.t. $\tau_{\mathcal{B}}, \tau_{\mathcal{H}}$. Then the existence of a minimizer of the Tikhonov-type regularization is only dependent on the properties of J and on whether (K, J) is mean coercive.

Proof Let $\tau_{\mathcal{B}}$ be the product topology of $\tau_{\mathcal{B}_1}, \ldots, \tau_{\mathcal{B}_N}$. We have assumed that $\mathcal{B}_1, \ldots, \mathcal{B}_N$ are compact spaces (w.r.t. the respective topologies), since the finite direct sum of compact spaces is compact we have that $\mathcal{B}$ is compact wrt $\tau_{\mathcal{B}}$. Since l.s.c. works piece-wise also the norm on $\mathcal{H}$ is l.s.c. w.r.t. to the product topology $\tau_{\mathcal{H}}$. By construction, K is linear and sequentially continuous w.r.t. the topologies. All other necessary conditions in Theorem 1 are related to J and the result follows. □

Remark 1 For uniqueness of the minimizer of (1) injectivity of K together with convexity of J or strict convexity of J is sufficient. We also consider the stability of the minimizer, i.e. where continuous perturbations in the data $d \in \mathcal{B}$ lead to continuous changes in the minimizer. The $\tau_{\mathcal{B}}$-l.s.c. of both J and the data term provide continuous dependence as long as $d_j \to d$ strongly implies $d_j \to d$ in $\tau_{\mathcal{H}}$ [11].

In order to apply Theorem 1 we first observe that J as defined in (2) is proper and bounded from below by construction. Furthermore, for $\epsilon > 0$ it is coercive. The remaining ingredient is to establish lower semi-continuity of functionals of the form (2) in $W^{m,p}(\Omega)$ and $SBV(\Omega)$. These are included below in Theorems 6 and 7. Here the notions of a function being Carathéodory, quasiconvex or satisfying certain growth conditions come into play and are defined below.

First, as compactness of the topology $\tau_{\mathcal{B}}$ on $\mathcal{B}$ is necessary in Theorem 1 and $SBV(\Omega)$ with the weak-$*$ topology is not compact, a technical result about the subset of $SBV(\Omega)$ that is compact (w.r.t. to the weak-$*$ topology) is included in Theorem 5. Note that $W^{m,p}(\Omega)$ is compact (w.r.t. to the weak topology) for $p \in (1, \infty)$, $m \in \mathbb{N}$ and Ω Lipschitz, so there are no complications in this case.

To apply Theorem 1 in the case of optimizing over $SBV(\Omega)$ the sizes of the jump sets need to be curtailed in the weak limit for there to be a minimizer.

Theorem 5 (Compactness of $SBV(\Omega)$ [2]) *Let $\Omega \subset \mathbb{R}^n$ be open, bounded with $\partial\Omega$ Lipschitz. Let $\varphi : [0, \infty) \to [0, \infty], \theta : (0, \infty) \to (0, \infty]$ be lower semi-continuous increasing functions with*

$$\lim_{t \to \infty} \frac{\varphi(t)}{t} = \infty, \quad \lim_{t \to 0} \frac{\theta(t)}{t} = \infty.$$

Let $(u_h)_h \subset SBV(\Omega)$, with $\|u_h\|_\infty$ uniformly bounded in h. Additionally,

$$\sup_h \left\{ \int_\Omega \varphi(|\nabla u_h|) dx + \int_{J_{u_h}} \theta(|u_h^+ - u_h^-|) dH^{n-1} \right\} < \infty,$$

where the jump set J_{u_h} is the set where $D^j u_h \neq 0$ and we integrate against the $n-1$-dimensional Hausdorff measure H^{n-1}. Then there exists a subsequence $(u_{h(k)})_k$ weakly$-*$ converging in $BV(\Omega)$ to $u \in SBV(\Omega)$. Additionally, the approximate gradients $\nabla(u_h)_h$ weakly converge to ∇u in $[L^1(\Omega)]^N$ and $(D^j u_h)_h$ weakly$-*$ converge to $D^j u$ in Ω.

So any subset $(u_h)_h \subset SBV(\Omega)$ that is uniformly bounded in ∞-norm with corresponding functions φ, θ that satisfy the conditions above is an admissible component $\mathcal{B}_i, i = 1, \ldots, N$ with the weak-$*$ topology. As φ, θ respectively relate to the two separate parts of the weak gradient D^a, D^j. Theorem 5 can be used solely using one of the functions φ, θ to determine convergence in their respective parts.

Next, we need a few definitions to help us define conditions under which functionals of the form (2) are l.s.c.

Definition 1 (Carathéodory functions [2]) Let $f : S \subseteq \mathbb{R}^N \times \mathbb{R}^{N\times n} \to \mathbb{R}$ be a function. We say that f is Carathéodory if f is $\mathcal{L}^N \times \mathbb{B}(\mathbb{R}^{N\times n})$ measurable and $s \to f(x, s)$ is continuous in $\mathbb{R}^{N\times n}$ for $\mathcal{L}^N$ almost every $x \in S$. Here $\mathcal{L}$ denotes the Lebesgue measure and $\mathcal{B}$ the Borel σ-algebra.

Remark 2 $S = \emptyset$ is also allowed and since the functions considered depend only on $\nabla u = \xi$ the definition above naturally includes functions of type $f : \mathbb{R}^{N\times n} \to \mathbb{R}$.

Definition 2 (Joint convexity [8]) A function $f : \mathbb{R}^{N\times n} \to \mathbb{R}^+_\infty$ is said to be jointly convex if

$$f(\lambda\xi + (1-\lambda)\eta) \leq \lambda f(\xi) + (1-\lambda) f(\eta),$$

for every $\lambda \in [0, 1], \xi, \eta \in \mathbb{R}^{N\times n}$.

Note that this is the usual formulation of convexity with $N = 1$ or $n = 1$, in this case Definitions 2, 3, and 4 are all equivalent.

Definition 3 (Polyconvexity [8]) A function $f : \mathbb{R}^{N\times n} \to \mathbb{R}^+_\infty$ is said to be polyconvex if there exists $F : \mathbb{R}^{\tau(n,N)} \to \mathbb{R}^+_\infty$ jointly convex, such that

$$f(\xi) = F(T(\xi)),$$

where $T : \mathbb{R}^{N\times n} \to \mathbb{R}^{\tau(n,N)}$ is such that

$$T(\xi) := \left(\xi, \operatorname{adj}_2 \xi, \ldots, \operatorname{adj}_{n\wedge N} \xi\right). \tag{5}$$

Here $\operatorname{adj}_s \xi$ stands for the matrix of all $s \times s$ minors of the matrix $\xi \in \mathbb{R}^{N\times n}$ (See Definition 6), $2 \leq s \leq n \wedge N := \min\{n, N\}$ and

$$\tau(n, N) := \sum_{s=1}^{n\wedge N} \sigma(s), \text{ where } \sigma(s) := \binom{N}{s}\binom{n}{s} = \frac{N!n!}{(s!)^2(N-s)!(n-s)!}. \tag{6}$$

Definition 4 (Quasiconvexity [8]) A Borel measurable and locally bounded function $f : \mathbb{R}^{N\times n} \to \mathbb{R}^+_\infty$ is said to be quasiconvex if

$$f(\xi) \leq \frac{1}{\mathcal{L}^n(\Theta)} \int_\Theta f(\xi + \nabla\varphi(x))dx, \tag{7}$$

for every bounded open set $\Theta \subset \mathbb{R}^n$, for every $\xi \in \mathbb{R}^{N\times n}$ and for every $\varphi \in W^{1,\infty}_0(\Theta)$.

