A comparative study of structural similarity and regularization for joint inverse problems governed by PDEs

IOP Publishing Inverse Problems

Inverse Problems **35** (2019) 024003 (27pp) https://doi.org/10.1088/1361-6420/aaf129

A comparative study of structural similarity and regularization for joint inverse problems governed by PDEs

Benjamin Crestel[1], **Georg Stadler**[2] **and Omar Ghattas**[1,3]

[1] Institute for Computational Engineering & Sciences, The University of Texas at Austin, Austin, TX, United States of America
[2] Courant Institute of Mathematical Sciences, New York University, New York, United States of America
[3] Department of Geological Sciences and Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX, United States of America

E-mail: ben.crestel@utexas.edu, stadler@cims.nyu.edu and omar@ices.utexas.edu

Received 1 March 2018, revised 31 August 2018
Accepted for publication 9 November 2018
Published 27 December 2018

Abstract

Joint inversion refers to the simultaneous inference of multiple parameter fields from observations of systems governed by single or multiple forward models. In many cases these parameter fields reflect different attributes of a single medium and are thus spatially correlated or structurally similar. By imposing prior information on their spatial correlations via a joint regularization term, we seek to improve the reconstruction of the parameter fields relative to inversion for each field independently. One of the main challenges is to devise a joint regularization functional that conveys the spatial correlations or structural similarity between the fields while at the same time permitting scalable and efficient solvers for the joint inverse problem. We describe several joint regularizations that are motivated by these goals: a cross-gradient and a normalized cross-gradient structural similarity term, the vectorial total variation, and a joint regularization based on the nuclear norm of the gradients. Based on numerical results from three classes of inverse problems with piecewise-homogeneous parameter fields, we conclude that the vectorial total variation functional is preferable to the other methods considered. Besides resulting in good reconstructions in all experiments, it allows for scalable, efficient solvers for joint inverse problems governed by PDE forward models.

Keywords: joint inversion, multi-physics inverse problem, structural similarity prior, vectorial total variation, cross-gradient, nuclear norm

(Some figures may appear in colour only in the online journal)

1. Introduction

In a joint inverse problem one seeks to reconstruct multiple parameter fields from observational data and forward models that map the parameter fields to the data. In many cases these parameter fields reflect different attributes of a single medium and are thus spatially correlated or structurally similar. By imposing prior information on their spatial correlations via a joint regularization term, we seek to improve the reconstruction of the parameter fields relative to inversion for each field independently.

We formulate the joint inverse problem as an optimization problem with a regularized data misfit objective, governed by a forward model that represents a single or multiple physical phenomena. In the following, we restrict ourselves to forward models that take the form of partial differential equations (PDEs) characterized by two unknown parameter fields, m_1 and m_2, which we seek to reconstruct from observational data $\mathbf{d}$. The parameter-to-observable map $\mathcal{F}(m_1, m_2)$ typically involves solution of the forward PDEs given the parameter fields, followed by application of the observation operator, which restricts the PDE solution to the space of observables. The optimization problem is thus

$$\min_{(m_1,m_2)} \left\{ \frac{1}{2}|\mathcal{F}(m_1, m_2) - \mathbf{d}|^2 + \mathcal{R}(m_1, m_2) \right\}. \tag{1}$$

The role played by $\mathcal{R}$ in (1) is discussed in the next paragraph. Here, we address two specific settings for (1). In the first, the forward model in $\mathcal{F}(m_1, m_2)$ describes a single physical phenomenon. An example of such a joint inverse problem is inversion for the primary and secondary wave speeds in the Earth given measurements of the acceleration at the surface. Obtaining high quality reconstructions for both parameter fields is known to be difficult without incorporating some form of prior knowledge that couples the two fields [1–3]. We refer to formulation (1) as a single physics joint inverse problem.

In the second type of joint inverse problem, we consider observations $\mathbf{d}_1$ and $\mathbf{d}_2$ stemming from two distinct physical phenomena respectively, each depending on a single parameter field. In this case the forward models of the physical phenomena are uncoupled, and coupling occurs only via the inverse problem. The corresponding parameter-to-observable maps are denoted by $\mathcal{F}_1(m_1)$ and $\mathcal{F}_2(m_2)$, resulting in

$$\min_{(m_1,m_2)} \left\{ \frac{1}{2}|\mathcal{F}_1(m_1) - \mathbf{d}_1|^2 + \frac{1}{2}|\mathcal{F}_2(m_2) - \mathbf{d}_2|^2 + \mathcal{R}(m_1, m_2) \right\}. \tag{2}$$

This formulation emerges from the general case above by defining $\mathcal{F}(m_1, m_2) = [\mathcal{F}_1(m_1), \mathcal{F}_2(m_2)]^T$ and $\mathbf{d} = [\mathbf{d}_1, \mathbf{d}_2]^T$. In the context of subsurface exploration, just a few of the different physical phenomena that can be combined in (2) include electromagnetic and seismic waves [4, 5], radar and seismic waves [6], DC resistivity and seismic waves [7], and current resistivity and groundwater flow [8].

The *joint regularization term* $\mathcal{R}(m_1, m_2)$ in (1) and (2) acts to impose regularity on m_1 and m_2 individually to combat ill-posedness, but can also express structural similarity or spatial correlations between the two parameter fields. The remainder of this section introduces several different choices for $\mathcal{R}$. To isolate regularization from structural similarity, we decompose the joint regularization term $\mathcal{R}(m_1, m_2)$ into

$$\mathcal{R}(m_1, m_2) = \gamma_1\mathcal{R}_1(m_1) + \gamma_2\mathcal{R}_2(m_2) + \gamma\hat{\mathcal{R}}(m_1, m_2),$$

with $\gamma, \gamma_1, \gamma_2 > 0$. The terms $\mathcal{R}_1$ and $\mathcal{R}_2$ are regularization terms for each parameter field; here we take them to be total variation (TV) regularizations, since our target media are

piecewise-homogeneous (i.e. blocky). The term $\hat{\mathcal{R}}(m_1, m_2)$ incorporates the structural similarity between m_1 and m_2. We now discuss several choices for $\hat{\mathcal{R}}$. In [7], the authors introduce the *cross-gradient* term

$$\hat{\mathcal{R}}(m_1, m_2) = \frac{1}{2}\int_\Omega |\nabla m_1 \times \nabla m_2|^2 \, \mathrm{d}x,$$

which seeks to align gradients of the two parameter fields at each point in the medium, i.e. level sets that have the same shape. This seems to be the most popular choice in geophysics [4–8], and is discussed in section 2.1. Instead of the gradients of the parameter fields, one can use normalized gradients. This results in the normalized cross-gradient term

$$\hat{\mathcal{R}}_{\mathrm{ncg}}(m_1, m_2) = \frac{1}{2}\int_\Omega \left| \frac{\nabla m_1}{|\nabla m_1|} \times \frac{\nabla m_2}{|\nabla m_2|} \right|^2 \mathrm{d}x.$$

The *normalized cross-gradient* was first used in the context of image registration [9], and is discussed in section 2.2. Alternatively, when an empirical relationship between both parameters is known, one could use it in place of the structural similarity term $\hat{\mathcal{R}}$ [4, 10]; this approach, however, can be problematic in practice as these relationships are typically uncertain (thus introducing bias) and the resulting optimization problems can be difficult to solve [10, 11].

Alternatively, a single joint regularization term can impose regularity on both parameter fields while also expressing a preference for structural similarity. In particular, we consider the *vectorial total variation* (VTV) functional,

$$\mathcal{R}(m_1, m_2) = \gamma \int_\Omega \sqrt{|\nabla m_1|^2 + |\nabla m_2|^2} \, \mathrm{d}x,$$

with $\gamma > 0$. The VTV functional was introduced in the context of multi-channel imaging [12, 13], and later used in PDE-constrained joint inverse problems [10]; it is discussed in section 3. A second term we consider is the *nuclear norm*, which was used in [14, 15] to promote gradient alignment of a vector-valued image. Building on this idea, in section 4 we introduce a nuclear norm-based joint regularization term for PDE-constrained joint inverse problems.

The objective of this article is to construct and assess joint regularization terms that are (1) efficient for inverse problems governed by PDEs with infinite-dimensional parameter fields (and are thus large-scale after discretization) and (2) perform well in reconstructing sharp interfaces in the truth parameter fields. Indeed, targeting large-scale inverse problems entails several unique challenges that limit choices of the joint regularization term. Nonlinear inverse problems such as (1) and (2) must be solved iteratively, which requires gradient- (and Hessian-) based optimization methods to limit the number of optimization iterations, along with adjoint methods to limit the number of PDE model solutions that must be carried out at each iteration. Moreover, the adjoint method efficiently provides only directional second derivatives rather than full Hessians, the construction of which would require as many PDE solves as there are parameters (or observations). For these reasons, unless otherwise specified, we employ an inexact Hessian-free Newton-conjugate gradient method with backtracking line-search [1, 16, 17]. That is, we compute the Newton search direction using the preconditioned conjugate gradient method, with early termination to guarantee a descent direction and to avoid over-solving [18]. The efficient solution of the Newton system depends crucially on the choice of preconditioner; we detail our choices for each joint regularization functional in sections 2–4. An overview of the numerical methods we employ to solve large-scale inverse problems governed by PDEs can be found in appendix A.

Besides practicality and efficiency, our comparison of joint inversion methods focuses on the quality of the reconstructions. Truth parameter fields in geophysical exploration and medical imaging problems often present sharp contrasts within parameter fields. We focus on joint regularization terms that can best preserve sharp edges in the reconstructed images. Motivated by these criteria and a literature review, we identified the four candidates discussed above, namely (1) the cross-gradient and (2) its normalized variant, both paired with individual TV regularizations, (3) the VTV joint regularization, and (4) a nuclear norm-based joint regularization.

1.1. Contributions

The main contributions of this article are as follows: (1) We review three joint regularization terms commonly found in the literature (cross-gradient paired with TV, normalized cross-gradient paired with TV, and VTV joint regularization), and discuss their practical use for large-scale joint inverse problems governed by PDEs. We derive their first and second derivatives, and use them to study properties of the different joint regularization terms. (2) We adapt a nuclear norm joint regularization term to the context of joint inverse problems governed by PDEs. We discuss some of the resulting computational challenges, and propose a solver to address them. (3) We carry out a detailed comparison of all four joint regularization terms over a broad range of applications, and discuss their practical performance to reconstruct parameter fields with sharp interfaces.

1.2. Paper overview

In the next three sections, we introduce the four joint regularization terms. The cross-gradient and normalized cross-gradient are discussed in sections 2.1 and 2.2, the vectorial total variation in section 3, and the nuclear norm joint regularization in section 4. Section 5 summarizes our numerical experiments. In section 5.1, we report on several multiple physics joint inverse problems of the form (2), in which the two parameters fields arise as coefficients in two independent Poisson equations, respectively. We use this example to illustrate some key features of each joint regularization term. In section 5.2, we consider a single physics joint inverse problem of the form (1) for the acoustic wave equation, in which we invert for the bulk modulus and the density. Finally, in section 5.3, we study a multiple physics joint inverse problem with two different forward models, one an elliptic PDE and the other an acoustic wave equation. Section 6 provides concluding remarks.

2. Cross-gradient terms

In this section, we introduce the cross-gradient term and its normalized version. The main idea behind both of these structural similarity terms is to express the preference that the level sets of the inversion parameter fields m_1 and m_2 align. As illustrated in figure 1, alignment of the level sets is equivalent to the alignment of the gradients ∇m_1 and ∇m_2 at each point. By definition of the cross-product of two vectors, the vectors ∇m_1 and ∇m_2 are aligned when $|\nabla m_1 \times \nabla m_2|^2$ vanishes.

