Surv Geophys (2013) 34:675–695
DOI 10.1007/s10712-013-9232-4

Model Fusion and Joint Inversion

Eldad Haber · Michal Holtzman Gazit

Received: 25 September 2012 / Accepted: 15 April 2013 / Published online: 23 August 2013

Abstract Inverse problems are inherently non-unique, and regularization is needed to obtain stable and reasonable solutions. The regularization adds information to the problem and determines which solution, out of the infinitely many, is obtained. In this paper, we review and discuss the case when a priori information exists in the form of either known structure or in the form of another inverse problem for a different property. The challenge is to include such information in the inversion process. To use existing known structure, we review the concept of model fusion, where we build a regularization functional that fuses the inverted model to a known one. The fusion is achieved by four different techniques. Joint inversion of two data sets is achieved by using iterative data fusion. The paper discusses four different methods for joint inversion. We discuss the use of correspondence maps or the petrophysics of the rocks, as well as structure. In particular, we suggest to further stabilize the well-known gradient cross product and suggest a new technique, Joint Total Variation, to solve the problem. The Joint Total Variation is a convex functional for joint inversion and, as such, has favorable optimization properties. We experiment with the techniques on the DC resistivity problem and the borehole tomography and show how model fusion and joint inversion can significantly improve over existing techniques.

Keywords Joint inversion · Model fusion · Total variation · Cross product

1 Introduction

Consider the problem of recovering a model function from the noisy incomplete data obtained by the forward problem

E. Haber
Department of Mathematics and Earth and Ocean Science, University of British Columbia, Vancouver, BC, Canada
e-mail: haber@math.ubc.ca

M. Holtzman Gazit (✉)
Department of Mathematics, University of British Columbia, Vancouver, BC, Canada
e-mail: mikih@cs.ubc.ca

$$F(m) + \epsilon = d. \tag{1}$$

where $F(\cdot)$ is a forward operator that maps the model function $m(x)$ into the discrete data vector d and ϵ is random measurement noise assumed to be Gaussian independent and identically distributed (iid) with standard deviation σ^2. Assume that the problem is ill-posed, that is, the solution is not unique and discontinuous with respect to data perturbations. Small perturbations in d can lead to large changes in m. Such problems are typically solved by regularization where *meaningful* a priori information is added to the inversion process. This is achieved by solving an optimization problem of the form

$$\min_m \mathcal{J}(m) = \frac{1}{2}\|F(m) - d\|^2 + \alpha R(m) \tag{2}$$

where $R(m)$ is an appropriate regularization operator and α is a regularization parameter.

Typical choices for the regularization operator are based on smoothness requirements of the solution (Jackson 1979; Parker 1994). Using ℓ_2 norm on the gradient of m as well as total variation is among the popular choices (Vogel, 2001).

In geophysical applications, the operator F can be the direct current resistivity method, the electromagnetic imaging method, or the seismic forward problem. The model $m(x)$ can be a conductivity or the seismic velocity of the Earth, and the optimization problem (2) can be used to approximately recover these physical properties.

To be more concrete, in this work, we experiment with two model problems. The first is the seismic ray tomography, and the second is the 2D DC (Direct Current) resistivity. We now review these problems and demonstrate how non-uniqueness in the recovered model arises.

1.1 Borehole Tomography

Assume that the Earth is probed using two boreholes. Sources are put in one borehole, and receivers are placed in the other. Assuming straight rays, the data at the jth receiver due to signal omitted at the ith source are given by

$$D_{ij} = \int_{\Gamma_{ij}} m(x)\mathrm{d}\ell$$

where Γ_{ij} is the (assumed straight) path from the ith source to the jth receiver and $m(x)$ is a function that represents the slowness. To obtain a discrete approximation, we first organize the data, D, in a vector d where d_k is one datum that corresponds to the kth ray. Next, the model is discretized into n cells. Let A_{ij} be the length of the ith ray in the jth cell. Then, the discrete problem can be written as

$$Am + \epsilon = d$$

where m is the discretized slowness at each cell[1]. This is a linear inverse problem for the model m given the data d.

[1] Throughout the paper we somewhat abuse the notations between continuous and discrete variables. The understanding of what type of variable should be clear from the context.

1.2 Direct Current Resistivity

Direct Current (DC) resistivity is an imaging technique that uses static electric fields in order to image the Earth's interior. In the experiment, source electrodes are placed at x_{s1} and x_{s2} and direct current is injected through them into the Earth. The electric potential due to the current can be modeled as a partial differential equation

$$\nabla \cdot \exp(m(x))\nabla u_i = \delta(x^i_{s_1}) - \delta(x^j_{s_2}) \quad u_i \in \Omega \quad \nabla u_i \cdot \mathbf{n} = 0 \quad u_i \in \partial\Omega \quad i = 1, \ldots, n_s$$

Here, $m(x)$ is the log conductivity, and we assume to have n_s sources. Given the potential field u_i, one measures the datum D_{ij} by another set of electrodes, that is, the data are given by

$$D_{ij} = \int_\Omega \left(\delta(x^j_{r_1}) - \delta(x^j_{r_2})\right) u_i(x)\mathrm{d}x$$

Upon discretization, one obtains a discrete forward problem of the form

$$D = P^\top A(m)^1 Q + \mathcal{E}$$

where $A(m)$ is the discretization of the differential operator and the boundary conditions, Q is a discretization of the sources, and P is a discretization of the receivers.

The ill-posedness of these problems is demonstrated in the following examples. We consider surface data for the DC resistivity and borehole data for the tomography (the exact setting is described in the last section of the paper). Upon using a smooth recovery (that is $R(m) = \int_\Omega |\nabla m|^2 \mathrm{d}x$), we obtain the results shown in Fig. 1. Clearly, the recovery of each of these models is far from perfect. While the conductivity recovery has very limited depth recovery, the tomography "smeared" the model horizontally. This should not come as a surprise; the conductivity is obtained from a DC resistivity survey that is sensitive to variations close to the surface, while the tomography is obtained by rays that are integrating the model laterally.