Polyconvexity is more general than joint convexity but stricter than quasiconvexity. Polyconvexity is often easier to prove as only a valid F has to be found instead of providing estimates of the "energy" functional in (7). A notable example of a polyconvex function that is not jointly convex are determinants of square matrices. Quasiconvexity turns out to be the right generalization of usual convexity if one wants to preserve correspondence with the l.s.c. of integral functionals(see Theorems 6 and 7).

Definition 5 (Growth conditions [8]) Let $f : \mathbb{R}^{N\times n} \to \mathbb{R}^+_\infty$ and $1 \le p \le \infty$. Then f is said to satisfy growth condition $\left(C_p\right)$ for $1 < p < \infty$ if

$$-\alpha\,(1 + |\xi|^q) \le f(\xi) \le \alpha\,(1 + |\xi|^p) \quad \forall \xi \in \mathbb{R}^{N\times n},$$

for some $\alpha \ge 0$, $1 \le q < p$.

For $p = \infty$, the condition is

$$|f(\xi)| \le \eta(|\xi|) \quad \forall \xi \in \mathbb{R}^{N\times n},$$

where $\eta : [0, \infty) \to \mathbb{R}$ is a continuous and increasing function.

For $p = 1$, the condition is

$$|f(\xi)| \le \alpha(1 + |\xi|) \quad \forall \xi \in \mathbb{R}^{N\times n}$$

where $\alpha \ge 0$.

Theorem 6 (l.s.c. in $W^{1,p}$ [8]) *Let $p \in [1, \infty]$. Let $f : \mathbb{R}^{N\times n} \to \mathbb{R}^+_\infty$ be Carathéodory, quasiconvex and satisfying growth condition $\left(C_p\right)$. Let $\Omega \subset \mathbb{R}^n$ be a bounded open set and*

$$J(u) := \int_\Omega f(\nabla u(x))\mathrm{d}x.$$

Then J is weakly lower semi-continuous in $\left[W^{1,p}(\Omega)\right]^N$ (weak-$*$ lower semi-continuous if $p = \infty$), i.e.

$$\liminf_{u_\nu \to u} J\left(u_\nu\right) \ge J(u).$$

Note the minor conditions on Ω, namely it being bounded and open. There are results that also hold for unbounded Ω but the growth restrictions become more cumbersome in this case [8]. For Sobolev spaces with higher order $m > 1$, we can apply the same result because of the compact embeddings $W^{m,p} \hookrightarrow W^{1,p}$.

The requirements for l.s.c. in $SBV(\Omega)$ are a bit more complicated, with an additional bound on the size of the jump sets.

Theorem 7 (l.s.c. in SBV [2]) *Let $f : \mathbb{R}^{N\times n} \to [0, \infty)$ be Carathéodory with*

$$c|\xi|^p \le f(\xi) \le C(1 + |\xi|^p), \tag{8}$$

for all $\xi \in \mathbb{R}^{N\times n}$ for some $p > 1, c, C > 0$. If f is quasiconvex with respect to ξ then

$$J(u) := \int_\Omega f(\nabla u(x))dx$$

is weakly lower semi-continuous for sequences $u_h \in [SBV(\Omega)]^N$ converging to $u \in [SBV(\Omega)]^N$ with Hausdorff measure $\sup_h H^{N-1}(J_{u_h}) < \infty$.

Remark 3 For functions f only depending on the gradient ξ the upper bound in (8) is equivalent to the upper bound of the C_p growth condition. Whereas when working in $W^{m,p}$, for C_p we can have negative q-growth with $q < p$, for functions in SBV the lower bound in the equation above is much more strict. However, applying Theorem 7 to

$$f_\varepsilon(x,u,\xi) := \varepsilon|\xi|^r + f(x,u,\xi) \tag{9}$$

and then taking the limit $\varepsilon \downarrow 0$ we get the equivalent lower bound

$$0 \leq f(x,u,\xi).$$

So for non-negative functions f only depending on the gradient, we have $C_p \implies$ (8) with the given p.

Lemma 8 shows that including the TV term preserves l.s.c. in $W^{1,p}$ and SBV.

Lemma 8 (TV contribution preserves l.s.c.) *Let $p \in [1,\infty]$, $\Omega \subset \mathbb{R}^n$ be open and bounded. Let $f : \mathbb{R}^{N\times n} \to \mathbb{R}^+_\infty$ be as in Theorem 6 such that*

$$J(u) := \int_\Omega f(\nabla u(x))\mathrm{d}x.$$

is l.s.c. in $W^{1,p}(\Omega;\mathbb{R}^N)$. Then the functional given by

$$J(u) := \int_\Omega f(\nabla u(x))\mathrm{d}x + \varepsilon \sum_{i=1}^N TV(u_i)$$

with $\varepsilon \geq 0$ is also l.s.c in $W^{1,p}(\Omega)$. In addition, the same is true when we replace $W^{1,p}(\Omega;\mathbb{R}^N)$ by $SBV(\Omega)$.

Proof In Appendix A □

Remark 4 There is a straightforward extension for the lower semi-continuity (l.s.c.) results Theorems 6, 7 and Lemma 8 to include f also being dependent on $x, u(x)$, see [8, 33].

3 Structural Regularisation

Now that we have established generic conditions under which functionals of the type (2) are l.s.c. in either $W^{m,p}(\Omega)$ or $SBV(\Omega)$, we will give an overview of specific choices for $f : \mathbb{R}^{N\times n} \to \mathbb{R}^+$. In this paper only the choices for the structural regularizer f that lead to a well-posed problem are included (Although an extended list of regularizers investigated that do not give well-posedness is included in Appendix C). The variable $\xi \in \mathbb{R}^{N\times n}$ is indexed over both components with superscripts ξ^j, $j = 1,\ldots,N$ denoting the column vector and subscripts ξ_i, $i = 1,\ldots,n$ denoting the row vector.

We first define the Total Variation-based regularizers (Total Nuclear Variation, Total Vectorial Variation, and Total Spectral Variation):

$$f_{TNV}(\xi) := \|\xi\|_1, \tag{10}$$

$$f_{VTV}(\xi) := \|\xi\|_2, \tag{11}$$

$$f_{TSV}(\xi) := \|\xi\|_\infty, \tag{12}$$

where

$$\|G\|_{\ell^p} := \left(\sum_{i=1}^{\min\{n,N\}} \sigma_i^p \right)^{1/p} \tag{13}$$

is the Schatten-p norm. These have been used before mainly in image enhancement [9, 19, 22, 32].