2.1. The cross-gradient term

The cross-gradient term $\hat{\mathcal{R}}_{\text{cg}}$, defined as

$$\hat{\mathcal{R}}_{\text{cg}}(m_1, m_2) := \frac{1}{2}\int_\Omega |\nabla m_1 \times \nabla m_2|^2 \, \mathrm{d}x, \tag{3}$$

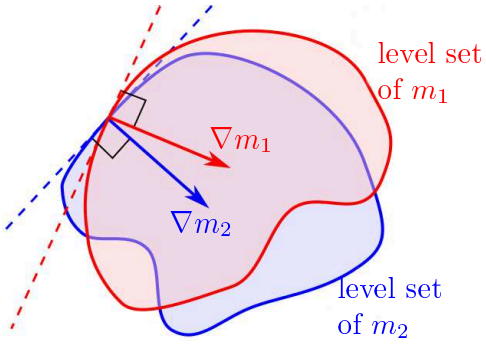


Figure 1. Sketch of a level set of the parameter fields m_1 (red) and m_2 (blue), with their respective gradients at a point.

was introduced in [7] and has become a popular choice in geophysical applications, particularly in seismic imaging. Although the formulation (3) is intuitive, it is inconvenient for discretization and computation of derivatives. Hence, using vector calculus, we re-write (3) as

$$\hat{\mathcal{R}}_{\mathrm{cg}}(m_1, m_2) = \frac{1}{2}\int_\Omega |\nabla m_1|^2 |\nabla m_2|^2 - (\nabla m_1 \cdot \nabla m_2)^2 \,\mathrm{d}x. \tag{4}$$

Combining the cross-gradient term (4) with independent TV regularizations for m_1 and m_2, we obtain the joint regularization

$$\mathcal{R}(m_1, m_2) = \gamma_1 \mathcal{R}_{\mathrm{TV},\varepsilon}(m_1) + \gamma_2 \mathcal{R}_{\mathrm{TV},\varepsilon}(m_2) + \gamma \hat{\mathcal{R}}_{\mathrm{cg}}(m_1, m_2), \tag{5}$$

where and $\gamma, \gamma_1, \gamma_2 > 0$, and here and in the remainder of this paper, we use the notation

$$\mathcal{R}_{\mathrm{TV},\varepsilon}(m) := \int_\Omega \sqrt{|\nabla m|^2 + \varepsilon}\,\mathrm{d}x \text{ for } \varepsilon > 0. \tag{6}$$

In [10] the authors propose a different formulation, in which each independent TV regularization is weighted by a non-linear function of the gradient of the other parameter. The goal of this weighting is to apply TV regularization only for points in the parameter space where the cross-gradient term by itself is not sufficient to prevent oscillatory solutions. Such oscillations may occur where the gradient of one parameter is very small, resulting in an (almost) vanishing cross-gradient term. Because this formulation further increases the nonlinearity of the problem, we instead use (5).

Next, we derive first and second derivatives of the cross-gradient regularization, interpret these derivatives as PDE operators, and draw analogies with the derivatives of the TV functional $\mathcal{R}_{\mathrm{TV}}$ or its regularized version (6). For this purpose, we first derive the first and second variation of the TV functional as follows:

$$\delta_m \mathcal{R}_{\mathrm{TV}}(m; \tilde{m}) = \int_\Omega |\nabla m|^{-1} (\nabla m \cdot \nabla \tilde{m})\,\mathrm{d}x,$$
$$\delta_m^2 \mathcal{R}_{\mathrm{TV}}(m; \hat{m}, \tilde{m}) = \int_\Omega |\nabla m|^{-1} (\nabla \hat{m} \cdot \nabla \tilde{m}) - |\nabla m|^{-3} (\nabla m \cdot \nabla \tilde{m})(\nabla m \cdot \nabla \hat{m})\,\mathrm{d}x,$$

where $\tilde{m}$ and $\hat{m}$ are arbitrary directions. Using integration by parts, the fact that $\tilde{m}$ in the expression for δ_m^2 is arbitrary, and the vector identity $(a \cdot b)(c \cdot d) = b \cdot (ac^T) \cdot d$, one finds that the Hessian $\mathcal{H}$ is the following second-order elliptic PDE operator

$$\mathcal{H}_{\mathrm{TV}}\hat{m} := -\nabla\cdot(A_{\mathrm{TV}}(m)\nabla\hat{m}),$$

with the anisotropic coefficient tensor

$$A_{\mathrm{TV}}(m) = \frac{1}{|\nabla m|}\left(I - \frac{\nabla m \nabla m^T}{|\nabla m|^2}\right). \tag{7}$$

This interpretation as diffusion operator shows that $\mathcal{H}_{\mathrm{TV}}$ acts very differently at different points $x \in \Omega$. In particular, let us consider a point x where the norm of ∇m is large, e.g. x is located at an interface in the parameter field m. Then, in directions parallel to ∇m (i.e. directions normal to an interface), A_{TV} vanishes and thus the elliptic operator does not smooth the reconstruction m in these directions. In contrast, in directions that are orthogonal to ∇m (i.e. directions that are tangent to interfaces), $\mathcal{A}_{\mathrm{TV}}$ does not vanish, thus smoothing the reconstruction m along interfaces. This explains the anisotropic smoothing properties of the TV functional and, in particular, its ability to recover sharp interfaces in parameter fields. Away from interfaces, where ∇m is small, $\mathcal{A}_{\mathrm{TV}}$ behaves like a scaled identity, thus smoothing m in all directions, much as H^1 norm-based Tikhonov regularization does.

We now turn to the derivation of the derivatives of the cross-gradient term $\hat{\mathcal{R}}_{\mathrm{cg}}$. Following similar arguments as for the scalar TV regularization above, this will provide us with insight regarding the regularization properties. Additionally, these derivatives are useful for devising a Newton-type algorithm for the inverse problem solution and for preconditioning the linear systems that arise.

Starting from (4), we now compute the gradient, and the action of the Hessian in a given direction for the cross-gradient term. We perform the computations using weak forms and then use integration by parts to derive the corresponding strong forms. The directional derivative at $m := (m_1, m_2)$ in a direction $\tilde{m} := (\tilde{m}_1, \tilde{m}_2)$ is given by

$$\delta_{m_1}\hat{\mathcal{R}}_{\mathrm{cg}}(m;\tilde{m}_1) = \int_\Omega |\nabla m_2|^2(\nabla\tilde{m}_1\cdot\nabla m_1) - (\nabla m_1\cdot\nabla m_2)(\nabla\tilde{m}_1\cdot\nabla m_2)\,\mathrm{d}x,$$
$$\delta_{m_2}\hat{\mathcal{R}}_{\mathrm{cg}}(m;\tilde{m}_2) = \int_\Omega |\nabla m_1|^2(\nabla\tilde{m}_2\cdot\nabla m_2) - (\nabla m_1\cdot\nabla m_2)(\nabla\tilde{m}_2\cdot\nabla m_1)\,\mathrm{d}x.$$

Taking another variation, we find that the action of the Hessian of the cross-gradient term in a direction $\hat{m} = (\hat{m}_1, \hat{m}_2)$ is given by

$$\delta^2_{m_1}\hat{\mathcal{R}}_{\mathrm{cg}}(m;\hat{m}_1,\tilde{m}_1) = \int_\Omega |\nabla m_2|^2(\nabla\tilde{m}_1\cdot\nabla\hat{m}_1) - (\nabla\tilde{m}_1\cdot\nabla m_2)(\nabla m_2\cdot\nabla\hat{m}_1)\,\mathrm{d}x,$$
$$\delta^2_{m_1,m_2}\hat{\mathcal{R}}_{\mathrm{cg}}(m;\hat{m}_2,\tilde{m}_1) = \int_\Omega 2(\nabla\tilde{m}_1\cdot\nabla m_1)(\nabla m_2\cdot\nabla\hat{m}_2) - (\nabla m_1\cdot\nabla m_2)(\nabla\tilde{m}_1\cdot\nabla\hat{m}_2) - (\nabla\tilde{m}_1\cdot\nabla m_2)(\nabla m_1\cdot\nabla\hat{m}_2)\,\mathrm{d}x,$$
$$\delta^2_{m_2}\hat{\mathcal{R}}_{\mathrm{cg}}(m;\hat{m}_2,\tilde{m}_2) = \int_\Omega |\nabla m_1|^2(\nabla\tilde{m}_2\cdot\nabla\hat{m}_2) - (\nabla\tilde{m}_2\cdot\nabla m_1)(\nabla m_1\cdot\nabla\hat{m}_2)\,\mathrm{d}x.$$

In strong form and neglecting boundary conditions, the Hessian $\mathcal{H}$ acts, in a direction $\hat{m}$, like an anisotropic vector diffusion operator, i.e.

$$\mathcal{H}\hat{m} = -\nabla\cdot(A_{\mathrm{cg}}(m)\nabla\hat{m}),$$

where A_{cg} is a diffusion tensor given by

$$A_{\mathrm{cg}}(m) = \begin{bmatrix} D(m_2) & B(m) \\ B(m)^T & D(m_1) \end{bmatrix}, \tag{8}$$

with, for $i = 1, 2$,

$$D(m_i) := |\nabla m_i|^2 I - \nabla m_i \nabla m_i^T,$$
$$B(m) := 2\nabla m_1 \nabla m_2^T - (\nabla m_1 \cdot \nabla m_2) I - \nabla m_2 \nabla m_1^T.$$

The block-diagonal part of A_{cg} indicates a TV-like behavior but where parameter m_1 (resp. m_2) preserves interfaces in directions where parameter m_2 (resp. m_1) presents an interface; this illustrates the coupling between both parameters. As we show numerically in figure 2, the Hessian of the cross-gradient term can be indefinite. The TV regularization being a convex functional, its Hessian is guaranteed to be positive semidefinite. Therefore, the Hessian obtained by retaining the block diagonal parts of the diffusion tensor (8), i.e. $\mathcal{H}_d \hat{m} := -\nabla \cdot (A_{\mathrm{cg},d}(m) \nabla \hat{m})$, with

$$A_{\mathrm{cg},d}(m) := \begin{bmatrix} D(m_2) & 0 \\ 0 & D(m_1) \end{bmatrix}, \tag{9}$$

is also guaranteed to be positive semidefinite. For this reason, when using the cross-gradient paired with two independent TV regularizations, we precondition the Newton system with a block-diagonal matrix containing the Hessian of the TV regularizations, combined with a small multiple of the identity in each block, and the block-diagonal part of the Hessian of the cross-gradient term (9).

2.2. Normalized cross-gradient

A disadvantage of the cross-gradient term (4) is that it vanishes where one of the inversion parameter fields is constant, hence potentially ignoring sharp discontinuities in the other. A remedy, proposed in the context of image registration in [9], is to normalize the gradient of both inversion parameters in the formulation of the cross-gradient. The normalized cross-gradient is given by

$$\hat{\mathcal{R}}(m_1, m_2) = \frac{1}{2} \int_\Omega \left| \frac{\nabla m_1}{|\nabla m_1|} \times \frac{\nabla m_2}{|\nabla m_2|} \right|^2 \mathrm{d}x = \frac{1}{2} \int_\Omega 1 - \left(\frac{\nabla m_1 \cdot \nabla m_2}{|\nabla m_1||\nabla m_2|} \right)^2 \mathrm{d}x.$$

Since this formulation is non-differentiable where $|\nabla m_1| = 0$ or $|\nabla m_2| = 0$, we use the modified normalized cross-gradient,

$$\hat{\mathcal{R}}_{\mathrm{ncg}}(m_1, m_2) := \frac{1}{2} \int_\Omega 1 - \left(\frac{\nabla m_1 \cdot \nabla m_2}{\sqrt{|\nabla m_1|^2 + \varepsilon} \sqrt{|\nabla m_2|^2 + \varepsilon}} \right)^2 \mathrm{d}x, \tag{10}$$

with $\varepsilon > 0$. In the rest of this paper, we refer to (10) when discussing the normalized cross-gradient. Combining the normalized cross-gradient term (10) with two TV regularizations, we obtain the joint regularization

$$\mathcal{R}(m_1, m_2) = \gamma_1 \mathcal{R}_{\mathrm{TV},\varepsilon}(m_1) + \gamma_2 \mathcal{R}_{\mathrm{TV},\varepsilon}(m_2) + \gamma \hat{\mathcal{R}}_{\mathrm{ncg}}(m_1, m_2), \tag{11}$$

where $\gamma, \gamma_1, \gamma_2 > 0$. Compared to the cross-gradient term, the derivatives of the normalized cross-gradient term give less obvious insight into its regularization behavior. Instead, we illustrate numerically that the normalized cross-gradient often behaves as a concave operator. In