The question that is at the core of every inverse problem algorithm is how to improve these results. If no prior information is available, then one may claim that if the models are geologically reasonable and fit the data, then they are a sufficiently good solution to the problem. However, if more information exists, then the goal of the inversion algorithm is to somehow incorporate it into the inversion. While generic regularization techniques can

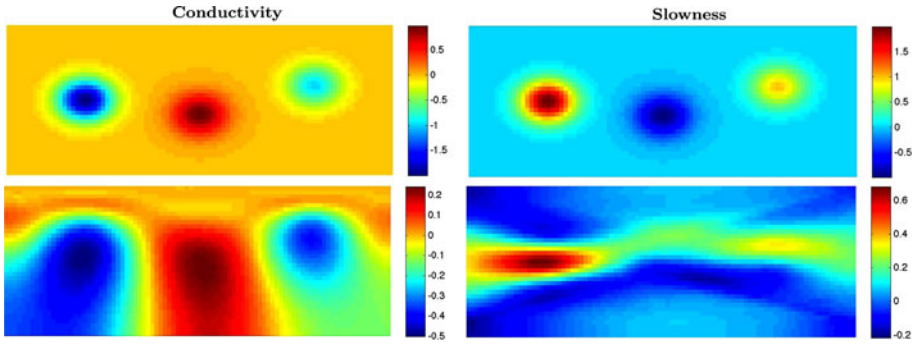

Fig. 1 True and smooth recovery of slowness and conductivity

do well for the incorporation of smoothness information, they fail to incorporate many types of a-priori information that is often available and needs to be represented.

Our simple example in Fig. 1 above is in fact one such case. It is clear that the slowness and the conductivity have similar "structure." Each has three anomalies located in the same area. Such information is natural if we consider that the change in physical properties represents a change in the underlying geology. The question of how to incorporate such structure into an inversion algorithm is at the heart of this paper.

Thus, in this work, we assume that we have structural information that needs to be respected. Two cases are to be considered.

1. In the simpler case, we assume the availability of a known function $s(x)$ such that the model, $m(x)$, has "similar structure" to $s(x)$. *We define the incorporation of information in a known model s(x) into the inversion of m(x) as* **model fusion**.
2. If the function $s(x)$ is unknown and is obtained by a different inverse problem of the form

$$y = G(s) + \epsilon \tag{3}$$

where $G(s)$ is a different forward modeling operator that is sensitive to changes in the parameter function $s(x)$, then we can attempt to invert for both m and s simultaneously in a process that is referred to as **joint inversion**.

The idea of model fusion commonly arises when two different imaging methods are used. For example, in geophysical imaging, seismic methods lead to structural information that needs to be respected by another imaging technique, say gravity. Similarly, in medical imaging, MRI or CT images can lead to structures that should be respected and incorporated by other imaging techniques with lower resolution such as proton emission tomography (PET). Generally, fusion and joint inversion are useful whenever both methods are sensitive to different parts of the model space.

Another common use of fusion and structure arises in geophysical imaging when geological sections and maps that are based on drilling and geological interpretation are given. Adding such information to the inversion process is difficult when using standard regularization techniques.

Joint inversion naturally arises when the model being used for structure is unknown; however, it shares many properties with the model fusion problem. In fact, as we show later in the paper, joint inversion can be achieved by iteratively considering model fusion.

The difficulty in fusing, merging, or jointly inverting m and s stems from the following three observations

- m and s have very different physical meaning (and units). As such, for many problems, we do not know an exact a priori relation between m and s, often referred to as a correspondence map (that is the map $m(s)$ or vice versa) or the petrophysics of the rocks.
- In many cases, when s involves a geological model, it does not have physical units and describes morphology.
- The structure of m and s can be similar at some regions but might be different at others. Thus, a method that can honor structure when possible is needed.

The idea of using a model function s to aid the recovery of m either in a static mode, that is, assuming known s or in a dynamic mode, that is, inverting for jointly s and m is not new. There are three main approaches for such integration

1. In some cases, petrophysical measurements that yield an empirical relation (correspondence map) between m and s are given. In this case, it is possible to incorporate the relations directly into the inversion algorithm of Roux et al. (2011) or Moorkamp et al. (2011). This approach is discussed in Sect. 2
2. In medical image registration, a gold standard to fuse two imaging modalities is Mutual Information (MI) (Pluim et al. 1999). MI can be viewed as a method to obtain the correspondence map directly from the data. We discuss the method in Sect. 3
3. In Haber and Oldenburg (1997) and Zhang and Morgan (1996), structure was defined by edges by using the zero crossings of second-order derivatives. This idea was substantially improved in Gallardo and Meju (2003), Gallardo and Meju (2004, 2011), Gallardo et al. (2012), Gallardo (2007), Jilinski et al. (2012), and Moorkamp et al. (2007) that defined structure using the cross products of the gradient fields. Similarly, other methods were based on variations of this method include Tryggvason and Linde (2006), Linde et al. (2008), Hu et al. (November–December 2009), and De Stefano et al. (2011). Other methods which depend on a model but not its edges include Lelivre and Oldenburg (2009) which uses a Bayesian approach and Cardiff and Kitanidis (2009) which depends on the distance from a reference model. We discuss using morphological structure, that is, derivatives, in Sect. 4 and introduce a new technique that we refer to as joint total variation in Sect. 5.

Throughout the paper, we discuss a number of new methodologies that can improve the results of the different techniques. In particular, in Sect. 5, we propose a new approach to the problem that uses a *convex* regularization term in both m and s. This new regularization enables us to solve the joint inversion problem using classical optimization techniques and has favorable numerical properties.

We discuss implementation details and suggest simple algorithms for the solution of the optimization problems that rise from the different formulations in Sect. 6. In Sect. 7, we perform numerical experiments on borehole tomography as well as the DC resistivity to demonstrate and compare the different approaches. Finally, in Sect. 8, we summarize the paper and discuss future research directions.