Next, we define the Gradient Difference, Cross-Gradient, and Nambu functionals, which are only defined for $N = 2$.

$$f_{GD}(\xi) := \sum_{i=1}^{n} \left|\xi_i^1 - \xi_i^2\right|^2 = \sum_{i=1}^{n} \|\xi_i^1\|^2 + \|\xi_i^2\|^2 - 2\langle \xi_i^1, \xi_i^2 \rangle \tag{14}$$

$$f_{CG}(\xi) := (\xi^1 \times \xi^2)^2 := \|\xi^1\|^2 \|\xi^2\|^2 - \langle \xi^1, \xi^2 \rangle^2, \tag{15}$$

$$f_{\mathrm{Nambu}}(\xi) := |\xi^1 \times \xi^2|. \tag{16}$$

We can readily extend these for $N \geq 2$ by applying them pair-wise:

$$f_{jGD} := \frac{1}{2} \sum_{i=1,\ldots,N} \sum_{j=1,\ldots,i} f_{GD}(\xi^i, \xi^j), \tag{17}$$

$$f_{jCG}(\xi) := \frac{1}{2} \sum_{i=1,\ldots,N} \sum_{j=1,\ldots,i} f_{CG}(\xi^i, \xi^j), \tag{18}$$

$$f_{jNambu}(\xi) := \frac{1}{2} \sum_{i=1,\ldots,N} \sum_{j=1,\ldots,i} f_{Nambu}(\xi^i, \xi^j). \tag{19}$$

Finally, we present the following generalizations of the Cross Gradient and Nambu functionals for $N \geq 2$

$$f_{gCG}(\xi) := \sum_{\beta \in I^N_{\min\{N,n\}}} \sum_{\alpha \in I^n_{\min\{N,n\}}} \left((\mathrm{adj}_{\min\{N,n\}} \xi)^\beta_\alpha \right)^2, \tag{20}$$

$$f_{gNambu}(\xi) := \sqrt{f_{gCG}(\xi)}, \tag{21}$$

where

$$I^a_b := \{(\alpha_1, \ldots, \alpha_b) \in \mathbb{N}^b : 1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_b \leq a\},$$

is the set of all increasing b-tuples up to a and we have the following definition.

Definition 6 (Adjugate matrices) The adjugate matrix of order s, $\mathrm{adj}_s \xi \in \mathbb{R}^{\binom{N}{s} \times \binom{n}{s}}$ is defined to be

$$\mathrm{adj}_s \xi = \begin{pmatrix} (\mathrm{adj}_s \xi)^1_1 & \ldots & (\mathrm{adj}_s \xi)^1_{\binom{n}{s}} \\ \vdots & \ddots & \vdots \\ (\mathrm{adj}_s \xi)^{\binom{N}{s}}_1 & \ldots & (\mathrm{adj}_s \xi)^{\binom{N}{s}}_{\binom{n}{s}} \end{pmatrix}.$$

Where

$$\left(\mathrm{adj}_s\xi\right)^i_\alpha = (-1)^{i+\alpha}\det\begin{pmatrix}\xi^{i_1}_{\alpha_1} & \dots & \xi^{i_1}_{\alpha_s}\\ \vdots & \ddots & \vdots\\ \xi^{i_s}_{\alpha_1} & \dots & \xi^{i_s}_{\alpha_s}\end{pmatrix},$$

and $(i_1,\dots,i_s)$, $(\alpha_1,\dots,\alpha_s)$ are such that $\varphi^N_s(i)=(i_1,\dots,i_s)$, $\varphi^n_s(\alpha)=(\alpha_1,\dots,\alpha_s)$.

The cross-gradient and Nambu functional (also f_{jGD} to a lesser extent) have been widely used before in medical imaging and geosciences [1, 13, 15, 17, 18, 20, 23] but only in the case $N=2$ where only two coupled inverse problems are considered at a time and a sum over all pairs is used as cost function (see definition of f_{jCG}, f_{jNambu}). There have also been investigations for $N=3$ using RGB-images, [9, 26] via the Nambu functional and parallel level sets. The parallel level sets and other widely used structural quantifiers unfortunately do not lead to jointly convex f and hence should be tackled with other methods (See Appendix C). The generalized definitions for f_{gCG} and f_{gNambu} that preserve quasi-convexity are novel to the knowledge of the authors.

The necessary properties for the integrand functions f are proven in the lemma below. Note that these functions can be seen to be Carathéodory from the definitions.

Lemma 9 *Let $N, n\in\mathbb{N}$, then*

1. f_{TNV}, f_{TSV}, f_{VTV} *satisfy growth condition C_1 and are jointly convex.*
2. f_{jGD} *satisfies growth condition C_2 and is jointly convex.*
3. f_{jCG} *satisfies growth condition C_4 and is polyconvex.*
4. f_{jNambu} *satisfies growth condition C_2 and is polyconvex.*
5. f_{gCG} *satisfies growth condition $C_{\min\{N,n\}^2}$ and is polyconvex.*
6. f_{gNambu} *satisfies growth condition $C_{\min\{N,n\}}$ and is polyconvex.*

Note that polyconvexity or joint convexity imply separate convexity for the functions above.

Proof See Appendix B □

4 Main Results

Theorem 10 (Well-posedness in $W^{1,p}(\Omega)$) *Assume 3. Take as Banach space $\mathcal{B}=\bigoplus_{i=1}^N W^{m,p}(\Omega)$, $p\in(1,\infty)$ with the corresponding weak topology on each component. $d\in\mathcal{H}$, $u=(u_1,\dots,u_N)\in\mathcal{B}$, $\alpha>0$, $\epsilon>0$. Let $J:\mathcal{B}\to\mathbb{R}^+_\infty$ be given by*

$$J(u) := \int_\Omega f(\nabla u(x))\mathrm{d}x + \epsilon\sum_{i=1}^N TV(u_i),$$

with $f\in\{f_{VTV}, f_{TNV}, f_{TSV}, f_{jGD}, f_{jCG}, f_{jNambu}, f_{gCG}, f_{gNambu}\}$. Then

$$\underset{u\in\mathcal{B}}{\arg\min}\ \|Ku-d\|^2_{\mathcal{H}} + \alpha J(u)$$

admits a solution.

Secondly, this minimizer is unique if K is injective and $f \in \{f_{TNV}, f_{TSV}, f_{VTV}\}$.

Furthermore, if $\tau_{\mathcal{H}_i}$ is weaker than the norm topology in $\mathcal{H}$ for all $i = 1, \ldots N$ then the problem is stable. In the sense that if $d_j \to d$ in $\mathcal{B}$, then

$$\operatorname*{arg\,min}_{u\in\mathcal{B}} \left(\|Ku - d_j\|_{\mathcal{H}}^2 + \alpha J(u)\right) \to \operatorname*{arg\,min}_{u\in\mathcal{B}} \left(\|Ku - d\|_{\mathcal{H}}^2 + \alpha J(u)\right)$$

in $\tau_{\mathcal{B}}$.