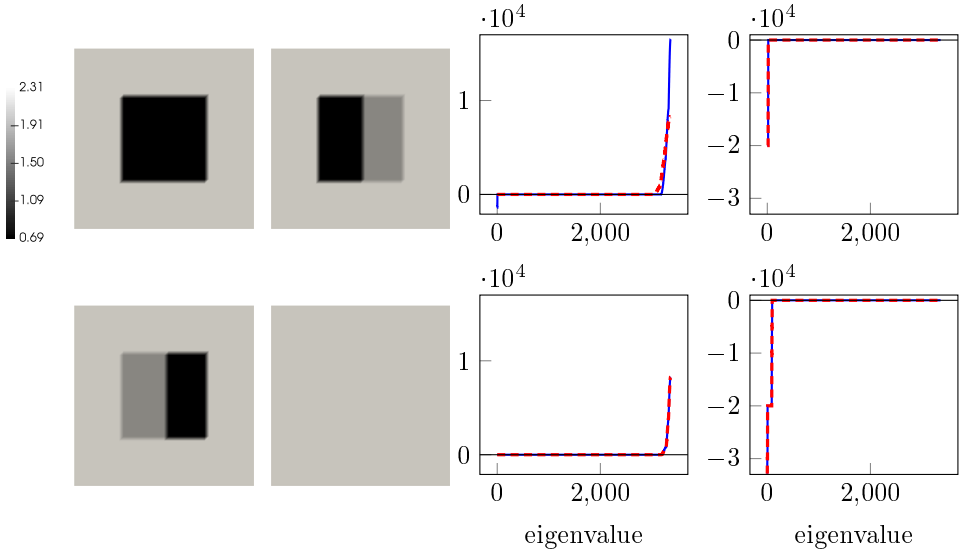


Figure 2. Eigenvalues of the Hessian operator (blue) and block-diagonal part of the Hessian operator (red) for the (iii) cross-gradient term (3) and the (iv) normalized cross-gradient term (10) with $\varepsilon = 10^{-4}$, for two combinations of truth parameter fields (i) m_1 and (ii) m_2. The domain is a unit square discretized by a 40×40 mesh of squares subdivided into triangles, and the parameter fields m_1 and m_2 are discretized using continuous piecewise linear finite elements.

figure 2, we plot the eigenvalues of its Hessian and of the block-diagonal part of its Hessian for different parameter fields m_1 and m_2, and observe that most eigenvalues are negative. The main practical consequence of this observation is that the Hessian of the joint regularization (11) may be indefinite. For this reason, the preconditioner for the Newton system is formed by the Hessians of the TV regularizations alone.

3. Vectorial total variation

The vectorial total variation functional [13], or color TV [12], is the multi-parameter equivalent of the total variation functional. It was first introduced for multi-channel imaging applications [12, 13], and later applied to joint inverse problems [10]. The VTV functional is convex, and unlike the cross-gradient and normalized cross-gradient, it serves as a regularization by itself, i.e. it does not require additional regularization terms. It is given by

$$\mathcal{R}(m_1, m_2) = \gamma \int_\Omega \sqrt{|\nabla m_1|^2 + |\nabla m_2|^2}\, \mathrm{d}x, \tag{12}$$

with $\gamma > 0$. Since this formulation is non-differentiable where $|\nabla m_1| = |\nabla m_2| = 0$, we introduce a modified VTV regularization given by

$$\mathcal{R}_{\mathrm{VTV}}(m_1, m_2) := \gamma \int_\Omega \sqrt{|\nabla m_1|^2 + |\nabla m_2|^2 + \varepsilon}\, \mathrm{d}x, \tag{13}$$

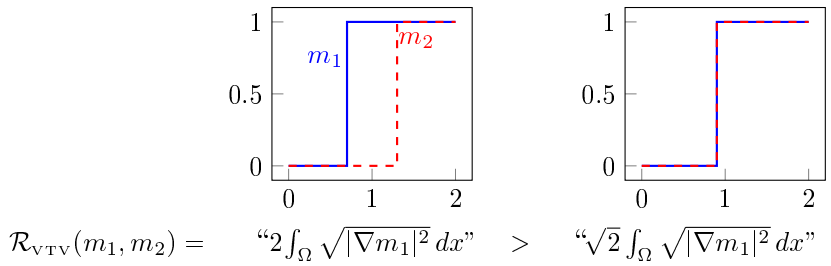


Figure 3. Values of the VTV regularization (12), for two parameter fields m_1 and m_2 defined over $\Omega = [0, 2]$, with both parameter fields having a single jump of the same amplitude, and $\mathcal{R}_{\mathrm{TV}}(m_1) = \mathcal{R}_{\mathrm{TV}}(m_2)$. This informal argument can be made rigorous by using piecewise linear functions for m_1 and m_2.

with $\varepsilon, \gamma > 0$. Whereas the cross-gradient terms (see section 2) work by aligning the level sets of the inversion parameter fields, VTV favors superimposition of discontinuities. An intuitive way to explain this, given the understanding of the TV regularization [19], is sketched in figure 3. Given two parameter fields with a single jump of same amplitude, the VTV functional is minimum when both jumps occur at the same location.

The derivatives of the VTV regularization resemble those of the TV regularization. For simplicity, we set $\gamma \equiv 1$ in the rest of this section. The directional derivative at a point $m = (m_1, m_2)$ in a direction $\tilde{m} = (\tilde{m}_1, \tilde{m}_2)$ is given by

$$\delta_{m_i}\mathcal{R}_{\mathrm{VTV}}(m; \tilde{m}_i) = \int_\Omega \frac{\nabla m_i \cdot \nabla \tilde{m}_i}{\sqrt{|\nabla m_1|^2 + |\nabla m_2|^2 + \varepsilon}}\, \mathrm{d}x, \quad \text{for } i = 1, 2. \tag{14}$$

We again interpret the Hessian of the VTV as a diffusion tensor to study its anisotropic diffusion behavior. In strong form (see section 2.1), it is given by

$$A_{\mathrm{VTV}}(m) := \frac{1}{|\nabla m|_\varepsilon} \begin{bmatrix} I - \frac{\nabla m_1 \nabla m_1^T}{|\nabla m|_\varepsilon^2} & -\frac{\nabla m_1 \nabla m_2^T}{|\nabla m|_\varepsilon^2} \\ -\frac{\nabla m_2 \nabla m_1^T}{|\nabla m|_\varepsilon^2} & I - \frac{\nabla m_2 \nabla m_2^T}{|\nabla m|_\varepsilon^2} \end{bmatrix}, \tag{15}$$

where $|\nabla m|_\varepsilon^2 = |\nabla m_1|^2 + |\nabla m_2|^2 + \varepsilon$. Comparing with the diffusion tensor for the Hessian of the TV regularization (7), we find similar terms along the block diagonal, with the exception of the normalization factor in the denominator. It is $|\nabla m_i|$ in the case of TV, and $|\nabla m|_\varepsilon$ in the case of VTV, i.e. it involves the gradient of both parameters, hence introducing coupling between the parameter fields. The eigen-decomposition of the diffusion tensor of the Hessian provides further insights. For simplicity, we use $\varepsilon = 0$ in this analysis. Skipping details that can be found in [20], the eigenpairs for the diffusion tensor A_{VTV} are

$$\left(\begin{bmatrix} \nabla m_1 \\ \nabla m_2 \end{bmatrix}, 0\right), \left(\begin{bmatrix} (\nabla m_1)^\perp \\ 0 \end{bmatrix}, \frac{1}{|\nabla \mathbf{m}|}\right), \left(\begin{bmatrix} 0 \\ (\nabla m_2)^\perp \end{bmatrix}, \frac{1}{|\nabla \mathbf{m}|}\right), \left(\begin{bmatrix} \nabla m_2 \\ -\nabla m_1 \end{bmatrix}, \frac{1}{|\nabla \mathbf{m}|}\right).$$

The kernel of the diffusion tensor contains parameter field directions that are not smoothed out by the regularization. Reconstructions in these directions can display sharp edges. It is informative to compare the eigenpairs of the diffusion tensor arising from the VTV Hessian with those arising from $\mathcal{R}_{\mathrm{TV}}(m_1) + \mathcal{R}_{\mathrm{TV}}(m_2)$, the sum of two independent TV regularizations. In this case, the eigenpairs are

$$\left(\begin{bmatrix}\nabla m_1\\ 0\end{bmatrix},0\right),\left(\begin{bmatrix}0\\ \nabla m_2\end{bmatrix},0\right),\left(\begin{bmatrix}(\nabla m_1)^{\perp}\\ 0\end{bmatrix},\frac{1}{|\nabla m_1|}\right),\left(\begin{bmatrix}0\\ (\nabla m_2)^{\perp}\end{bmatrix},\frac{1}{|\nabla m_2|}\right).$$

The sum of independent TV regularizations acts in the direction of each parameter m_i independently from the other parameters, analogously to the TV functional for a single inverse problem. That is, it preserves sharp interfaces in the parameter fields (large values of $|\nabla m_i|$) but smoothes along interfaces. This is in contrast with the kernel of the diffusion tensor of VTV, which favors parameter fields with sharp variations occurring at the same physical locations.

The use of TV regularization in PDE-constrained inverse problems increases the nonlinearity of the problem, and requires the use of customized solvers. Due to the similarity between TV and VTV, a similar challenging numerical behaviour can be expected for VTV. In [20] we tailor a primal-dual Newton method [21] for the efficient, scalable solution of PDE-constrained joint inverse problems regularized with VTV. Since the focus of the current paper is on a qualitative comparison of several joint regularization terms, we skip details of this solver here and instead refer to [20].

4. Nuclear norm joint regularization

The nuclear norm joint regularization seeks to promote gradient alignment by minimizing the rank of the Jacobian of the gradients of the parameter fields. Different versions of that idea have been used in various imaging applications. In color image denoising, this approach is often referred to as total nuclear variation [14]; the unified framework to discuss VTV and the total nuclear variation in [14] shows that the nuclear norm-based functional is a regularizer in itself. This can be simply justified by the equivalence of all norms in finite dimensions (here, on the space of matrices). In [15], the authors propose the pointwise nuclear norm of a matrix field as regularization to express a preference for alignment of image edges. Building on [14, 15], we propose a nuclear norm joint regularization suitable for large-scale PDE constrained optimization.