Throughout this paper, we move between continuous formulations where m is a function, to a discrete formulation where the function is discretized on a mesh. For brevity and simplicity of notations, we do not differentiate between the discrete and the continuous cases and assume that the reader can imply the nature of the variable from the context.

2 Inversion Through Correspondence Maps

In many cases, laboratory experiments are made in order to obtain an empirical relations between physical parameters. Such relations are often referred to as **petrophysics**. For example, a number of papers discuss empirical relations between seismic velocity and electrical conductivity (Jones et al. 2009) and seismic velocity and density (Barton 1986). Some theoretical relations such as Archie's (1942) law are also used to build such maps. Work that uses these maps is presented in Moorkamp et al. (2007, 2011), Zhang and Morgan (1996), and we summarize it below.

Assuming that the correspondence map is known exactly, it is simple to use it within the inversion process. In this case, one can write $m = m(s)$ and simply use this relation in the process of solving a single optimization problem, that is, no joint inversion is needed.

Note that fusion (that is inverting for m given s) does not make sense here and, thus, we discuss the joint inversion only. This can be done by simply setting

$$m = m(s)$$

and solving the optimization problem

$$\min_s \frac{1}{2}(F(m(s)) - d)^\top \sum_m^{-1}(F(m(s)) - d) + \frac{1}{2}(G(s) - y)^\top \sum_s^{-1}(G(s) - y) + \alpha R(s) \quad (4)$$

The covariance matrices Σ_m and Σ_s play an important role in the inversion process as they scale the two experiments correctly. Failing to use them may fit one inverse problem but not the other. The regularization parameter α needs to be estimated, and this can be done using "standard" techniques such as discrepancy principle or Generalized Cross Validation (Vogel 2001).

The idea of an exact correspondence map is not practical for most realistic cases. Consider the case of the generic correspondence map shown in Fig. 2, where direct data collected (typically laboratory data) about s are plotted against data collected about m.

The map (the red line) is obtained from fitting experimental data, and thus it is not certain. If one uses the map in Fig. 2, then this uncertainty is not taken into consideration and the resulting inversion can be erroneous.

To incorporate this uncertainty into the inversion, we assume that the map, C, is characterized by a few parameters, p, that is $C = C(s;p)$. For example, Archie's law suggests

$$\log(m) = p_1 - p_2 \log(s)$$

or, given m and s

$$A(s)p = b(m)$$

where

$$A(s) = \begin{pmatrix} 1 & -\log(s_1) \\ \vdots & \vdots \\ 1 & -\log(s_n) \end{pmatrix} \quad \text{and} \quad b(m) = \begin{pmatrix} \log(m_1) \\ \vdots \\ \log(m_n) \end{pmatrix}$$

and $p = [p_1 p_2]^\top$.

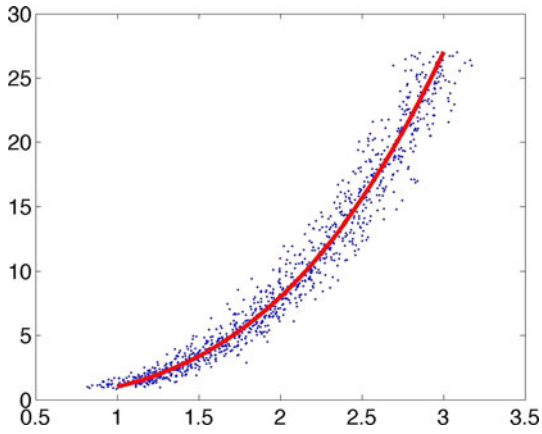

Fig. 2 A correspondence map between two physical properties and the fit obtained from the map

To recover the parameters p from laboratory measurements of m and s, in the first stage, we compute the correspondence map by solving the total least squares (TLS) problem (Van Huffel and Vandewalle 1991). The TLS assumes that both A and b contain noise, that is, the relations between the properties is

$$(A + \mathcal{E})p = b + \eta$$

where $\mathcal{E}$ and η are noise in the matrix and right-hand side, respectively. While it is possible to use regular least squares for the estimation of the correspondence map, it is not advisable. This is because both the matrix $A(s)$ and the right-hand side $b(m)$ depend on laboratory measurements that are noisy. After the correspondence map has been estimated, we obtain the parameter p and can write an approximate mathematical relation between m and s. Furthermore, we can estimate the residuals, that is

$$r_i = C(s_i; p) - m_i, \quad i = 1, \ldots, \#\text{ of samples}$$

and use the residuals to build the empirical variance

$$\Sigma_c = \frac{1}{N}\sum r_i^2.$$

Using this estimate, it is possible to add the relations into the inversion in terms of a another penalty and eliminate m from the inversion, obtaining a problem for a single model s and the noise in the map between m and s, r.

$$\begin{aligned}\min_{s,r} &\frac{1}{2}(F(C(s;p)+r)-d)^\top \sum_m^{-1}(F(C(s;p)+r)-d)\\ &+\frac{1}{2}(G(s)-y)^\top \sum_s^{-1}(G(s)-y)+\alpha_s R_s(s)+\frac{\gamma}{2}\|r\|^2\end{aligned} \tag{5}$$

The complexity of this inversion stems from the fact that we need to choose two different regularization parameters. The choice of γ can be done by using the discrepancy principle, where γ is chosen such that $\|r\| = \|C(s;p) - m\| \approx n\Sigma_c$ where n is the length of the m. Numerical strategies for the solution of the problem are discussed in Sect. 6.

One major pitfall of this approach is that the correspondence map is obtained from samples obtained from a very narrow geographical area and at very different scales. In general, the map may not be unique, that is, two values of m may be associated with a single value of s and vice versa. Furthermore, in many cases, the actual anomalies we are after are not modeled within the map. In this case, joint inversion may lead to wrongly biased results and it is difficult to detect that such errors have been made. One way to check the consistency of the results is to make sure that the m and s obtained through the inversion process are represented by the data. For example, if the laboratory data are limited to $0 \leq m \leq 1$, then are should be taken when applying the correspondence map to $1 \ll m$.