Remark 5 As f_{VTV}, f_{TNV} and f_{TSV} are themselves coercive the result still holds for $\epsilon = 0$. For the other f strict positivity of ϵ is necessary.

Proof The core argument rests on Theorem 1. It is necessary to check all conditions in this statement. Via Corollary 4 we can transport our properties on the components of $\bigoplus_{i=1}^N \mathcal{B}_i$ and $\bigoplus_{i=1}^N \mathcal{H}_i$ to properties of $\mathcal{B}$ and $\mathcal{H}$. The combination of Banach space $W^{m,p}(\Omega)$, $p > 1, m \in \mathbb{N}$ with the weak topology and the assumptions 3 result in the required properties of Theorem 1 regarding $\Omega, \mathcal{H}$, and K.

Note that by construction all f are real-valued and continuous on $\bar{\Omega}$ and since Ω is bounded, J is proper (and non-negative by non-negativity of f).

All $f \in \{f_{VTV}, f_{TNV}, f_{TSV}\}$ satisfy the necessary properties for weak l.s.c. in $W^{1,p}(\Omega)$ in Theorem 6 from Lemma 9 as joint convexity or polyconvexity imply quasiconvexity. Here we can directly decide mean coercivity of (K, J) depending on $\mathcal{N}(K)$ since $J(u)$ is coercive in ∇u.

For $f \in \{f_{jGD}, f_{gCG}, f_{gNambu}\}$, weak l.s.c. in $W^{1,p}(\Omega)$ comes from Lemma 8. Since there is no way to get mean coercivity from the structural part f the mean coercive TV part is required such that $\bar{f} \neq f$. Now we can decide mean coercivity depending on $\mathcal{N}(K)$ since then $J(u)$ is coercive in ∇u.

For uniqueness of the minimizer, we need strict convexity of either $\|Kb - d\|_{\mathcal{H}}^2$ or $J(b)$. Since in all particular choices for f there is no strict convexity of $J(b)$ in either b or ∇b, we only have to look at the first term. Taking into account the remark after Corollary 4 there is uniqueness of the minimizer if K is injective and J is jointly convex, this is true iff $f \in \{f_{TNV}, f_{TSV}, f_{VTV}\}$. In the same remark, if each component topology $\tau_{\mathcal{H}_i}$ is weaker than the norm on $\mathcal{H}_i$, then the topologies on $\mathcal{H}$ have the same. This gives by definition $d_j \to d$ strongly in $\mathcal{H}$ implies $d_j \to d$ in $\tau_{\mathcal{H}}$. Stability in the sense described follows automatically. □

Remark 6 Noteworthy is the fact that if K is injective, we can easily prove mean coercivity of (K, J) in the cases above as then $\mathcal{N}(K) = \{0\}$ and $b_0 = b_\Omega$ as the mean over our domain is a valid choice in Lemma 2. We have chosen here for the more general case, as K is not always injective and it is sometimes worth the effort to prove mean coercivity of (K, J) explicitly.

Theorem 11 (Well-posedness in $SBV(\Omega)$) *Assume 3. Take as Banach space $\mathcal{B}$ a subspace of the Banach space $[SBV(\Omega)]^N$, $N \in \mathbb{N}$ with the corresponding weak−∗ topology. Additionally, assume there is some l.s.c. increasing $\theta : (0, \infty) \to (0, \infty]$ with*

$$\lim_{t\to 0} \frac{\theta(t)}{t} = \infty,$$

where

$$\sup_{b\in\mathcal{B}} \left\{ \int_{J_u} \theta\left(|u^+ - u^-|\right) dH^{n-1} \right\} < \infty. \tag{22}$$

$d \in \mathcal{H}, u = (u_1, \ldots, u_N) \in \mathcal{B}, \alpha > 0, \epsilon > 0$. *Let* $J : \mathcal{B} \to \mathbb{R} \cup \{\infty\}$ *be given by*

$$J(u) := \int_\Omega f(\nabla u(x))\mathrm{d}x + \epsilon \sum_{i=1}^{N} TV(u_i), \varepsilon > 0$$

for $f \in \{f_{jGD}, f_{jCG}, f_{jNambu}, f_{gCG}, f_{gNambu}\}$. *Then*

$$\arg\min_{u \in \mathcal{B}} \|Ku - d\|_{\mathcal{H}}^2 + \alpha J(u)$$

admits a solution.

Proof The proof follows the same line of reasoning as the proof of the second J in Theorem 10 with $f + TV$. The only difference being the additional functional conditions of $\mathcal{B}$ and l.s.c. in $SBV(\Omega)$ instead of $W^{1,p}(\Omega)$. Here taking as integrands $f_{VTV}, f_{TNV}, f_{TSV}$ do not lead to well-posed problems as these have linear growth and subsequently cannot satisfy the lower bound $c|\xi|^p \leq f(\xi)$ with $p > 1$ for some $c > 0$.

Instead of Theorem 10 for l.s.c in $W^{1,p}(\Omega)$ for f we now use Theorem 7 and Lemma 8 to get weak l.s.c. in $[SBV(\Omega)]^N$.

Compactness of $\mathcal{B}$ is necessary with respect to the weak$-*$ topology. This is the case when we satisfy the conditions of Theorem 5. In the assumptions above, a satisfactory θ provides two things; a bound on both the size of the jump set $J_b, b \in \mathcal{B}$ and magnitude of the jumps $|b^+ - b^-|$ in J_b. We still need to prove existence of a suitable φ. The claim is that the growth condition necessary for the l.s.c. in $SBV(\Omega)$ provides us with a φ. By the l.s.c. in $SBV(\Omega)$ there is some $p > 1$ such that

$$0 \leq f(\xi) \leq \alpha(1 + |\xi|^p).$$

Take $\varphi : [0, \infty) \to [0, \infty]$ as $\varphi(x) = \alpha(1 + |x|^p)$. Then this is easily seen to be lower semi-continuous and increasing in x. Also

$$\lim_{t \to \infty} \frac{\alpha(1 + |t|^p)}{t} = \infty,$$

since $\alpha > 0, p > 1$. Finally, let $u_h \in SBV(\Omega)$ be a uniformly bounded sequences with $\|u_h\|_\infty < C$ for all h. Then since Ω is bounded,

$$\begin{aligned}\sup_h \left\{ \int_\Omega \varphi(|\nabla b_h|)dx \right\} &= \sup_h \left\{ \int_\Omega \alpha(1 + |\nabla b_h|^p)dx \right\} \\ &\leq \int_\Omega \alpha(1 + C^p)dx \leq \alpha(1 + C^p)|\Omega| < \infty.\end{aligned}$$

Hence we can dispense of the need to provide adequate φ for compactness of $\mathcal{B}$. For the rest of the argument we lead to reader towards the proof of Theorem 10. □

5 Summary

Theorems 10 and 11 give necessary conditions required for well-posedness of the studied variational problem. To be guaranteed existence (and uniqueness) of a minimizer to inverse problem when using a given structural quantifier $f(\nabla u_1, \ldots, \nabla u_N)$ several properties need to hold. Firstly, it is necessary for f to be Carathéodory, quasiconvex and satisfying a C_p growth condition as in Definition 5. This together with the standard assumptions of non-negativity, properness and (mean) coercivity are enough to apply Theorem 10 for well-posedness in $W^{1,p}(\Omega)$. If additionally, if f has super-linear growth, Theorem 11 is also applicable. Note that for $\varepsilon > 0$ there is automatic (mean) coercivity. Using the methods discussed, it is possible to extend the results to any admissible structural integrand $f(x, u(x), \nabla u)$ given f satisfies the extended criteria for l.s.c. of J in $W^{1,p}$ and SBV for such functions. On the theoretical side, further investigation of extensions of Theorem 1 for non-linear operators K or different Banach spaces $\mathcal{B}$ could lead to interesting results. On the practical side, it is the hope that researchers in mathematically adjacent disciplines can adapt the results to their own specific optimization problems and structural similarity quantifiers.