As for the methods discussed in section 2, this term seeks to promote alignment of parameter level sets by attaining its minimum value when gradients align. Let us introduce the matrix-valued function $\mathbf{G}:\Omega\to\mathbb{R}^{d\times 2}$, with $\Omega\subset\mathbb{R}^d$ the physical domain, defined by

$$\mathbf{G}(x):=\left[\nabla m_1|\nabla m_2\right]=\begin{bmatrix}\partial_{x_1}m_1 & \partial_{x_1}m_2\\ \vdots & \vdots\\ \partial_{x_d}m_1 & \partial_{x_d}m_2\end{bmatrix}.$$

The gradients ∇m_1 and ∇m_2 are aligned at $x\in\Omega$ if the columns of $\mathbf{G}(x)$ are multiples of each other, in which case the rank of $\mathbf{G}(x)$ is 1. One could seek to promote gradient alignment by minimizing $\int_\Omega \text{rank}\,(\mathbf{G}(x))\,\mathrm{d}x$. However, in practice, minimization of the rank of a matrix is notoriously difficult. The nuclear norm of a matrix, defined as the ℓ_1-norm of its singular values and denoted by $\|\cdot\|_*$, is often a good proxy for the rank [22]. We therefore define, with $\gamma>0$, the nuclear norm joint regularization as

$$\mathcal{R}_*(m_1,m_2):=\gamma\int_\Omega\|\mathbf{G}(x)\|_*\,\mathrm{d}x. \tag{16}$$

4.1. Gradient of the nuclear norm joint regularization

We now compute derivatives of (16) using the chain rule. Let us introduce the notation $f(M) := \|M\|_*$, for arbitrary $M \in \mathbb{R}^{d\times 2}$. Thus, $\mathcal{R}_*(m_1, m_2) = \gamma \int_\Omega f(\mathbf{G}(x))\,\mathrm{d}x$. Denoting the gradient of f with respect to the entries of the matrix M by $\nabla f(M) \in \mathbb{R}^{d\times 2}$, the first directional derivatives of (16) with respect to the inversion parameters m_i, $i = 1, 2$, in a direction $\tilde{m}_i$, are given by

$$\partial_{m_i}\mathcal{R}_*(m_1, m_2)\tilde{m}_i = \gamma \int_\Omega (\nabla f(\mathbf{G}), \partial_{m_i}\mathbf{G}(x)\tilde{m}_i)\,\mathrm{d}x, \tag{17}$$

where

$$\partial_{m_1}\mathbf{G}(x)\tilde{m}_1 = \begin{bmatrix} \partial_{x_1}\tilde{m}_1 & 0 \\ \vdots & \vdots \\ \partial_{x_d}\tilde{m}_1 & 0 \end{bmatrix} \text{ and } \partial_{m_2}\mathbf{G}(x)\tilde{m}_2 = \begin{bmatrix} 0 & \partial_{x_1}\tilde{m}_2 \\ \vdots & \vdots \\ 0 & \partial_{x_d}\tilde{m}_2 \end{bmatrix},$$

and the inner product for matrices $M = (m_{ij})_{ij}, N = (n_{ij})_{ij} \in \mathbb{R}^{d\times 2}$ is defined as $(M, N) = \sum_{i=1}^{d}\sum_{j=1}^{2} m_{ij}n_{ij}$.

We next compute the gradient of the nuclear norm $\nabla f(M)$. Given a full-rank matrix $M \in \mathbb{R}^{n\times m}$, i.e. $r := \operatorname{rank}(M) = \min(m, n)$, and singular values $\{\sigma_k\}_{k=1}^r$, we define its (reduced) singular value decomposition (SVD) by $M = U\Sigma V^T$, with $U \in R^{n\times r}$, $V \in \mathbb{R}^{m\times r}$, and $\Sigma \in \mathbb{R}^{r\times r}$ a diagonal matrix containing the singular values of M, i.e. $\Sigma_{kk} = \sigma_k > 0, k = 1, \ldots, r$. The (i, j)-entry of the gradient of the nuclear norm is given by

$$(\nabla f(M))_{ij} = \sum_{k=1}^{r} \frac{\partial \sigma_k}{\partial m_{ij}} = \sum_{k=1}^{r} u_{ik}v_{jk},$$

where the second equality uses the singular value sensitivity [23]. The gradient of the nuclear norm with respect to the entries of M is then given by

$$\nabla f(M) = UV^T.$$

4.2. Modified nuclear norm joint regularization

The nuclear norm $f(M)$ is non-differentiable when the matrix M is not full-rank, corresponding to the case where at least one of the singular values vanishes. To make it differentiable, similar to the treatment of TV regularization, we define the modified nuclear norm by

$$f_\varepsilon(M) := \|M\|_{*,\varepsilon} = \sum_{k=1}^{\min(m,n)} \sqrt{\sigma_k^2 + \varepsilon}, \tag{18}$$

where $\varepsilon > 0$. For $\gamma > 0$, we define the modified nuclear norm joint regularization as

$$\mathcal{R}_{*,\varepsilon}(m_1, m_2) := \gamma \int_\Omega f_\varepsilon(\mathbf{G}(x))\,\mathrm{d}x. \tag{19}$$

The (i, j)-entry of the gradient of the modified nuclear norm (18) is given by

$$(\nabla f_\varepsilon(M))_{ij} = \frac{\partial}{\partial m_{ij}} \sum_{k=1}^{\min(m,n)} \sqrt{\sigma_k^2 + \varepsilon} = \sum_{k=1}^{r} \frac{\sigma_k}{\sqrt{\sigma_k^2 + \varepsilon}} \frac{\partial \sigma_k}{\partial m_{ij}},$$

where in the last expression the sum is up to r since, by definition of the rank of a matrix, $\sigma_k = 0$ for all $k > r$. Let us now introduce the diagonal matrix $W_\varepsilon \in \mathbb{R}^{r\times r}$, with entries $(W_\varepsilon)_{ii} = \sigma_i/\sqrt{\sigma_i^2 + \varepsilon}$. Using the expression for the sensitivity of the singular values [23], the gradient of the modified nuclear norm is then given by

$$\nabla f_\varepsilon(M) = UW_\varepsilon V^T. \tag{20}$$

The first directional derivatives of (19) with respect to the inversion parameters m_i, $i = 1, 2$, in a direction $\tilde{m}_i$, are given by

$$\partial_{m_i}\mathcal{R}_{*,\varepsilon}(m_1, m_2)\tilde{m}_i = \gamma \int_\Omega (\nabla f_\varepsilon(\mathbf{G}), \partial_{m_i}\mathbf{G}(x)\tilde{m}_i)\,\mathrm{d}x. \tag{21}$$

The modified nuclear norm (18), however, is not twice differentiable when two singular values are equal (crossing singular values). This is because the second derivative requires the sensitivity of the individual singular vectors, which are not differentiable where singular values cross. We have not found a practical workaround for this singularity, and thus proceed with a gradient-based method to solve joint inverse problems regularized with the nuclear norm joint regularization; the solver is detailed in appendix A.2. In the rest of this paper, when using 'nuclear norm joint regularization', we refer to the modified nuclear norm joint regularization (19).

5. Numerical examples

In this section, we present a comprehensive numerical comparison of the four joint regularization approaches introduced in sections 2–4, i.e. the cross-gradient (5), the normalized cross-gradient (11), the vectorial total variation (13), and the nuclear norm (18) regularization. Reconstructions obtained with these joint regularization terms are compared with each other, and with the reconstructions obtained by solving a joint inverse problem with independent TV regularizations. The parameters for all joint regularization terms are selected empirically as leading to the best reconstructions. The values of ε are chosen small enough to provide reconstructions with sharp interfaces, but large enough to avoid numerical difficulties (see for instance the discussion in [24]).

The different regularizations are compared using three examples covering both types of joint inverse problems (1) and (2). In section 5.1, we combine two uncoupled Poisson inverse problems to form the joint inverse problem (2), where we invoke prior knowledge that the two truth parameter fields have similar structure. In section 5.2, we compare the ability of the joint regularization terms to improve the reconstruction of the bulk modulus and the density in an acoustic wave equation, an example of a joint inverse problem (1). Finally, in section 5.3, we formulate a multi-physics joint inverse problem (2), which combines an inverse problem governed by the Poisson equation with one governed by the acoustic wave equation. Here again, the Poisson parameter and the wave speed fields are assumed to have similar structure.

In all examples, the domain is a 2D unit square, with a uniform mesh of isosceles right triangles obtained by cutting in half $N \times N$ squares; we define the mesh size parameter $h := 1/N$. All data are generated synthetically from the truth parameter fields, and then polluted by adding independent and identically distributed Gaussian noise; the noise level is specific to each example. We use continuous Galerkin finite elements to discretize all field variables, with the state, adjoint, incremental state, and incremental adjoint variables using quadratic elements, and the parameter fields using linear elements. All examples are implemented in Python and built on the finite element library FEniCS [25, 26]. For the examples in section 5.1 and 5.3,

we used the optimization routines from hIPPYlib [27], a Python library for deterministic and Bayesian inverse problems. A short description of the numerical methods used for the solution of these problems can be found in appendix A. For details regarding the computation of the adjoint-based derivatives we refer to [20].

5.1. Joint Poisson inverse problems with different observation points

Here, we solve a joint inverse problem of the form (2) for the two coefficient fields m_1 and m_2. Considered separately, m_1 and m_2 are solutions to the (almost identical) TV-regularized inverse problems governed by the Poisson equation, i.e.

$$m_i := \arg\min_m \left\{ \frac{1}{2}|B_i u - \mathbf{d}_i|^2 + \gamma_i \int_\Omega \sqrt{|\nabla m|^2 + \varepsilon}\, \mathrm{d}x \right\}, \quad \text{where} \quad \begin{cases} -\nabla \cdot (\mathrm{e}^m \nabla u) = 1, \text{ in } \Omega, \\ u = 0, \text{ on } \partial\Omega. \end{cases} \tag{22}$$

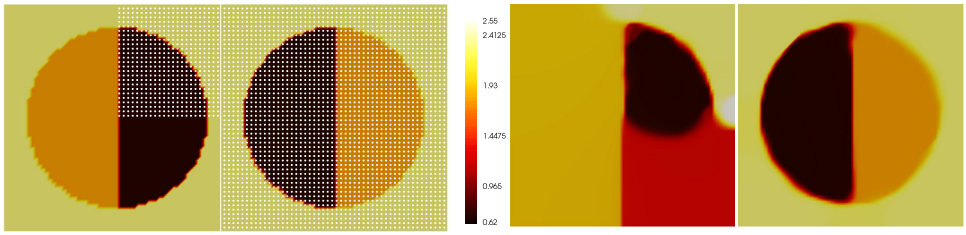
The operators B_i represent pointwise observation operators, and the data $\mathbf{d}_i$ are synthetic observations polluted with 2% Gaussian noise. The domain Ω is discretized with a mesh of 8192 triangles (i.e. $h = 1/64$). In all experiments presented in this section, the initial guesses for both parameter fields are constant zero over the domain, i.e. $m_1^0 \equiv 0$ and $m_2^0 \equiv 0$.

The differences between the inverse problems for m_1 and m_2 reside in the truth parameter fields, and in the observation operators B_i. In the first example (section 5.1.1), the truth parameter fields differ but have interfaces at the same spatial locations. In the second example (section 5.1.2), some interfaces in the truth parameter field for m_2 are not present in the truth parameter field for m_1. In both examples, the observation locations defined by B_1 only cover the top-right quadrant of the domain, whereas the observation locations defined by B_2 are distributed over the entire domain; see figures 4 and 6.

5.1.1. Truth parameter fields having identical interface locations. In the first example, the parameter fields have interfaces at the same locations. In figure 4, we show the truth parameter fields m_1 and m_2 and their reconstructions obtained by solving the inverse problems (22) independently. The reconstructions obtained with the four regularization methods are shown in figure 5, and the corresponding values of the relative medium misfit are given in table C1.

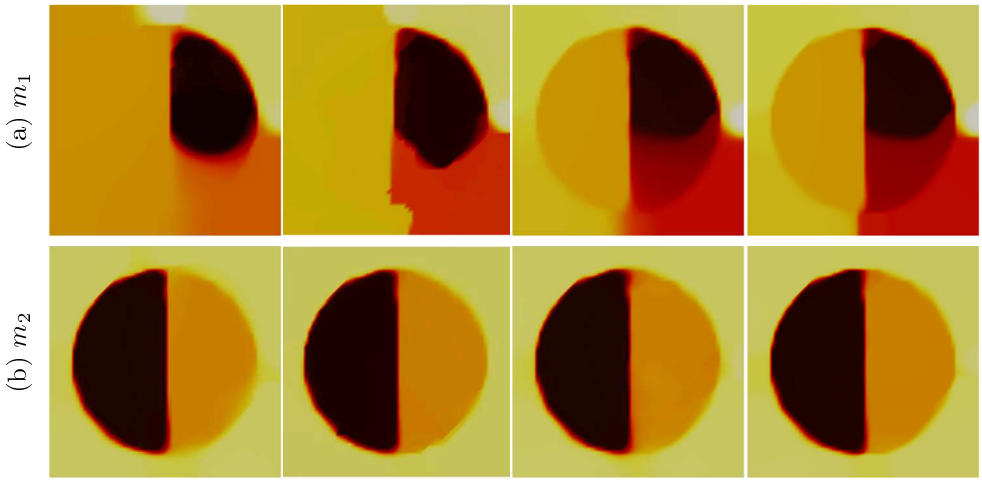
The reconstructions for parameter m_2 do not differ significantly (figure 5(b)). Due to the large number of observation points, this parameter is already well reconstructed in an independent inverse problem (figure 4(d)). We observe an improvement in the reconstruction of parameter m_1 for all four joint inverse problems compared to the independent reconstruction shown in figure 4(c). Using the cross-gradient only marginally improves the reconstruction for parameter m_1, most likely because the independent reconstruction for m_1 shows large areas of constant values, where the cross-gradient term vanishes; these areas therefore cannot be improved by the cross-gradient. The normalized cross-gradient improves over the cross-gradient but fails to recover the circular interface. Both the VTV joint regularization and the nuclear norm joint regularization perform better in this example, and lead to reconstructions that contain all sharp interfaces in the target image.