3 Inversion Through Mutual Information

Mutual information (MI) is the gold standard when comparing multi-modality images in medical image registration (Pluim et al. 1999; Viola 1995). Its roots are in information theory, and we shortly describe the method and its implementation.

The main idea of MI is to use the joint histogram of m and s in order to define a good similarity measurement. Since the classical definition of MI is in vector spaces, we assume that the inverse problem has been discretized and that m and s are vectors.

The mutual information distance function between the vectors m and s is defined by

$$MI(m,s) := H[\rho_\sigma(s)] + H[\rho_\sigma(m)] - H[\rho_\sigma(s,m)], \tag{6}$$

where the entropy H is defined by

$$H[\rho] := -\int \rho(t)\log(\rho(t))\mathrm{d}t, \tag{7}$$

$\rho_\sigma(s, m)$ is the estimated joint density function of the vectors (s, m), and $\rho_\sigma(s)$ and $\rho_\sigma(m)$ are the density functions of the two random variables s and m.

We thus have to evaluate the density function ρ_σ which depends on the discrete variables $s(\mathbf{x}_j)$ and $m(\mathbf{x}_j)$. Care must be taken when modeling the function ρ_σ, in particular, the function needs to be differentiable and compactly supported. A common way to estimate ρ and the associated MI function is to use the histogram method of Silverman (1986) setting

$$\rho_\sigma(s,m;g_1,g_2) := \sum_j K(\sigma, s(x_j) - g_1)K(\sigma, m(x_j) - g_2), \tag{8}$$

where K is a kernel function which integrates to one and σ is a parameter controlling the width of the kernel. The histograms are defined on the axis g_1,g_2 which are the dynamic values in s and m, respectively, that is, g_1 is a variable that spans the physical range of s, $\min(g_1) \leq \min(s)$ and $\max(g_1) \geq \max(s)$ and similarly, $\min(g_2) \leq \min(m)$ and $\max(g_2) \geq \max(m)$.

Note that ρ_σ is only an estimate of the true density ρ. The parameter σ is a smoothing parameter. For small σ, the Kernel resembles a delta function and for a larger one, it becomes smoother. It is possible to show that when the Kernel resembles a delta function, we implicitly assume that our sampling is infinite (Silverman 1986). For finite sampling, the density has to be smoothed and the value of σ should not be set to 0. To find the best smoothing parameter, various techniques have been suggested. Here, we use general cross validation methods (GCV) as presented in Silverman (1986).

Note that, in order to evaluate the integrals in (6) numerically, a further discretization of the joint density function $\rho_\sigma(g_1,g_2)$ is needed over the axis g_1 and g_2. For a detailed description about the discretization of MI and its implementation, see Modersitzki (2009).

After discretizing, the MI measure model fusion or joint inversion can be done in the "usual" way by solving the optimization problem

$$\begin{aligned}\min_{s,m} &\frac{1}{2}(F(m)-d)^\top \sum_m^{-1}(F(m)-d)\\ &+\frac{1}{2}(G(s)-y)^\top \sum_s^{-1}(G(s)-y) + \alpha_{MI}MI(m,s)\end{aligned} \tag{9}$$

While MI has very nice intuitive and theoretical properties, it suffers from being highly nonlinear. In fact, throughout our numerical experiments for joint inversion or fusion on synthetic examples, we had very little success in getting the method to avoid local minima. We believe that further research is needed in order to use this measure in a robust way in geophysical inverse problems.

4 Inversion Through Structure

In the previous sections, we used the idea of correspondence maps in order to "tie" two models with different physical properties. In this section, we discuss a different approach that has been successful in many applications. Rather than find a map, the idea is to use the topology of the models in order to connect them. Thus, we discuss the incorporation of so called "structural information" in inverse problems.

In order to use structure, we first need to define the meaning of structure similarity. Our exact definition is as follows:

Let m(x) and s(x) be continuously differentiable functions. Then we say that m(x) has the same structure as s(x) if the level sets of m(x) are parallel to the level sets of s(x).[2]

Recall that the level sets of a function are the contours (in 2D) and surfaces (in 3D) where the function is constant.

If the structure of m and s is approximately the same then one would like to measure the structure similarity and find models such that this measure is "small" by adding an appropriate regularization. To define such a measure, we recall that if the level sets are parallel then the normals to the level set are parallel. The (non-normalized) normals to the level sets are given by the gradient fields ∇m and ∇s. Using this observation, it was suggested in Gallardo and Meju (2003) to use the cross products of the gradient fields, that is, we define the similarity measure between m and s as

$$\mathcal{S}(m;s) = \frac{1}{2}\int_{\Omega} |\nabla m(x) \times \nabla s(x)|^2 \mathrm{d}x = \int_{\Omega} \frac{1}{2}|S_{\times}\nabla m|^2 \mathrm{d}x \tag{10}$$

where

$$S_{\times} = \begin{pmatrix} 0 & s_z & -s_y \\ -s_z & 0 & s_x \\ -s_x & s_y & 0 \end{pmatrix}.$$

The expression (10) involves cross derivatives of s and m and, therefore using it directly requires either low order accuracy or long differences that are notoriously unstable (Ascher 2010). This is because the vector $v = (1, -1, \ldots)^{\top}$ is in the null space of the long difference matrix, and thus, it does not penalize very high oscillations.