Appendix A: Proof Lemma 8

Proof Lemma 8

The total variation can also be written as an integral over Ω via

$$\mathrm{TV}(u_i) = \int_\Omega |\nabla u_i(x)| dx,$$

or

$$\mathrm{TV}(u_i) = \int_\Omega |\xi_i| dx,$$

So the conclusion follows if we can apply Theorem 6 to the function $\bar{f} := f + f_c := f + \epsilon \sum_{i=1}^N |\xi_i|$. $\bar{f}$ needs to be Carathéodory, quasiconvex and satisfy C_p.

$W^{1,p}(\Omega) \subset BV(\Omega)$, and as TV quantifies the variation, it follows from definition of bounded variation functions that $\mathrm{TV}(u_i) < \infty$ for all $i = 1, \dots, N$ and the sum $0 \le \epsilon \sum_{i=1}^N TV[u_i] < \infty$. So $f(\xi) \le \bar{f}(\xi) < \infty$ for all $\xi \in \mathbb{R}^{N\times n}$, in particular it is real-valued. From the explicit form of the continuous contribution $\sum_{i=1}^N |\xi_i|$ we can easily see $\bar{f}$ to be $\mathcal{L}^N \times \mathbb{B}(\mathbb{R}^{N\times n})$ measurable and continuous. Together this implies f to be Carathéodory.

The new contribution f_c is affine in ξ, this implies that it is convex and in particular quasiconvex. Now, the sum of quasiconvex functions is quasiconvex via

$$\begin{aligned}\bar{f}(\xi) = f(\xi) + f_c(\xi) &\le \frac{1}{|D|}\int_D f(\xi + \nabla\varphi) dx + \frac{1}{|D|}\int_D f_c(\xi + \nabla\varphi) dx \\ &= \frac{1}{|D|}\int_D \bar{f}(\xi + \nabla\varphi) dx\end{aligned}$$

for all $D \subset \mathbb{R}^n$ bounded and open, $\xi \in \mathbb{R}^{N\times n}$ and $\varphi \in W_0^{1,\infty}(D)$. So $\bar{f}$ is quasiconvex.

We prove that f_c satisfies C_1.

$$0 \le f_c(\xi) = \epsilon \sum_{i=1}^N |\xi_i| \le \epsilon \cdot \sum_{i=1}^N \sqrt{\sum_{j=1}^n (\xi_i^j)^2}.$$

Note that $\sqrt{\cdot}$ is a concave function, then applying Jensen's inequality for concave functions with $\phi(x) = \sqrt{x}$ with $x_i = \sum_{j=1}^{n}(\xi_i^j)^2$ we have

$$\phi\left(\frac{\sum_{i=1}^{N} x_i}{N}\right) \geq \frac{\sum_{i=1}^{N} \phi(x_i)}{N}.$$

Substitution gives

$$\sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{n}(\xi_i^j)^2}{N}} \geq \frac{\sum_{i=1}^{N} |\xi_i|}{N}.$$

Writing this as a norm we get for all $\xi \in \mathbb{R}^{N\times n}$

$$\sqrt{N}|\xi| \geq f_c(\xi).$$

Such that with $\epsilon_{C_1} := \sqrt{N}\epsilon \geq 0$ we have C_1 for f_c.

Let f satisfy C_p, $p \in (1, \infty)$. Then

$$f(\xi) + f_c(\xi) \leq \alpha(1 + |\xi|^p) + \bar{\alpha}|\xi| \leq \max\{\alpha, \bar{\alpha}, 1\}(1 + |\xi|^p + |\xi|),$$

for some $\alpha \geq 0, \bar{\alpha} = \sqrt{N}\epsilon$. For all ξ with $|\xi| > 1$ we have $|\xi|^p > |\xi|$ so

$$\bar{f}(\xi) \leq \max\{\alpha, \bar{\alpha}, 1\}(1 + |\xi|^p).$$

The set of all ξ with $|\xi| \leq 1$ is compact in $\mathbb{R}^{N\times n}$. Since $\bar{f}$ is continuous it is bounded on this set, denote the upper bound by $C > 0 \in \mathbb{R}^+$. Then for all ξ we have

$$\bar{f}(\xi) \leq \max\{C, \max\{\alpha, \bar{\alpha}, 1\}(1 + |\xi|^p)\} \leq a(1 + |\xi|^p)$$

with $a = \max\{C, \alpha, \bar{\alpha}, 1\}$ and $\bar{f}$ satisfies the upper bound of C_p. For the lower bound we notice that $f_c \geq 0$ such that the same q lower bound holds for $\bar{f}$ as for f. Together we have that $\bar{f}$ satisfies C_p.

If f satisfies C_1 we have easily

$$|\bar{f}(\xi)| \leq |f(\xi) + f_c(\xi)| = (\alpha + \bar{\alpha})(1 + |\xi|),$$

such that $\bar{f}$ satisfies C_1. If f satisfies C_∞, then the new $\bar{\eta}(|\xi|) := \eta(|\xi|) + \bar{\alpha}(1 + |\xi|)$ implies C_∞ for $\bar{f}$. We have all necessary properties for Theorem 6. Apply Theorem 6 to $\bar{f}$ and the lemma follows in $W^{1,p}(\Omega)$.

The conditions for l.s.c. of $SBV(\Omega)$ are also Carathéodory, quasiconvexity and a growth condition as described in Theorem 7. The same arguments can be used to conclude the result for $SBV(\Omega)$ instead of $W^{1,p}(\Omega)$ since the growth condition necessary is a special case of C_p as discussed in the remark after Theorem 7. □

Appendix B: Proof Lemma 9

Proof Lemma 9

1. Since f_{GD} is a sum of squares, it is non-negative. The lower bound C_p is then trivially true for any $\alpha \geq 0, q \geq 1$. For the upper p-bound, the highest order terms in $f_{GD}(\xi)$ are of order 2, hence

$$f_{GD}(\xi) \leq \sum_{j=1,2} \sum_{i=1}^{n} \left(\xi_i^j\right)^2 = |\xi|^2.$$

Take $\alpha = 1$, then f_{GD} satisfies C_2.