5.1.2. Truth parameter fields having different interface locations. Here, the only difference with the previous example is that the truth parameter field for m_1 no longer has a vertical discontinuity along the line $x = 0.5$ (see figure 6). In figures 6(c) and (d), we again show the reconstructions for parameters m_1 and m_2 obtained by solving two independent inverse



(a) Truth m_1 (b) Truth m_2 (c) Indep m_1 (d) Indep m_2

Figure 4. Parameter fields m_1 and m_2 in the example of section 5.1.1: truth parameter fields ((a), (b)) and reconstructions ((c), (d)) obtained by solving the inverse problem (22) with $\varepsilon = 10^{-3}$, $\gamma_1 = 3\cdot 10^{-7}$, $\gamma_2 = 4\cdot 10^{-7}$, and initial guesses $m_1^0 = m_2^0 = 0$. White dots in (a) and (b) indicate the location of the pointwise observations. The observation points defined through B_1 are a lattice of 25×25 points that cover only the top-right quadrant of the domain. The observation points for B_2 are a square lattice of 50×50 points distributed over the entire domain.



(a) m_1

(b) m_2

(i) cross-gradient (ii) norm. cross-gd (iii) vectorial TV (iv) nuclear norm

Figure 5. Reconstructions for the parameter fields (a) m_1 and (b) m_2, obtained by solving a joint inverse problem (2) regularized with (i) the cross-gradient ($\gamma = 2\cdot 10^{-8}$) combined with two independent TV regularizations, (ii) the normalized cross-gradient ($\gamma = 6\cdot 10^{-6}$ and $\varepsilon = 10^{-3}$) combined with the same independent TV regularizations, (iii) the VTV joint regularization ($\gamma = 3\cdot 10^{-7}$ and $\varepsilon = 10^{-3}$), and (iv) the nuclear norm joint regularization ($\gamma = 3\cdot 10^{-7}$ and $\varepsilon = 10^{-3}$). The parameters for the independent TV regularizations and the initial guesses for all problems are as for the independent inverse problems (see caption of figure 4). The legend for all plots is as in figure 4.

problems (22). The reconstructions for the four joint inverse problems are shown in figure 7, and the corresponding values of the relative medium misfit are given in table C1.

As in the previous example, for m_2 the reconstructions obtained with the different joint inverse problems do not differ significantly (see figure 7(b)). However, we observe differences among the reconstructions for parameter m_1. Using the cross-gradient only marginally improves the reconstruction for parameter m_1. The use of the normalized cross-gradient does not show improvement over the cross-gradient. As in the first example, both the VTV joint

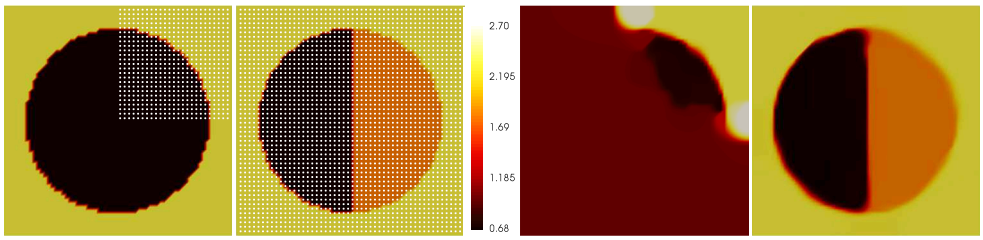


Figure 6. Parameter fields for m_1 and m_2 in the example of section 5.1.2: truth parameter field ((a), (b)) and reconstructions ((c), (d)) obtained by solving the inverse problem (22) with $\varepsilon = 10^{-3}$, $\gamma_1 = 4\cdot 10^{-7}$, $\gamma_2 = 4\cdot 10^{-7}$, and initial guesses $m_1^0 = m_2^0 = 0$. White dots in (a) and (b) indicate the location of the pointwise observations, as detailed in figure 4.

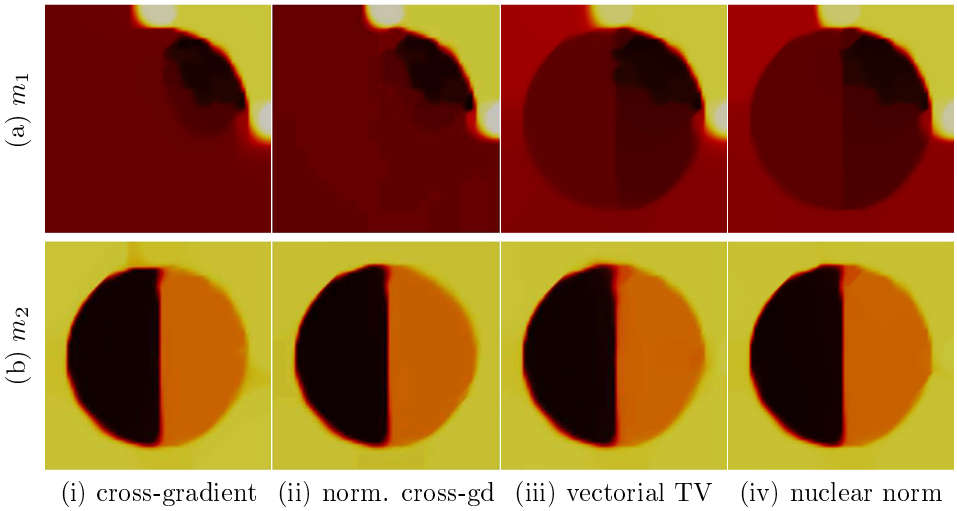


Figure 7. Reconstructions for the parameter fields (a) m_1 and (b) m_2, obtained by solving a joint inverse problem (2) regularized with (i) the cross-gradient combined with 2 independent TV regularizations ($\gamma = 5\cdot 10^{-9}$), (ii) the normalized cross-gradient combined with the same independent TV regularizations ($\gamma = 7\cdot 10^{-7}$ and $\varepsilon = 10^{-3}$), (iii) the VTV joint regularization ($\gamma = 4\cdot 10^{-7}$ and $\varepsilon = 10^{-3}$), and (iv) the nuclear norm joint regularization ($\gamma = 4\cdot 10^{-7}$ and $\varepsilon = 10^{-3}$). The parameters for the independent TV regularizations and all initial guesses are the same as used for the independent inverse problems (see caption in figure 6). The legend is as in figure 6.

regularization and the nuclear norm joint regularization perform the best, and their corresponding reconstructions contain all sharp interfaces present in the true image. However, in figures 7(a) (iii) and (iv) we also see a vertical discontinuity not present in the true image 6(c). This ghost interface in m_1 is due to the presence of such a discontinuity in m_2, and highlights the tendency of the VTV joint regularization and nuclear norm joint regularization to superimpose discontinuities in both parameters. Note, however, that the amplitude of this ghost interface is small compared to the amplitudes of the correctly recovered interfaces.

5.2. Joint inversion of bulk modulus and density in the acoustic wave equation

We now study a joint inverse problem of the form (1), i.e. both parameters enter the same equation, namely the acoustic wave equation.

5.2.1. Problem description. We start by defining the forward problem, i.e. the acoustic wave PDE. The propagation of acoustic waves depends on the bulk modulus κ and the density ρ of the medium of propagation. Let us define the acoustic pressure, $u(\mathbf{x},t) := -\kappa(\mathbf{x})\nabla\cdot\mathbf{u}(\mathbf{x},t)$, with $\mathbf{u}(\mathbf{x},t)$ the displacement vector at location $\mathbf{x}$ and time t. The time-domain acoustic wave equation with first order absorbing boundary condition [28] and initial conditions at rest is given by

$$\begin{aligned}
\frac{1}{\kappa}\ddot{u} - \nabla\cdot\left(\frac{1}{\rho}\nabla u\right) &= f, && \text{in } \Omega\times(0,T),\\
u(\mathbf{x},0) = \dot{u}(\mathbf{x},0) &= 0, && \text{in } \Omega,\\
\frac{1}{\rho}\nabla u\cdot\mathbf{n} &= 0, && \text{on } \partial\Omega_n\times(0,T),\\
\frac{1}{\rho}\nabla u\cdot\mathbf{n} &= -\frac{1}{\sqrt{\kappa\rho}}\dot{u}, && \text{on } \partial\Omega_a\times(0,T),
\end{aligned} \tag{23}$$

where f is a forcing term, $\dot{u}$ and $\ddot{u}$ are the first and second time derivatives of u, and the boundary of the domain $\partial\Omega$ is partitioned as $\partial\Omega = \partial\Omega_a\cup\partial\Omega_n$. The acoustic wave velocity of the medium is given by c, with the relation $\kappa = \rho c^2$. The PDE in (23) is the variable density form of the acoustic wave equation; when the density ρ is assumed constant, equation (23) reduces to $\frac{1}{c^2}\ddot{u} - \Delta u = \tilde{f}$.

Here, we assume that both the bulk modulus κ and the density ρ are unknown. Since they both appear in (23) through their inverse, we introduce the parameters $\alpha := 1/\kappa$ and $\beta := 1/\rho$, and formulate the inverse problem in terms of α and β. As common in seismic inversion, we consider N_s multiple experiments, characterized by their forcing terms f_i and datasets $\mathbf{d}_i$, which correspond to pointwise observations in space, recorded continuously in time. The acoustic wave inverse problem is then formulated as

$$\min_{\alpha,\beta>0}\left\{\frac{1}{2N_s}\sum_{i=1}^{N_s}\int_0^T |Bu_i(t) - \mathbf{d}_i(t)|^2\,\mathrm{d}t + \mathcal{R}(\alpha,\beta)\right\}, \tag{24}$$

where each u_i solves the forward problem (23) with forcing term f_i,

$$\begin{aligned}
\alpha\ddot{u}_i - \nabla\cdot(\beta\nabla u_i) &= f_i, && \text{in } \Omega\times(0,T),\\
u_i(\mathbf{x},0) = \dot{u}_i(\mathbf{x},0) &= 0, && \text{in } \Omega,\\
\beta\nabla u_i\cdot\mathbf{n} &= 0, && \text{on } \partial\Omega_n\times(0,T),\\
\beta\nabla u_i\cdot\mathbf{n} &= -\sqrt{\alpha\beta}\dot{u}_i, && \text{on } \partial\Omega_a\times(0,T).
\end{aligned}$$

In our experiments, the physical constraints $\alpha,\beta>0$ are never active, and therefore not enforced explicitly.

5.2.2. Solution of the acoustic wave joint inverse problem. Because the solution of the acoustic wave equation couples the parameters α and β, the inverse problem (24) could be regularized by two independent TV regularizations, i.e. $\mathcal{R}(\alpha,\beta) = \mathcal{R}_{\mathrm{TV},\varepsilon}(\alpha) + \mathcal{R}_{\mathrm{TV},\varepsilon}(\beta)$ [1].

However, the resulting problem can be difficult to solve and does not incorporate the structural correlation that usually exists between these parameters due to the types of rock occurring in the subsurface. Going beyond the use of ad hoc methods to handle both parameters at once, some researchers have addressed (24) as a joint inverse problem [2, 3]. Previous attempts have used the cross-gradient term, but not its normalized version, the VTV or the nuclear norm regularization. In this section, we study whether the use of joint regularization can improve reconstructions for α and β.