A mathematically equivalent expression that is easier to discretize is obtained by considering the dot product.

$$\mathcal{S}(m;s) = \int_{\Omega} |\nabla m(x)|^2 |\nabla s(x)|^2 - |\nabla m(x)^{\top} \nabla s(x)|^2 \mathrm{d}x \tag{11}$$

The equivalence between (11) and (10) corresponds to the vector identity

$$|\mathbf{a} \times \mathbf{b}|^2 = |\mathbf{a}|^2 ||\mathbf{b}|^2 - (|\mathbf{a}^{\top}\mathbf{b})^2$$

The dot product is zero for structures that are identical. The advantage of this expression is that it involves only normal derivatives products, and therefore, as we see next, it is easy to discretize it using short differences. Similar expressions to this one are studied in Jilinski et al. (2010) and Haber and Modersitzki (2006).

[2] Recall that the level sets of a function $m(x)$ are lines (in 2D) and surfaces (in 3D) that are defined implicitly by the equation $m(x) = \text{const}$.

Given $s(x)$ one can use $\mathcal{S}$ as a regularization operator and obtain a model that is close to a desired structure. Nonetheless, for many problems, this can be insufficient. Consider the case that $|\nabla s|$ is very small at some regions. An example for such a case is in our first tomographic problem in Fig. 1. The absolute gradient of the true model is plotted in Fig. 3 Note that there are many areas where $|\nabla m| \approx 0$ and, therefore, rough structure can be built in these areas without any penalty. This is demonstrated when we attempt to recover the model using the gradients obtained from the *true* model (that is never available in practice). The recovered model in Fig. 3 fits the data well but generate artifacts of high oscillations that are caused due to the lack of penalty and the long differences. As can be seen from this example, in this case, using $\mathcal{S}$ as regularization does not generally lead to a stable inversion algorithm. As a remedy, one can add a global regularization term such as smoothness to the problem in order to obtain sufficient regularization and avoid un-physical oscillations. In our experience, this does not yield a good regularization scheme in general, because global smoothness of the model often conflicts with the structural information (11) that allows non-smooth structures at some of the domain. We therefore propose the following regularization that stabilizes the problem in areas where $|\nabla s| \approx 0$.

$$R(m;s) = \int_\Omega |\nabla m(x)|^2 |\nabla s(x)|^2 - |\nabla m(x)^\top \nabla s(x)|^2 \mathrm{d}x + \frac{\beta}{2} \int_\Omega |\nabla m(x)|^2 \chi_\theta(|\nabla s(x)|) \mathrm{d}x \tag{12}$$

where β is the regularization coefficient $\chi_\theta(t)$ is a "smoothed" indicator function of the form

$$\chi_\theta(t) = \frac{1}{2}(\tanh(\theta t) + 1) \tag{13}$$

and the parameter θ is chosen based on the physical properties of ∇s. The idea is that $\theta\,|\nabla s|$ is large (say larger than 3) where $|\nabla s|$ is considered physically significant in its magnitude. The regularization (12) uses structure whenever possible and reverts to smoothing otherwise. Other functions can be used instead (e.g., total variation), but here, we concentrate on smoothness only.

When we have two different observations of the same structure, generated by two different physical experiments, we would like to jointly invert the data and thus reach a solution that fuses m with s. Consider the joint inversion problem:

$$F(m) + \epsilon_1 = d \quad \text{and} \quad G(s) + \epsilon_2 = y \tag{14}$$

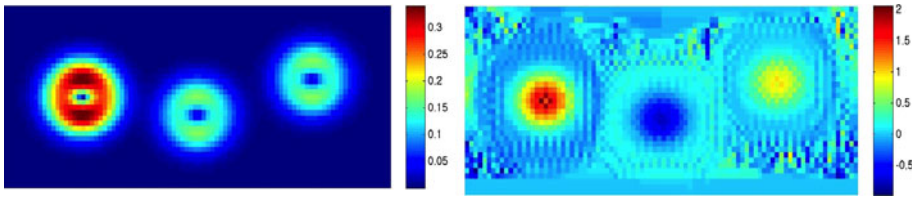

Fig. 3 The norm of the gradient $|\nabla m|$ for the true tomography model in Fig. 1 (*left*) and recovered slowness model using the unstabilized cross product (*right*). Note how the unregularized area in the model yields uncontrolled oscillations at the *bottom*

In this case, information that is contained in one model is relevant to the other model and vice versa. Therefore, the structures determined by one model can help identify structures in the other model, and the two models can correct each other in the joint inversion process, allowing a better solution for both. However, while we assume both models to have similar structures at similar locations, we do not impose that they share all the structures, and it is possible for one model to have a structure in a locations where the other model has none (e.g.$\|\nabla m\| \gg 0$ where $\|\nabla s\| = 0$).

We can obtain a solution to the problem using a block coordinate descent method where the model s is the structure for m, and the model m is the structure for s. This leads to the optimization problem

$$\begin{aligned}\min_{m,s} \mathcal{J}(m,s) = &\frac{1}{2}(F(m)-d)^\top \sum_m^{-1}(F(m)-d) + \frac{1}{2}(G(s)-y)^\top \sum_s^{-1}(G(s)-y) \\ &+ \alpha R(s;m) + \frac{\beta_1}{2}\int_\Omega |\nabla m(x)|^2 \chi_\theta(|\nabla s(x)|)\mathrm{d}x + \frac{\beta_2}{2}\int_\Omega |\nabla s(x)|^2 \chi_\theta(|\nabla m(x)|)\mathrm{d}x\end{aligned} \tag{15}$$

The problem involves many parameters to be determined, and this is discussed in Sect. 6.

We will now shortly discuss the discretization of the structure operator. First, assume that m is discretized at cell centers and approximate the first-order derivative using the usual short difference

$$\begin{aligned}(m_x)_{i+\frac{1}{2},j} &= \frac{1}{h}(m_{i+1,j} - m_{i,j}) + O(h^2) \\ (m_y)_{i,j+\frac{1}{2}} &= \frac{1}{h}(m_{i,j+1} - m_{i,j}) + O(h^2)\end{aligned} \tag{16}$$

The approximation leads to a staggered grid given in Fig. 4, where, m_x, m_y, and m are discretized in different locations.