We can successively unpack f_{jGD} to get an easy convexity condition. As addition of convex functions preserves convexity, f_{jGD} is convex for a given n, N iff f_{GD} is convex for a given n. Again by the addition property of convexity, we can look only at one term in the summation over i in f_{GD}, which is the function $|\xi_i^1 - \xi_i^2|^2$ for a fixed $i = 1, \ldots, n$. Since $x \to x^2$ and $x \to |x|$ are both convex mappings we can consider only $\xi_i^1 - \xi_j^2$. This is a linear mapping in two variables and clearly convex. Following this reasoning in the opposite direction concludes the joint convexity of f_{jGD}.

2. From Lagrange's identity, given by

$$\sum_{i=1}^{k} a_i^2 \sum_{i=1}^{k} b_i^2 - \left(\sum_{i=1}^{k} a_i b_i\right)^2 = \sum_{1 \leq i < j \leq k} (a_i b_j - b_i a_j)^2, \tag{23}$$

for two sets $(a_1, \ldots, a_k), (b_1, \ldots, b_k), k \in \mathbb{N}$ it can be seen that f_{CG} can be written as a sum of squares, and thus is always non-negative. Any lower bound for C_p is satisfied. We have

$$|\xi|^4 = \left(\sum_{i=1}^{n} |\xi_i^1|^2 + |\xi_i^2|^2\right)^2$$

$$= \sum_{i=1}^{n} \left(|\xi_i^1|^2 + |\xi_i^2|^2\right)^2 + \sum_{i,j=1, i \neq j}^{n} \left(|\xi_i^1|^2 + |\xi_i^2|^2\right)\left(|\xi_j^1|^2 + |\xi_j^2|^2\right)$$

$$\geq \sum_{i,j=1}^{n} \|\xi^1\|^2 \|\xi^2\|^2 \geq f_{CG}(\xi).$$

Hence taking into account the upper bound, we have C_p for $p \geq 4$. Since implicitly $N = 2, n > 1$, we get $s = \min\{N, n\} = 2$ and $\tau(n, 2) = \binom{2}{1}\binom{n}{1} + \binom{2}{2}\binom{n}{2} = 2n + \frac{n(n-1)}{2}$. In particular we only need to worry about the minors $\mathrm{adj}_2\xi$ of order 2. We have variables

$$T(\xi) = \left(\xi, (\mathrm{adj}_2\xi)_1^1, \ldots, (\mathrm{adj}_2\xi)_{\binom{n}{2}}^1\right) \in \mathbb{R}^{2n + \frac{n(n-1)}{2}}.$$

From Definition 6 we have in particular $I_2^2 = \{(1, 2)\}$ with $|I_2^2| = 1$ and I_2^n the set of all $\binom{n}{2}$ increasing 2-tuples up to n. Hence the upper indices of our 2×2 minors are fixed and equal to $(1, 2)$. Let $\alpha \in \{1, \ldots, \binom{n}{2}\}$ such that $\varphi_2^n(\alpha) = (\alpha_1, \alpha_2)$ with $(\alpha_1, \alpha_2) \in I_2^n$. Then

$$\left(\mathrm{adj}_2\xi\right)_\alpha^1 = (-1)^{1+\alpha} \det \begin{pmatrix} \xi_{\alpha_1}^1 & \xi_{\alpha_2}^1 \\ \xi_{\alpha_1}^2 & \xi_{\alpha_2}^2 \end{pmatrix},$$

and

$$\left(\left(\mathrm{adj}_2\xi\right)_\alpha^1\right)^2 = \left(\xi_{\alpha_1}^1 \xi_{\alpha_2}^2\right)^2 + \left(\xi_{\alpha_2}^1 \xi_{\alpha_1}^2\right)^2 - 2\xi_{\alpha_1}^1 \xi_{\alpha_2}^1 \xi_{\alpha_1}^2 \xi_{\alpha_2}^2$$

$$= \left(\xi^1_{\alpha_1}\xi^2_{\alpha_2} - \xi^1_{\alpha_2}\xi^2_{\alpha_1}\right)^2.$$

Note that this is the same form as in Lagrange's Identity (23). It can be seen that the terms under the sum in the RHS of (23) are the adjugate minors of the matrix

$$\begin{pmatrix} a_1 & \dots & a_n \\ b_1 & \dots & b_n \end{pmatrix}.$$

Our claim is that we can take $F : \mathbb{R}^{\tau(n,2)} \to \mathbb{R} \cup \{\infty\}$ as $F(T(\xi)) = \sum_{\alpha \in I_2^n} \left(\left(\mathrm{adj}_2\xi\right)^1_\alpha\right)^2$. It is a sum of squares of some of our variables $(\xi, \mathrm{adj}_2\xi)$. Since $x \mapsto x^2$ is convex and a sum of convex function is convex we have that F is a convex function in $T(\xi)$. The only difference between (23) and $F(T(\xi))$ is the summation over $\alpha \in I_s^n$ on the one hand and summation over $1 \leq i < j \leq n$ on the other hand. From the definition of I_2^n we have an equivalent index set as

$$I_2^n = \{(i, j) \in \mathbb{N}^2, 1 \leq i < j \leq n\}.$$

From Definition 3 we have that f_{CG} is polyconvex. Addition of convex functions preserves convexity, so via summation over pairs also f_{jCG} is polyconvex.

3. We can write

$$f_{Nambu} = \sqrt{f_{CG}},$$

and f_{CG} satisfies C_4 from (ii), so after taking roots on the lower and upper bound of f_{CG} f_{Nambu} satisfies $C_{\sqrt{4}} = C_2$ immediately. Plugging in,

$$f_{\text{Nambu}}(\xi) = \sqrt{\sum_{\alpha \in I_2^n} \left(\left(\mathrm{adj}_2\right)^1_\alpha\right)^2}.$$

This can be rephrased in terms of a norm. Define $\|\cdot\|_M$ to be the Euclidean norm on $(\mathrm{adj}_2\xi) \in \mathbb{R}^{\binom{n}{2}}$. Now extend this function via the standard embedding to $\mathbb{R}^{2n} \times \mathbb{R}^{\binom{n}{2}}$. Then

$$f_{\text{Nambu}}(\xi) = \|T(\xi)\|_M, \text{ where } T(\xi) \in \mathbb{R}^{2n} \times \mathbb{R}^{\binom{n}{2}}.$$

We can take $F_{\text{Nambu}} = \|\cdot\|_M$, which is a convex function since it is a (projected) norm. So directly from the definition we have f_{Nambu} polyconvex. Addition of convex functions preserves convexity, so via summation over pairs also f_{jNambu} is polyconvex.