In our numerical tests, we use 6 independent sources, $f_i(\mathbf{x}, t)$, located on the top boundary of the domain at 0.1, 0.25, 0.4, 0.6, 0.75, and 0.9 from the left boundary (yellow stars in figure 8(a)); each source is a point source in space, and a Ricker wavelet in time with a central frequency of 2 Hz. The data are recorded at 20 locations equally spaced along the top boundary (green triangles in figure 8(b)), and polluted by independent Gaussian noise with zero mean and variance corresponding to a signal-to-noise ratio of 20 dB. The boundary conditions are a homogeneous Neumann boundary condition along the top boundary $\partial\Omega_n = [0, 1] \times \{1\}$, and an absorbing boundary condition along the left, bottom, and right boundaries $\partial\Omega_a = \{0, 1\} \times [0, 1] \cup [0, 1] \times \{0\}$. The truth parameter fields for α and β are shown in figure 8; they correspond to an acoustic wave velocity varying from 2 km s^{-1} to 3 km s^{-1}[4], typical values for a shallow subsurface (see for instance [29, 30]). The finite-element mesh consists of 800 triangles ($h = 1/20$). The initial guesses for parameters α and β are smoothed versions of the truth parameter fields (see figure 8(ii)).

In figure 8(iii), we show the reconstructions of parameters α and β obtained by solving (24) with independent TV regularizations. Whereas parameter α is well reconstructed, the reconstruction for β is rather poor. We next solve (24) with the proposed joint regularization terms. The results are shown in figure 9, and the corresponding values of the relative medium misfit are given in table C1.

The different reconstructions for α (figure 9(a)) do not differ significantly from each other. However, the use of joint regularization improves the quality of the reconstruction for the parameter β (figure 9(b)). Whereas the use of the cross-gradient only results in marginal improvement compared to the reconstruction in figure 8(d), the use of the normalized cross-gradient allows recovery of the interfaces more clearly. The best reconstructions are obtained using the VTV or the nuclear norm joint regularizations.

5.3. Joint inverse problem with different physics

As a last problem, we study a joint inverse problem (2) governed by two different physics models; namely, we combine a Poisson inverse problem and an acoustic wave inverse problem (assuming the density ρ is known). This inverse problem is intended as a model problem for joint seismic-electromagnetic inversion in the electromagnetic low frequency limit. The Poisson inverse problem is identical to the one used in section 5.1,

$$\min_m \left\{ \frac{1}{2}|Bu - \mathbf{d}|^2 + \gamma_m \int_\Omega \sqrt{|\nabla m|^2 + \varepsilon}\, \mathrm{d}x \right\}, \quad \text{where} \quad \begin{cases} -\nabla \cdot (\mathrm{e}^m \nabla u) &= 1 \text{ in } \Omega, \\ u &= 0 \text{ on } \partial\Omega. \end{cases} \tag{25}$$

The observation operator B extracts the state u at 20×20 equally distributed points over the entire domain (white dots in figure 10(a)). The data are polluted with 1% Gaussian noise.

[4] The following units are used: distance in km, velocity in km s^{-1}, density in g cm^{-3}, and bulk modulus in GPa.

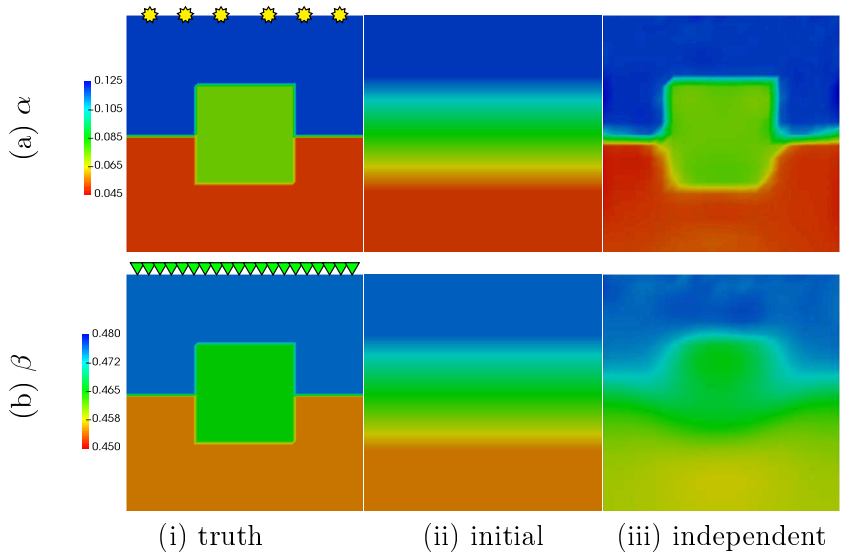


Figure 8. Parameter fields (a) α and (b) β in the joint acoustic inverse problem (24): (i) truth parameter fields, (ii) initial guesses, and (iii) reconstructions when solving (24) regularized with two independent TV regularizations ($\varepsilon = 10^{-3}$, $\gamma_\alpha = 5 \cdot 10^{-6}$, and $\gamma_\beta = 9 \cdot 10^{-6}$). The yellow stars in (a-i) and the green triangles in (b-i) indicate the locations of the point sources and observations, respectively.

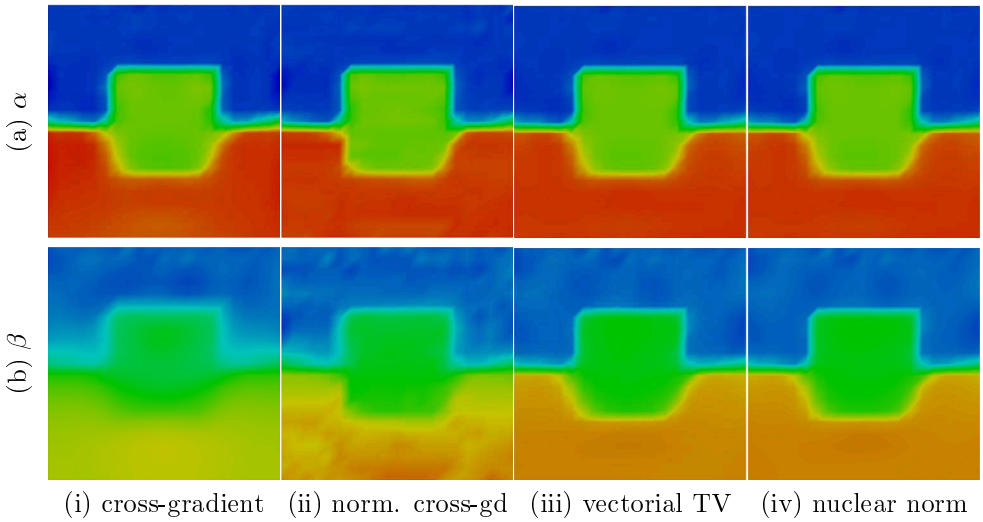


Figure 9. Reconstructions for the parameter fields (a) α and (b) β, obtained by solving (24) regularized with (i) the cross-gradient ($\gamma = 10^{-2}$) combined with two independent TV regularizations, (ii) the normalized cross-gradient ($\gamma = 9 \cdot 10^{-6}$ and $\varepsilon = 10^{-6}$) combined with the same independent TV regularizations, (iii) the VTV joint regularization ($\gamma = 7 \cdot 10^{-6}$ and $\varepsilon = 10^{-3}$), and (iv) the nuclear norm joint regularization ($\gamma = 7 \cdot 10^{-6}$ and $\varepsilon = 10^{-3}$). The parameters for the independent TV regularizations are the ones selected for the independent inverse problems (see caption in figure 8). The legend is as in figure 8.

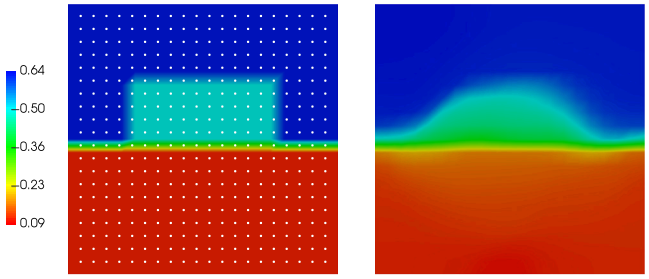


Figure 10. Plots of (a) truth parameter field for m in the Poisson inverse problem (25), and (b) its reconstruction ($\gamma_m = 2\cdot 10^{-8}$ and $\varepsilon = 10^{-3}$) with initial parameter field set to a constant value of 0.625. The white dots in (a) indicate the location of the pointwise observations.

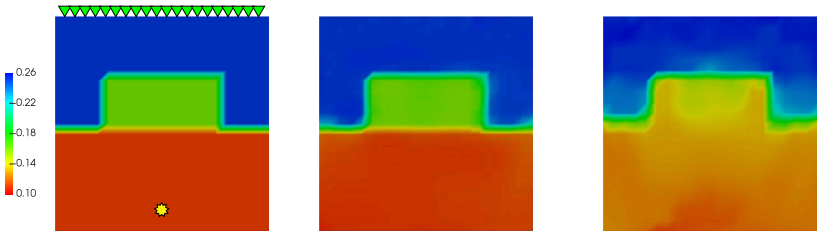
For the acoustic wave inverse problem, we set $\beta \equiv 1$, and invert only for the parameter $\alpha = 1/\kappa = 1/c^2$,

$$\min_{\alpha} \left\{ \frac{1}{2} \int_0^T |Bu(t) - \mathbf{d}(t)|^2 \, \mathrm{d}t + \gamma_\alpha \int_\Omega \sqrt{|\nabla \alpha|^2 + \varepsilon} \right\} \mathrm{d}x, \quad \text{where}$$
$$\begin{cases} \alpha \ddot{u} - \Delta u = f_\alpha, & \text{in } \Omega \times (0,T), \\ u(\mathbf{x},0) = \dot{u}(\mathbf{x},0) = 0, & \text{in } \Omega, \\ \nabla u \cdot \mathbf{n} = 0, & \text{on } \partial\Omega_n \times (0,T), \\ \nabla u \cdot \mathbf{n} = -\sqrt{\alpha}\dot{u}, & \text{on } \partial\Omega_a \times (0,T). \end{cases} \tag{26}$$

We use a single source f_α with frequency 2 Hz or 4 Hz, located at $(0.5, 0.1)$ (yellow star in figure 11(a)), and 20 pointwise observations equally spaced along the top boundary (green triangles in figure 11(a)). The boundary conditions, the noise level in the data, the mesh, and the numerical discretization are as in section 5.2. The initial guess for the Poisson parameter field m (resp. for the acoustic parameter field α) is set to a constant field with value 0.625 (resp. 0.25), corresponding to the value in the upper layer of the truth parameter field, in blue in figure 10(a) (resp. figure 11(a)).