Furthermore, we define averaging operators which interpolate edge quantities to cell centers.

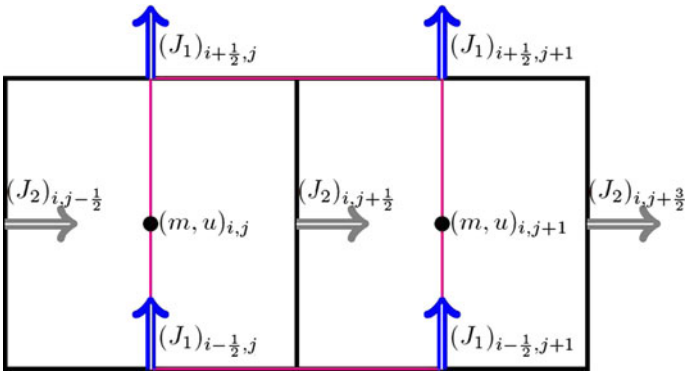

Fig. 4 Staggered grid in 2D

$$\begin{aligned} v_{i,j} &= \frac{1}{2}(v_{i+\frac{1}{2},j} + v_{i-\frac{1}{2},j}) + O(h^2) \\ v_{i,j} &= \frac{1}{2}(v_{i,j+\frac{1}{2}} + v_{i,j-\frac{1}{2}}) + O(h^2) \end{aligned} \tag{17}$$

The differences and averages can be combined into matrices. Let

$$D = \begin{bmatrix} D_x \\ D_y \end{bmatrix}$$

be the gradient matrix and similarly let

$$A_v = [A_x \quad A_y]$$

be the averaging matrix. Using these matrices, the structure operator can be approximated as

$$S_d(m;s) = \sum (A_v(Dm)^2) \odot (A_v(Ds)^2) - (A_v((Dm) \odot (Ds)))^2$$

where $\odot$ is the Hadamard product[3]. It is easy to verify that this is a second-order approximation to the continuous structure operator.

In order to use this operator in a numerical optimization scheme, we need to define the Jacobian and Hessian of this operator. It is straightforward to see that the Jacobian, $J_S = \nabla_m S_d$ and Hessian, $H_S = \nabla_m^2 S_d$ of the discretized structure operator are

$$\begin{aligned} J &= diag(A_v(Ds \odot Ds))A_v diag(Dm)D - diag(A_v((Dm) \odot (Ds)))A_v diag(Ds)D \\ H_S &= D^\top \left(diag\left(A_v^\top A_v (Ds)^2\right) - diag(Ds)A_v^\top A_v diag(Ds)\right) D \end{aligned} \tag{18}$$

5 Inversion Through Joint Total Variation

In the previous section, we have discussed a regularization term based on the geometry of structure. In this section, we use ideas from joint sparse recovery of van den Berg and Friedlander (2010) in order to achieve similar goals.

To do that, we use the notion of a mixed $\ell_{1,2}$ norm. Let X be a $n \times 2$ matrix, we define the norm $\|X\|_{1,2}$ as

$$\|X\|_{1,2} = \sqrt{X_{\downarrow,1}^2 + X_{\downarrow,2}^2}$$

where $X_{\downarrow,i}$ is the ith column of X and

$$X_{\downarrow,i}^2 = X_{\downarrow,i} \odot X_{\downarrow,i}$$

Assume that each column of X is sparse, that is, assume that each column of X has only a few non-zero entries. Then, it is straightforward to verify that the $\ell_{1,2}$ norm of X is smaller if the non-zero entries correspond to the same row of X. This observation is in the core of the ideas of joint recovery of sparse signals that appeared in the work of van den Berg and Friedlander (2010).

[3] Recall that the Hadamard product of a vector y with a vector z is defined as $(z \odot y)_i = z_i y_i$

Here, we use the idea for model fusion and joint inversion. by extending the definition of the $\ell_{1,2}$ to continuous variables. Assume that we have two models m and s and let $|\nabla m|$ and $|\nabla s|$ be the absolute value of their gradients. We now define the joint total variation of the two models as

$$JTV(m,s) = \int_{\Omega} \sqrt{|\nabla m|^2 + |\nabla s|^2}\mathrm{d}x \tag{19}$$

Note that the models are coupled through the square root function and, as can be observed below, the gradients are completely coupled.

We can use this definition for joint inversion by solving the optimization problem

$$\begin{aligned}\min_{m,s} \mathcal{J}(m,s) = &\frac{1}{2}(F(m)-d)^{\top}\sum_{m}^{-1}(F(m)-d) + \frac{1}{2}(G(s)-y)^{\top}\sum_{s}^{-1}(G(s)-y)) \\ &+ \alpha JTV(m,s)\end{aligned} \tag{20}$$

As previously discussed, for model fusion, we fix the model s and invert for m alone, while for joint inversion, we solve the optimization problem for both m and s.

We will now briefly discuss the discretization of the joint total variation (JTV) regularization. Consider the gradient and average matrices D and A_v introduced in the previous section. Then, the absolute value of the gradient can be discretized as

$$\|\nabla m\| \approx \sqrt{A_v|Dm|^2} \qquad \|\nabla s\| \approx \sqrt{A_v|Ds|^2}$$

where $|t|^2 = t \odot t$. This implies that

$$JTV(m,s) \approx JTV^h(m,s) = v^{\top}\sqrt{A_v|Dm|^2 + A_v|Ds|^2 + \epsilon} \tag{21}$$

where v is a vector of the volume of each voxel in the models and ϵ is a small number that is added to address the lack of differentiability at 0 (see Ascher et al. 2005). The derivatives of $JTV^h(m, s)$ are

$$\nabla_m JTV^h(m,s) = D^{\top}\left(\frac{A_v^{\top}v}{(A_v|Dm|^2 + A_v|Ds|^2 + \epsilon)^{\frac{1}{2}}}\right)Dm$$

$$\nabla_s JTV^h(m,s) = D^{\top}\left(\frac{A_v^{\top}v}{(A_v|Dm|^2 + A_v|Ds|^2 + \epsilon)^{\frac{1}{2}}}\right)Ds$$

Inversion through JTV has one remarkable property that is unique in comparison with other joint inversion methodologies. While all previously introduced methods are, in general, nonconvex, the JTV term is convex in both m and s. This allows for much better optimization algorithms and for more robust convergence results. In particular, if both forward models are convex then the complete optimization problem is convex, and a unique solution is obtained independent of the starting point.