4. The definition of adjugate matrices $(\mathrm{adj}_s\xi)^\beta_\alpha$ gives that it is $(-1)^{\beta+\alpha}$ times the determinant of a $s \times s$ matrix. From a function-based perspective we can see $(\mathrm{adj}_s\xi)^\beta_\alpha$ as a polynomial of variables $\xi^i_j, i = 1, \dots, N, j = 1, \dots n$ of order s as we know how to compute determinants. Now in f_{gCG} we have that each term is a square of a multi-variable polynomials of order $\min\{N, n\}$. Hence f_{gCG} is a multi-variable polynomial of order $\min\{N, n\}^2$. From each variable ξ^i_j having

$$\xi^i_j \leq |\xi|,$$

we get for any arbitrary product of length $\min\{N,n\}^2$ of coefficients of ξ denoted by $x_k, k = 1, \dots, \min\{N,n\}$ where every x_k is equal to some coefficient ξ^i_j the following bound.

$$\prod_{k=1,\dots,\min\{N,n\}^2} x_k \leq |\xi| \cdot \dots \cdot |\xi| = |\xi|^{\min\{N,n\}^2}.$$

So each polynomial is bounded from above by $|\xi|^{\min\{N,n\}^2}$ and adding up all contributions we get that f_{gCG} satisfies $C_{\min\{N,n\}^2}$ with $\alpha = |I^N_{\min N,n}||I^n_{\min N,n}|$.

By construction this is a sum of squares over $(\mathrm{adj}_{\min N,n}\xi)^\beta_\alpha$. By the definition of adjugate matrices these variables above are the coefficients of $\mathrm{adj}_{\min N,n}\xi$ with $\beta = 1, \dots, \binom{N}{\min\{N,n\}}, \alpha = 1, \dots, \binom{n}{\min\{N,n\}}$. By the definition of polyconvexity we have a convex function $F : \mathbb{R}^{\tau(n,N)} \to \mathbb{R}$ in $T(\xi) = (\xi, \dots, \mathrm{adj}_{\min\{N,n\}}\xi)$. Hence f_{gCG} is polyconvex.

5. The proof is along the same lines as (iii). Namely, we can regard the square root of the double sum over I^N_s, I^n_s with $s = \min\{N,n\}$ as a norm over $(\mathrm{adj}_s\xi) \in \mathbb{R}^{\binom{N}{s}\binom{n}{s}}$. Then extend it to the entire space $T(\xi) \in \mathbb{R}^{\tau\{N,n\}}$ as projected norm. Then f_{gNambu} is a norm over this Euclidean space and hence convex. Since the variables are given in terms of the coefficients of the adjugate minors this gives polyconvexity of f_{gNambu}. Polyconvexity implies separate convexity. In addition, the order of f_{gNambu} is given by a square root of the order of f_{gCG}. So via a similar reasoning as in (iii) of each term under the square root, f_{gNambu} satisfies $C_{\min\{N,n\}}$.
6. By definition $|\xi| := \sqrt{\sum_{i,j} |\xi^i_j|^2} = \|\xi\|_{l^2}$. The C_1 condition in this case asks for existence of an $\alpha \geq 0$ such that for all $G \in \mathbb{R}^{N\times n}$ the following holds.

$$|\|G\|_{l^p}| \leq \alpha(1 + |G|).$$

Note that both $|G|$ and $\|G\|_{l^p}$ are matrix norms on a finite dimensional vector space, hence there is equivalence between them. By non-negativity of the norm, we automatically satisfy all lower bounds for C_p. The equivalence of matrix norms gives a constant $c_p \geq 0$ such that

$$\|G\|_{l^p} \leq c_p|G|, \forall G \in \mathbb{R}^{N\times n}.$$

Picking $\alpha = c_p$ implies that condition C_1 holds. $\|\xi\|_{l^p}$ is trivially jointly convex, since any norm on a vector space is convex. □

Appendix C: Other Quantifiers of Similarity

Included in Table 1 is a list of integrands $f : \xi \in \mathbb{R}^{2n} \to \mathbb{R}^+$ that quantify structural regularization that were previously used in the literature and that could not give well-posedness of the variational problem via Theorem 1. The structural quantifiers that give well-posedness are laid out explicitly in Theorems 10 and 11. The critical difference between functions that do and do not give well-posedness is the property of quasiconvexity. For completeness sake and as help for researchers that are interested in other structural regularizers there is included a list below of the integrands studied that are not quasiconvex. This can be done by using the Legendre-Hadamard conditions to disprove rank one convexity (and subsequently quasiconvexity) [8]. Noteworthy are the expressions based on the dot-product as these are often used in geoscience [18].

Table 1 Investigated structural quantifiers

Name	Formula
Matched Difference	$\min_{w\in\mathbb{R}^2} \|w_1\xi^1 - \xi^2\|_2^2 + \|\xi^1 - w_2\xi^2\|_2^2$
Dot Product	$\langle \xi^1, \xi^2\rangle^2$
Adapted Dot Product	$\lvert\langle \xi^1, \xi^2\rangle\rvert$
Normalised Dot Product	$\left\lvert\left\langle \frac{\xi^1}{\lvert\xi^1\rvert}, \frac{\xi^2}{\lvert\xi^2\rvert}\right\rangle\right\rvert^2$
Linear Parallel Level Sets	$\lvert\xi^1\rvert\lvert\xi^2\rvert - \langle \xi^1, \xi^2\rangle$
Quadratic Parallel Level Sets	$\sqrt{1+\lvert\xi^1\rvert^2\lvert\xi^2\rvert^2} - \langle \xi^1, \xi^2\rangle^2$

Acknowledgements We thank Prof. Dr. Carolin Kreisbeck (KU Eichstätt-Ingolstadt) for her helpful suggestions and insightful discussions.

Declarations

Competing Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abubakar, A., Gao, G., Habashy, T.M., Liu, J.: Joint inversion approaches for geophysical electromagnetic and elastic full-waveform data. Inverse Probl. **28** (2012). https://doi.org/10.1088/0266-5611/28/5/055016
2. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems (2000)
3. Ballester, C., Caselles, V., Igual, L., Verdera, J., Rougé, B.: A variational model for p+xs image fusion. Int. J. Comput. Vis. **69**, 43–58 (2006). https://doi.org/10.1007/S11263-006-6852-X
4. Bathke, C., Kluth, T., Brandt, C., Maass, P.: Improved image reconstruction in magnetic particle imaging using structural a priori information. Int. J. Magn. Part. Imag. IJMPI **3**, 10 (2017). https://doi.org/10.18416/ijmpi.2017.1703015
5. Bilgic, B., Goyal, V.K., Adalsteinsson, E.: Multi-contrast reconstruction with Bayesian compressed sensing. Magn. Reson. Med. **66**, 1601–1615 (2011). https://doi.org/10.1002/MRM.22956
6. Bland, J., Mehranian, A., Belzunce, M.A., Ellis, S., Costa-Luis, C., McGinnity, C.J., Hammers, A., Reader, A.J.: Intercomparison of mr-informed pet image reconstruction methods. Med. Phys. (2019). https://doi.org/10.1002/mp.13812
7. Bowsher, J.E., Yuan, H., Hedlund, L.W., Turkington, T.G., Akabani, G., Badea, A., Kurylo, W.C., Wheeler, C.T., Cofer, G.P., Dewhirst, M.W., Johnson, G.A.: Utilizing mri information to estimate f18 - fdg distributions in rat flank tumors. IEEE Nucl. Sci. Symp. Conf. Rec. **4**, 2488–2492 (2004). https://doi.org/10.1109/NSSMIC.2004.1462760
8. Dacorogna, B.: Direct Methods in the Calculus of Variations. Applied Mathematical Sciences, vol. 78 (2007). https://doi.org/10.1007/978-0-387-55249-1
9. Ehrhardt, M.J., Arridge, S.R.: Vector-valued image processing by parallel level sets. IEEE Trans. Image Process. **23**, 9–18 (2014). https://doi.org/10.1109/TIP.2013.2277775