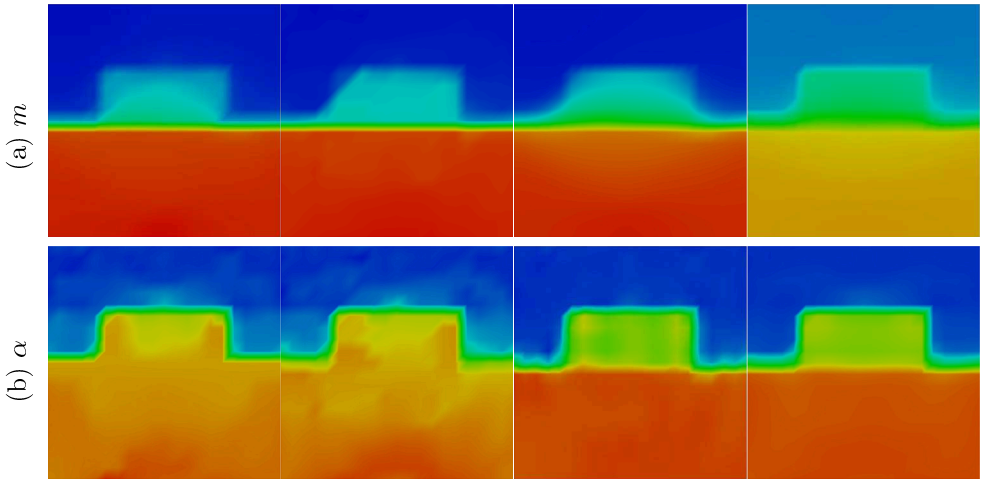
As reference, we first solve the inverse problem for the parameters m and α when (25) and (26) are solved independently. The results for the Poisson inverse problem (26) are shown in figure 10(b), where it can be seen that the horizontal interface is well reconstructed, but the shape of the rectangular perturbation is smeared out. For the acoustic wave inverse problem (26), we show two reconstructions in figure 11, one with a source f_α of frequency 2 Hz (figure 11(b)), and one with a source f_α of frequency 4 Hz (figure 11(c)). While the reconstruction at 2 Hz is excellent, the reconstruction at 4 Hz lacks sufficient low-frequency information and appears to converge toward a local minimum, missing the horizontal discontinuity present in the truth parameter field (figure 11(a)). The reconstructions for all four joint inverse problems, with a seismic source f_α of frequency 4 Hz, are shown in figure 12, and the corresponding values of the relative medium misfit are given in table C1.

The use of the cross-gradient or its normalized variant improves the reconstruction for the Poisson parameter m (figures 12(a), (i) and (ii)), compared to the reconstruction from the Poisson inverse problem (25) alone (figure 10(b)). However, neither of the cross-gradient terms brings any improvement to the reconstruction of the acoustic wave velocity



(a) truth (b) reconstruction 2 Hz (c) reconstruction 4 Hz

Figure 11. Plots of (a) truth parameter field for α in the acoustic inverse problem (26), and ((b), (c)) its reconstructions ($\gamma_\alpha = 3\cdot 10^{-8}$ and $\varepsilon = 10^{-3}$) with initial value for the parameter field set to 0.25, and a source f_α of frequency (b) 2 Hz, and (c) 4 Hz. The green triangles in (a) indicate the locations of the pointwise observations, and the yellow star in (a) indicates the location of the source.



(i) cross-gradient (ii) norm. cross-gd (iii) vectorial TV (iv) nuclear norm

Figure 12. Reconstructions for the parameter fields (a) m in (25) and (b) α in (26), obtained by solving a joint inverse problem with seismic source f_α of frequency 4 Hz, and regularized with (i) the cross-gradient ($\gamma = 8\cdot 10^{-7}$) combined with two TV regularizations, (ii) the normalized cross-gradient ($\gamma = 8\cdot 10^{-8}$ and $\varepsilon = 10^{-5}$) combined with the same TV regularizations, (iii) the VTV joint regularization ($\gamma = 4\cdot 10^{-8}$ and $\varepsilon = 10^{-3}$), and (iv) the nuclear norm joint regularization ($\gamma = 5\cdot 10^{-7}$ and $\varepsilon = 10^{-3}$). The parameters for the independent TV regularizations are as for the independent inverse problems (see captions of figures 10 and 11). Legend is the same as in figures 10 and 11.

(figures 12(b), (i) and (ii)); in particular, the reconstructions do not show the horizontal discontinuity that was missing in the reconstruction of the acoustic wave velocity alone (figure 11(c)). On the other hand, the use of either the VTV joint regularization, or the nuclear norm joint regularization, leads to significant improvements in the reconstruction of the acoustic wave velocity (figures 12(b), (iii) and (iv)). Both reconstructions contain all features of the

truth parameter field (figure 11(a)); most noticeably, the horizontal discontinuity that was missing in the independent reconstruction (figure 11(c)) is now fully reconstructed. The use of the VTV joint regularization provides only marginal improvement to the reconstruction of the Poisson parameter m, in terms of relative medium misfit (table C1); however, the shape of the rectangular perturbation, which was smeared out in the reconstruction from the Poisson inverse problem alone (figure 10(b)), is clearer in figure 12(a) (iii). The reconstruction of the Poisson parameter obtained with the nuclear norm joint regularization indicates that the optimization converged to a local minimum. Despite all discontinuities present in the truth parameter field (figure 10(a)) being clearly reconstructed in figure 12(a) (iv), the values of the parameters are significantly different. Similar, or worse, performance was observed when setting H_0 to be a multiple of the identity matrix in the BFGS solver [17]. Moreover, almost identical results were obtained when solving the Poisson-acoustic joint inverse problem, regularized by VTV, using the BFGS method described in appendix A.2. We therefore conjecture that the poor performance of the nuclear norm joint regularization, in the case of a multi-physics joint inverse problem, can be attributed to the use of a gradient-based method for the solution of the joint inverse problem. The significant difference in the structure of the gradients, coming from the Poisson and acoustic wave inverse problems, dictate the use of a Newton method, which is affine-invariant, in order to balance the individual search directions. This conjecture is supported by previous results found in the literature. For instance, in the context of a joint full waveform inversion for the conductivity and permittivity of a medium, the authors in [31] found the reconstructions obtained using the L-BFGS method to be highly sensitive to the scaling of the parameter fields. The authors of [32] report similar difficulties when employing a quasi-Newton method on a cross-well example, inverting for compressibility and anisotropy, and study alternative formulations to remedy this problem.

6. Conclusion

We conducted a systematic review of regularization terms for joint inverse problems governed by PDEs with infinite-dimensional parameter fields. We considered two types of joint inverse problems: (1) those coupling several uncoupled physics forward problems via joint regularization terms, and (2) those in which all inversion parameters depend on the same physics. Based on a review of the literature, we identified three joint regularization terms for this study that are tractable for large-scale PDE constrained joint inverse problems. The cross-gradient is a popular choice in geophysical applications and seeks to align level sets of the parameter fields. The normalized cross-gradient was designed to overcome some of the potential weaknesses of the cross-gradient term. The vectorial total variation is an extension of total variation regularization to joint inverse problems, and originated from the imaging community. In addition, we introduced a fourth novel joint regularization term based on the nuclear norm of a gradient matrix. The comparison of these joint regularization terms was carried out for three problems: (1) a joint Poisson inverse problem for which the truth parameter fields are known to share a similar structure, (2) an acoustic wave inverse problem in which we invert for the bulk modulus and the density, and (3) a joint Poisson-acoustic wave inverse problem, providing an example of multiple physics joint inversion.

Based on this study, we recommend use of the vectorial total variation joint regularization. It leads to superior reconstructions in all our examples. Moreover, we have available a scalable, efficient primal-dual nonlinear optimization solver and Hessian preconditioner for

joint inverse problems regularized with this term [20]. The nuclear norm joint regularization showed encouraging results, even leading to slightly better reconstructions than the vectorial total variation for some examples. However, its numerical realization is challenging since it is not twice differentiable as required by Newton's method. For piecewise-homogeneous parameter fields, the cross-gradient similarity term does not improve significantly over independent reconstructions. In particular, it can fail to reconstruct some edges entirely, since the cross-gradient term vanishes at points where one parameter field is constant. The normalized cross-gradient similarity term leads to a joint inverse problem that is challenging to solve numerically. Even though it improves on the cross-gradient, the improvement is generally minimal, and the reconstructions do not compare favorably with the ones obtained with vectorial total variation. Compared to the cross-gradient approaches, an additional advantage of the VTV and nuclear norm functionals is that they also act as regularizations, making individual regularization functionals unnecessary. This reduces the number of hyperparameters or regularization weights that must be chosen (see table B1), thereby simplifying the inverse problem.

Acknowledgments

The authors would like to thank David Keyes (KAUST) and George Turkiyyah (AUB) for very helpful discussions that inspired this work. They would also like to thank Sergey Fomel (UT-Austin) for referring them to [2, 3], Jan Modersitzki (Lübeck) for directing their attention to [9], Nick Alger (UT-Austin) for help with figure 1, and two anonymous referees whose comments helped significantly to improve the manuscript. This work was partially supported by AFOSR grant FA9550-17-1-0190, DOE grant DE-SC0009286, KAUST award OSR-2016-CCF-2596, and NSF grants ACI-1550593, DMS-1723211, CBET-1507009, and CBET-1508713.

Appendix A. Summary of numerical optimization techniques for the solution of regularized inverse problems

In this section, we describe the large-scale numerical optimization methods used for our numerical examples. As already discussed in the introduction, the solution of PDE-constrained optimization problems typically requires iterative methods. These methods require first (and ideally, also second) derivatives of the objective function with respect to the parameter fields [17, 33]. These derivatives can be computed efficiently using adjoint methods [34–36]. In particular, the computation of a gradient requires one solve of the governing state equation, the solution of an adjoint equation and the evaluation of the expression for the gradient. Moreover, adjoint methods allow the computation of directional second derivatives by solving two linear PDEs, one a linearization of the state equation, and the other one a linearization of the adjoint. Since these PDE solves usually dominate all other required operations, one often measures the complexity of PDE-constrained optimization algorithms by the number of required PDE solves. Line search and trust-region methods are employed to globalize local optimization methods [17]. We restrict our description here to the former, since we use line search methods in this paper. In the remainder of this section, we denote the parameter field pair by $m = (m_1, m_2)$, the objective function by $\mathcal{J}(m)$ and use upper indices to denote iteration numbers.

A.1. Line-search Newton-CG for cross-gradient and VTV regularizations

In the kth iteration, we update the medium parameters $m^{(k)}$ along a search direction $p^{(k)}$ by computing $m^{(k+1)} = m^{(k)} + \alpha^{(k)} p^{(k)}$ with an appropriate step length $\alpha^{(k)} > 0$. To ensure convergence, the search direction must be a descent direction, i.e. it must satisfy $\langle g^{(k)}, p^{(k)} \rangle < 0$, where $g^{(k)}$ is the gradient of $\mathcal{J}$ with respect to m evaluated at $m^{(k)}$, and $\langle \cdot, \cdot \rangle$ is an appropriate inner product. The step length $\alpha^{(k)}$ could be chosen to minimize the objective functional along this search direction $p^{(k)}$. However, solving this minimization problem exactly is too expensive for large-scale applications, since a single evaluation of the objective functional requires the solution of the state PDE, potentially multiple times (e.g. N_s times in the example of section 5.2 which has multiple sources). Instead, we seek an approximate minimizer that satisfies the following Armijo condition to ensure sufficient descent,

$$\mathcal{J}(m^{(k)} + \alpha^{(k)} p^{(k)}) \leqslant \mathcal{J}(m^{(k)}) + c_1 \alpha^{(k)} \langle g^{(k)}, p^{(k)} \rangle, \tag{A.1}$$

with $0 < c_1 < 1$. To ensure sufficiently large step lengths, we use backtracking line search [17] to find a step length that satisfies (A.1). That is, the step length is computed by starting from an initial guess $\alpha_0^{(k)} > 0$, and is reduced until the sufficient descent condition (A.1) is satisfied. When computing the search direction for a Newton-type method (see next paragraph), we use $\alpha_0^{(k)} = 1$, since this is guaranteed to be a successful step length in a neighborhood of a minimizer [17].