6 Numerical Solution of the Optimization Problems

We have introduced four different model fusion or joint inversion formulations. The first (5) is based on correspondence maps and is noted as CM, the second, (9), is based on

Mutual Information and is noted MI, the third, (15), is based on structure and is noted as SI, and the last (20) is based on joint total variation and is referred to as JTV. All four formulations lead to large scale optimization problems. In this section, we briefly review how such optimization problems are solved and discuss some of the practical details that arise when numerically solving these problems. We do not intend to give a full numerical treatment to the problem and point to text-books such as Parker (1994), Vogel (2001), and Hansen (1997) that review basic treatment of inverse problems.

6.1 Numerical Optimization for Model Fusion

Recall that, in model fusion, the model s is assumed known and m is unknown. In this case, the optimization problems have the common form

$$\min_m \mathcal{J}(m) = misfit(m) + \alpha\, regularization(m; s)$$

where the data misfit is the difference between the observed and predicted data and the regularization term involves the known parameter function s. The problem can be viewed as a "standard" inverse problem with a different regularization term, and thus, common strategies for the solution of inverse problems directly apply.

The problem can be solved using some descent method, where each iteration requires four main steps.

1. Computation of $\mathcal{J}(m)$ and $\nabla_m \mathcal{J}(m)$
2. Approximation of the Hessian $H \approx \nabla_m^2 \mathcal{J}$
3. Computation of a step that solves the linear system $H\delta m = -\nabla_m \mathcal{J}(m)$
4. (Safe) update of the model $m \leftarrow m + \mu\delta m$ where μ is a line search parameter.

Roughly speaking, different algorithms differ in their approximation to the Hessian $\nabla_m^2 \mathcal{J}$. In our experiments, we use the limited memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) and the Gauss–Newton method (Nocedal and Wright 1999). We have found that these two methods perform well for most geophysical inverse problems. Gauss–Newton is more computationally involved, but it tends to converge faster. The regularization parameter can be tuned such that the data fit to the discrepancy principle (Tikhonov and Arsenin 1977) or generalized cross-validation (GCV) (Haber and Oldenburg 2000; Golub et al. 1979).

It is important to differentiate between the structural approaches, SI and JTV and the others, namely CM and MI. The remarkable feature for SI is that it leads to a quadratic regularization and JTV leads to convex regularization. Therefore, for convex forward problems, we obtain a convex inverse problem that is solved using very few iterations. In fact, solving the optimization problem does not impose any additional difficulties when comparing to other simple regularization techniques, such as smoothness or smallness. On the other hand, MI and, in some cases, CM can be highly nonlinear and often nonconvex with multiple minima. Thus, model fusion can be easily performed using SI and JTV but requires special attention when using other methods.

When using SI, it is important to use some other regularization in space in locations where $\|\nabla s\| \approx 0$. In our experience, if the extra regularization is not added to the problem then the overall optimization problem is under-regularized and is often still ill-posed, and this can lead to unreasonable solutions. JTV on the other hand does not require any additional terms and can be used automatically even if the gradients of s vanish.

Nonetheless, although SI has clear advantages from a computational approach, using a map between m and s, and assuming that the relationship is well known and close to exact,

will inject much more information into the problem. In this case, the problem can be solved obtaining a feasible "geological" model that is often of better quality compared with any other method. If on the other hand, the relations between m and s are highly inaccurate then the injection of such information can be disastrous.

Finally, in our attempts to work with MI, we have found it to be highly nonlinear and nonconvex. While we managed to solve some individual problems by "tweaking" parameters, we did not manage to obtain a sufficiently robust code that works well in general. We believe that using MI for model fusion and joint inversion remains an interesting area of research.

6.2 Numerical Optimization for Joint Inversion

In joint inversion, we jointly invert for m and s by solving the generic optimization problem

$$\min_{m,s} \mathcal{J}(m,s) = misfit_m(m) + misfit_s(s) + \alpha\, regularization(m,s)$$

where $misfit_m$ and $misfit_s$ are the misfits for the parameters m and s, respectively. The difficulty here stems from the fact that inverting for both models simultaneously requires access to the computational components of each problem, and the ability to integrate them. Since the components may have very different scales the coupling requires new codes that jointly solve the problem. Our approach has been to utilize existing codes for the solution of the problem by making small changes that allow for joint inversion.

To this end, we use a block coordinate descent method. While block descent may be slow, it separates the joint inversion problem into a number of model fusion problems that are easier to solve. A general joint inversion algorithm is as follows, at iteration k

1. Approximately solve the optimization problem

$$m_k = \arg\min_m misfit_m(m) + \alpha\, regularization(m, s_{k-1})$$

2. Approximately solve the optimization problem

$$s_k = \arg\min_s misfit_s(s) + \alpha\, regularization(m_k, s)$$

At this point, we stress that working one model at a time performs actual joint inversion. It is just an algorithmic choice and in the absence of local minima leads to the same answer as an algorithm that simultaneously inverts for m and s.

The advantage of the block descent is that, as previously discussed, each of these problems can be solved using "standard" inversion algorithms and does not require special machinery. This makes the solution of joint inversion problems simple, if the fusion problem can be easily solved. In particular, for SI and JTV joint inversion can be achieved with ease since each optimization problem involves a convex regularization term. Using a coordinate descent has another added advantage. In many if not most cases, inversion codes are designed to obtain an individual regularization parameter. Thus, when using coordinate descent, it is possible to use existing algorithms for the estimation of regularization parameters.