10. Ehrhardt, M.J., Betcke, M.M.: Multi-contrast mri reconstruction with structure-guided total variation. SIAM J. Imaging Sci. **9**, 1084–1106 (2015). https://doi.org/10.1137/15M1047325
11. Ehrhardt, M.J., Lang, L.F.: Inverse Problems (2018)
12. Ehrhardt, M.J., Thielemans, K., Pizarro, L., Atkinson, D., Ourselin, S., Hutton, B.F., Arridge, S.R.: Joint reconstruction of pet-mri by exploiting structural similarity. Inverse Probl. **31**, 23 (2015). https://doi.org/10.1088/0266-5611/31/1/015001
13. Ehrhardt, M.J., Markiewicz, P., Liljeroth, M., Barnes, A., Kolehmainen, V., Duncan, J.S., Pizarro, L., Atkinson, D., Hutton, B.F., Ourselin, S., Thielemans, K., Arridge, S.R.: Pet reconstruction with an anatomical mri prior using parallel level sets. Fac. Eng. Inf. Sci. - Pap: Part A **35**, 2189 (2016). https://doi.org/10.1109/TMI.2016.2549601
14. Ford, J.F., Mead, J.L.: Joint inversion of compact operators. J. Inverse Ill-Posed Probl. **28**(1), 105–118 (2019). https://doi.org/10.1515/jiip-2019-0068
15. Franz, G., Moorkamp, M., Jegen, M., Berndt, C., Rabbel, W.: Comparison of different coupling methods for joint inversion of geophysical data: a case study for the Namibian continental margin. J. Geophys. Res., Solid Earth **126**, 2021–022092 (2021). https://doi.org/10.1029/2021JB022092
16. Gallardo, L.A.: Multiple cross-gradient joint inversion for geospectral imaging. Geophys. Res. Lett. **34** (2007). https://doi.org/10.1029/2007GL030409
17. Gallardo, L.A., Meju, M.A.: Joint two-dimensional dc resistivity and seismic travel time inversion with cross-gradients constraints. J. Geophys. Res., Solid Earth **109**, 3311 (2004). https://doi.org/10.1029/2003JB002716
18. Gallardo, L.A., Meju, M.A.: Structure-coupled multiphysics imaging in geophysical sciences. Rev. Geophys. **49**, 1003 (2011). https://doi.org/10.1029/2010RG000330
19. Goldluecke, B., Strekalovskiy, E., Cremers, D.: The Natural Vectorial Total Variation Which Arises from Geometric Measure Theory (2012). https://doi.org/10.1137/110823766
20. Haber, E., Holtzman-Gazit, M.: Model fusion and joint inversion. Surv. Geophys. **34**, 675–695 (2013). https://doi.org/10.1007/S10712-013-9232-4/TABLES/1
21. Haber, E., Oldenburg, D.: Joint inversion: a structural approach. Inverse Probl. **13**, 63 (1997). https://doi.org/10.1088/0266-5611/13/1/006
22. Holt, K.M.: Total nuclear variation and Jacobian extensions of total variation for vector fields. IEEE Trans. Image Process. **23**, 3975–3989 (2014). https://doi.org/10.1109/TIP.2014.2332397
23. Hu, W., Abubakar, A., Habashy, T.M.: Joint electromagnetic and seismic inversion using structural constraints. Geophysics **74** (2009). https://doi.org/10.1190/1.3246586
24. Kaipio, J.P., Kolehmainen, V., Vauhkonen, M., Somersalo, E.: Inverse problems with structural prior information. Inverse Probl. **15**, 713–729 (1999). https://doi.org/10.1088/0266-5611/15/3/306
25. Kazantsev, D., Lionheart, W.R.B., Withers, P.J., Lee, P.D.: Multimodal image reconstruction using supplementary structural information in total variation regularization. Sens. Imaging **15**, 1–18 (2014). https://doi.org/10.1007/S11220-014-0097-5
26. Kimmel, R., Malladi, R., Sochen, N.: Images as embedding maps and minimal surfaces: movies, color, and volumetric medical images. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 350–355 (1997). https://doi.org/10.1109/CVPR.1997.609348
27. Knoll, F., Dong, Y., Langkammer, C., Hintermüller, M., Stollberger, R.: Total variation denoising with spatially dependent regularization (2010)
28. Knoll, F., Holler, M., Koesters, T., Otazo, R., Bredies, K., Sodickson, D.K.: Joint mr-pet reconstruction using a multi-channel image regularizer. IEEE Trans. Med. Imaging **36**(1), 1–16 (2017). https://doi.org/10.1109/TMI.2016.2564989
29. Li, F., Fang, F., Shen, C., Zhang, G.: A variational approach for pan-sharpening. IEEE Trans. Image Process. **22** (2013). https://doi.org/10.1109/TIP.2013.2258355
30. Linde, N., Binley, A., Tryggvason, A., Pedersen, L.B., Revil, A.: Improved hydrogeophysical characterization using joint inversion of cross-hole electrical resistance and ground-penetrating radar traveltime data. Water Resour. Res. **42** (2006). https://doi.org/10.1029/2006WR005131
31. Linde, N., Tryggvason, A., Peterson, J.E., Hubbard, S.S.: Joint inversion of crosshole radar and seismic traveltimes acquired at the south oyster bacterial transport site. Geophysics **73** (2008). https://doi.org/10.1190/1.2937467
32. Moeller, M., Cremers, D.: Image Denoising—Old and New. Denoising of Photographic Images and Video. Advances in Computer Vision and Pattern Recognition. Springer, Berlin (2018). https://doi.org/10.1007/978-3-319-96029-6_3
33. Rindler, F.: Calculus of Variations (2018). https://doi.org/10.1007/978-3-319-77637-8
34. Saunders, J.H., Herwanger, J.V., Pain, C.C., Worthington, M.H., Oliveira, C.R.E.D.: Constrained resistivity inversion using seismic data. Geophys. J. Int. **160**, 785–796 (2005). https://doi.org/10.1111/j.1365-246X.2005.02566.x

35. Scherders, E.M.L.: Joint Inversion of Acoustic and Electromagnetic Wave Fields for Subsurface Imaging (2022). https://repository.tudelft.nl/islandora/object/uuid
36. Tryggvason, A., Linde, N.: Local earthquake tomography for p- and s-velocity structure using 3d cross-gradients constraints. EAGE (2007) https://doi.org/10.1029/2005GL025485
37. Vunckx, K., Atre, A., Baete, K., Reilhac, A., Deroose, C.M., Laere, K.V., Nuyts, J.: Evaluation of three mri-based anatomical priors for quantitative pet brain imaging. IEEE Trans. Med. Imaging (2012). https://doi.org/10.1109/TMI.2011.2173766