The choice of good search directions is crucial in PDE-constrained optimization. In the steepest descent method, one chooses the search direction as the negative gradient, i.e. $p^{(k)} = -g^{(k)}$. Unfortunately, the resulting algorithm usually converges slowly in the presence of stretched contour lines of the objective $\mathcal{J}$, a consequence of the typical ill-posedness of inverse problems. The Newton direction is given by the solution of the linear system $H(m^{(k)})\, p^{(k)} = -g^{(k)}$, where $H(m^{(k)})$ is the Hessian, i.e. the second derivative of $\mathcal{J}$, evaluated at $m^{(k)}$. The direction $p^{(k)}$ arising as solution of this equation is a descent direction only if the Hessian is positive definite, which may not be the case, in particular far away from the minimizer. When the Hessian is indefinite, one solution is to replace the Hessian with a positive definite approximation, a common choice being the Gauss–Newton Hessian [17]. This approximation is obtained by setting the adjoint variables to zero in the computation of the Hessian. Another option is to retain the full Hessian but solve the Newton system approximately, in a way that guarantees the computed solution to be a descent direction. Since for large-scale problems exactly constructing the Hessian is infeasible, we solve the Newton system using the conjugate gradient (CG) method. This requires only Hessian-vector products as provided by the adjoint method. CG is a Krylov subspace iterative method that solves linear systems of equations for symmetric positive definite matrices. When applied to a Newton system with an indefinite Hessian it will eventually encounter a negative curvature direction and fail. However, one can use the previous iterate before the iteration failed as the search direction, which is guaranteed to be a descent direction [17]. Additionally, regardless of the definiteness of the Hessian, we terminate the CG iterations prematurely to avoid over-solving, that is, we solve the Newton system with a coarse termination tolerance, thus applying just a few iterations of the CG method [18]. As the optimization iteration converges, the tolerance is gradually decreased to allow increasingly accurate computation of the Newton search

direction in order to benefit from the fast local convergence properties of the Newton method. For our experiments with cross-gradient regularization, its normalized version and the vector-TV regularization, we use the Newton-CG method with backtracking line search described above. For the nuclear norm regularization, we do not use directional second derivatives, but instead approximate the Newton direction using a BFGS method, as summarized next.

A.2. BFGS method for nuclear norm regularization

To solve joint inverse problems regularized with the nuclear norm joint regularization (section 4), we use a BFGS quasi-Newton method with damped update [17]. That is, we find the search direction $p^{(k)}$ by computing $p^{(k)} = -B^{(k)}g^{(k)}$, where $g^{(k)}$ is again the gradient of the objective function and $B^{(k)}$ is a positive definite approximation of the inverse of the Hessian. This approximation is updated at each iteration with the rank-2 update

$$B^{(k+1)} = (I - \rho_k r^{(k)}(y^{(k)})^T)B^{(k)}(I - \rho_k y^{(k)}(r^{(k)})^T) + \rho_k r^{(k)}(r^{(k)})^T, \tag{A.2}$$

where $y^{(k)}$ is the difference between the gradient at steps $k+1$ and k, $\rho_k := 1/(y^{(k)})^T r^{(k)}$, and $r^{(k)}$ is the damped form of $s^{(k)}$, the difference between the parameter at steps $k+1$ and k, and is defined as $r^{(k)} := \theta_k s^{(k)} + (1-\theta_k)B^{(k)}y^{(k)}$, with

$$\theta_k := \begin{cases} 1, & \text{if } (s^{(k)})^T y^{(k)} \geqslant \alpha (y^{(k)})^T B^{(k)} y^{(k)}, \\ \frac{(1-\alpha)(y^{(k)})^T B^{(k)} y^{(k)}}{(y^{(k)})^T B^{(k)} y^{(k)} - (s^{(k)})^T y^{(k)}}, & \text{otherwise.} \end{cases}$$

The classical BFGS method requires the curvature condition $(s^{(k)})^T y^{(k)} > 0$ to be satisfied at all steps. This condition is necessary to maintain positive definiteness of $B^{(k)}$ for all k. However, the curvature condition can be guaranteed to be satisfied only when the objective function is strictly convex, which is typically not the case for nonlinear inverse problems. Using a damped update allows us to apply a backtracking line search, while avoiding skipping some updates of $B^{(k)}$ entirely. In our numerical experiments, we found that $\alpha = 0.2$ worked well. The BFGS formula (A.2) requires the initialization $B^{(0)}$. BFGS-type methods perform well when the difference between the initial Hessian approximation and the true Hessian is a compact operator [37]. Thus, we take $B^{(0)}$ as the inverse of the Hessian of the regularization. This quantity is not available for the nuclear norm joint regularization. However, VTV and the nuclear norm joint regularization come from the same family of joint regularizations, differing only by the matrix norm employed [14]. Since matrix norms are equivalent in finite dimensions, we set $B^{(0)}$ to the inverse of the Hessian of the VTV joint regularization at the parameter $m^{(k)}$.

Appendix B. Number of hyperparameters for each joint regularization

Table B1. Number of hyperparameters for a joint inverse problem with 2 parameter fields.

	Joint regularization			
	Cross-grad	n-cross-grad	Vectorial TV	Nuclear norm
γ				
TV	2	2	—	—
joint	1	1	1	1
ε				
TV	1	1	—	—
joint	1	1	1	1
Total	5	5	2	2

Appendix C. Table of relative medium misfits for examples

In table C1, the relative misfits for the examples presented in section 5 are summarized.

Table C1. Relative medium misfits (in L^2-norm) for the examples in section 5.

	Section 5.1.1		Section 5.1.2		Section 5.2		Section 5.3	
	m_1 (%)	m_2 (%)	m_1 (%)	m_2 (%)	α (%)	β (%)	m (%)	α (%)
Independent	23.2	5.1	46.9	5.1	2.8	0.8	9.0	9.9
Cross-grad	22.3	5.2	46.1	5.6	3.1	0.7	4.9	11.0
n-cross-grad	21.2	5.0	46.7	5.0	2.5	0.4	4.9	10.7
Vectorial TV	20.2	5.1	41.1	5.2	2.4	0.2	8.9	3.3
Nuclear norm	20.2	4.8	40.8	5.0	2.4	0.2	20.6	4.5

ORCID iDs

Georg Stadler https://orcid.org/0000-0001-7762-6544

References

[1] Epanomeritakis I, Akçelik V, Ghattas O and Bielak J 2008 A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion *Inverse Problems* **24** 034015
[2] Manukyan E, Maurer H and Nuber A 2018 Improvements to elastic full waveform inversion using cross-gradient constraints *Geophysics* **83** R105–15
[3] Li M, Liang L, Abubakar A and van den Berg P M 2013 Structural similarity regularization scheme for multiparameter seismic full waveform inversion *SEG Technical Program Expanded Abstracts* (Los Angeles, CA: Society of Exploration Geophysicists) pp 1089–94

[4] Abubakar A, Gao G, Habashy T M and Liu J 2012 Joint inversion approaches for geophysical electromagnetic and elastic full-waveform data *Inverse Problems* **28** 055016
[5] Semerci O, Pan G, Li M, Liang L and Habashy T 2014 Joint electromagnetic and seismic inversion for petrophysical parameters using multi-objective optimization *SEG Annual Meeting (Denver, CO, 26–31 October 2014)* (SEG) (https://doi.org/10.1190/segam2014-1245.1)
[6] Feng X, Ren Q, Liu C and Zhang X 2017 Joint acoustic full-waveform inversion of crosshole seismic and ground-penetrating radar data in the frequency domain *Geophysics* **82** H41–56
[7] Gallardo L A and Meju M A 2003 Characterization of heterogeneous near-surface materials by joint 2D inversion of DC resistivity and seismic data *Geophys. Res. Lett.* **30** 1658
[8] Steklova K and Haber E 2016 Joint hydrogeophysical inversion: state estimation for seawater intrusion models in 3D *Comput. Geosci.* **1** 75–94
[9] Haber E and Modersitzki J 2006 Intensity gradient based registration and fusion of multi-modal images *Proc. of Medical Image Computing and Computer-Assisted Intervention* pp 726–33
[10] Haber E and Gazit M H 2013 Model fusion and joint inversion *Surv. Geophys.* **34** 675–95
[11] Gallardo L A and Meju M A 2011 Structure-coupled multiphysics imaging in geophysical sciences *Rev. Geophys.* **49** RG1003
[12] Blomgren P and Chan T F 1998 Color TV: total variation methods for restoration of vector-valued images *IEEE Trans. Image Process.* **7** 304–9
[13] Bresson X and Chan T F 2008 Fast dual minimization of the vectorial total variation norm and applications to color image processing *Inverse Problems Imaging* **2** 455–84
[14] Holt K M 2014 Total nuclear variation and jacobian extensions of total variation for vector fields *IEEE Trans. Image Process.* **23** 3975–89
[15] Knoll F, Holler M, Koesters T, Otazo R, Bredies K and Sodickson D K 2017 Joint MR-PET reconstruction using a multi-channel image regularizer *IEEE Trans. Med. Imaging* **36** 1–16
[16] Akçelik V, Biros G and Ghattas O 2002 Parallel multiscale Gauss–Newton–Krylov methods for inverse wave propagation *Proc. of IEEE/ACM SC2002 Conf. (Baltimore, MD)* (https://doi.org.10.1109/SC.2002.10002)
[17] Nocedal J and Wright S J 2006 *Numerical Optimization* 2nd edn (Berlin: Springer)
[18] Dembo R S, Eisenstat S C and Steihaug T 1982 Inexact Newton methods *SIAM J. Numer. Anal.* **19** 400–8
[19] Chambolle A, Caselles V, Cremers D, Novaga M and Pock T 2010 An introduction to total variation for image analysis *Theoretical Foundations and Numerical Methods for Sparse Recovery* (Berlin: Walter de Gruyter) pp 263–340
[20] Crestel B 2017 Advanced techniques for multi-source, multi-parameter, and multi-physics inverse problems *PhD Thesis* University of Texas at Austin
[21] Hintermüller M and Stadler G 2006 An infeasible primal-dual algorithm for total variation-based inf-convolution-type image restoration *SIAM J. Sci. Comput.* **28** 1–23
[22] Fernandez-Granda C 2016 Optimization-based data analysis *Lecture Notes at NYU-CIMS*
[23] Papadopoulo T and Lourakis M I A 2000 *Estimating the Jacobian of the Singular Value Decomposition: Theory and Applications* (Berlin: Springer) pp 554–70
[24] Ascher U M, Haber E and Huang H 2006 On effective methods for implicit piecewise smooth surface recovery *SIAM J. Sci. Comput.* **28** 339–58
[25] Logg A, Mardal K-A and Wells G 2012 *Automated Solution of Differential Equations by the Finite Element Method: the FEniCS Book* vol 84 (Berlin: Springer)
[26] Logg A and Wells G N 2010 DOLFIN: automated finite element computing *ACM Trans. Math. Softw.* **37** 20
[27] Villa U, Petra N and Ghattas O 2018 hIPPYlib: an extensible software framework for large-scale inverse problems *J. Open Source Software* **3** 940
[28] Engquist B and Majda A 1977 Absorbing boundary conditions for the numerical simulation of waves *Math. Comput.* **31** 629–51
[29] Martin G S, Wiley R and Marfurt K J 2006 Marmousi2: an elastic upgrade for Marmousi *Leading Edge* **25** 156–66
[30] Virieux J and Operto S 2009 An overview of full-waveform inversion in exploration geophysics *Geophysics* **74** WCC1–26
[31] Lavoué F, Brossier R, Métivier L, Garambois S and Virieux J 2014 Two-dimensional permittivity and conductivity imaging by full waveform inversion of multioffset GPR data: a frequency-domain quasi-Newton approach *Geophys. J. Int.* **197** 248–68

[32] Peters B and Herrmann F J 2014 A sparse reduced Hessian approximation for multi-parameter wavefield reconstruction inversion *SEG Technical Program Expanded Abstracts* (Los Angeles, CA: Society of Exploration Geophysicists) pp 1206–10
[33] Vogel C R 2002 *Computational Methods for Inverse Problems* (*Frontiers in Applied Mathematics*) (Philadelphia, PA: SIAM)
[34] De Los Reyes J C 2015 *Numerical PDE-Constrained Optimization* (Berlin: Springer)
[35] Hinze M, Pinnau R, Ulbrich M and Ulbrich S 2009 *Optimization with PDE Constraints* (Berlin: Springer)
[36] Tröltzsch F 2010 *Optimal Control of Partial Differential Equations: Theory, Methods and Applications* (*Graduate Studies in Mathematics* vol 112) (Providence, RI: American Mathematical Society)
[37] Griewank A 1987 The local convergence of Broyden-like methods on Lipschitzian problems in Hilbert space *SIAM J. Numer. Anal.* **24** 684–705