7 Numerical Experiments

We now experiment with the different algorithms on our model problems. We start by assuming that one model is known which leads to model fusion and then experiment with the case that both models are unknown which leads to joint inversion.

7.1 Numerical Experiments for Model Fusion

In the first part of our experiments, we experiment with the different techniques on model fusion, where we assume that s is known and m is unknown. In our first experiment, we use the borehole tomography problem. To this end, we simulate a data set of 32 sources and 32 receivers that are spaced 96 m apart. Each source (or receiver) is spaced one meter away from its adjacent neighboring sources. All receivers share all sources and the data is polluted with 1 % noise. The model is discretized on a mesh size 32×96.

The true model corresponds to the structural information in two structures and has an additional new smooth anomaly. When inverting the data using smoothness (that is, regularizing by the first derivative), we obtain a smeared model where the middle anomaly is not well represented. When using both SI and JTV, we obtain a clear improvement where all anomalies are well represented. Finally, when using MI we obtain a result that is only marginally better than smoothness. From a numerical perspective, the SI yields a quadratic problem that is solved in one iteration while the JTV requires a few iterations. MI is very unstable and the result shown here was obtained by manually changing parameters to keep the method stable. The results clearly demonstrate the effectiveness of JTV and SI for model fusion; see Fig. 5.

In our next experiment, we repeat the model fusion experiment but with the DC problem. The results are shown in Fig. 6. DC resistivity has a limited resolution at depth, and this is clearly seen in the smooth inversion, where the anomalies are smeared toward depth. Using SI and JTV yield slightly different results but, overall, both known structures are well represented while the unknown structure is recovered with some depth smearing. Model fusion using both SI and JTV works well for this problem as well. On the other hand, MI did not converge to any reasonable result (and thus we avoid showing the result). This problem clearly demonstrates the power of convex structural regularizers over the nonconvex ones.

7.2 Joint Inversion

In this section, we demonstrate our experiments of joint inversion techniques. Here, we compare between CM (Correspondence Maps) where the maps are obtained from noisy data, SI, and JTV. Our attempts to use MI fail for this experiment as well.

For the experiment, we generate a slowness model and use the simple relation

$$m = \log(\sigma) = -s$$

to define both slowness and conductivity. We then recover the slowness and log conductivity models using the different techniques. The different inversions are plotted in Fig. 8. All three variants experimented with here gave a substantial improvement over individual inversions, which indicates the usefulness of joint inversion.

When analyzing the results, it is important to realize that in SI and JTV we do not use the direct expression that links between m and s but rather assume that the models have

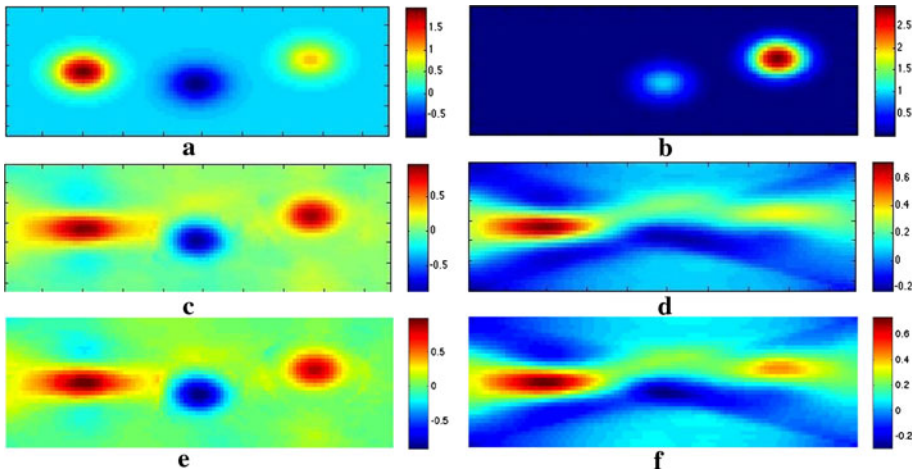

Fig. 5 Numerical experiment for model fusion of slowness. **a** True slowness model, **b** structural information, **c** structural inversion using SI, **d** smooth inversion without structural information, **e** structural inversion using JTV, **f** structural inversion using MI

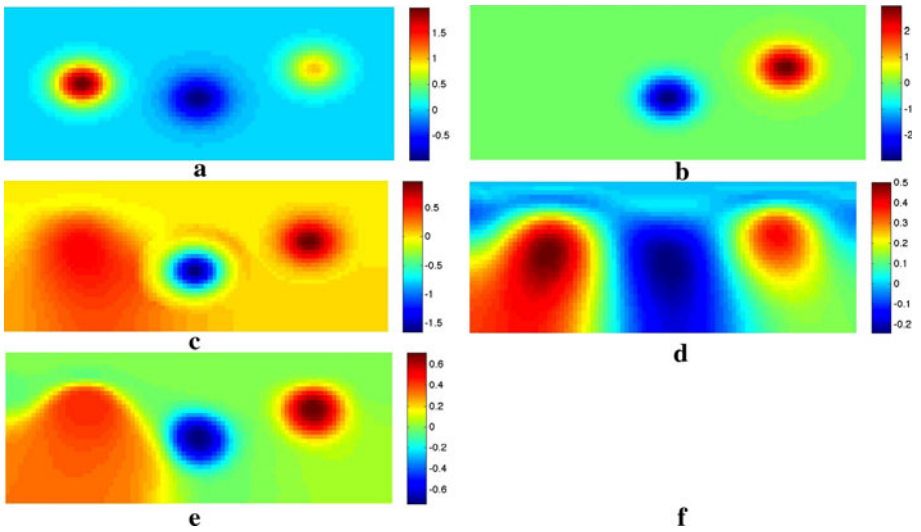

Fig. 6 Numerical experiment for model fusion of conductivity. **a** True slowness model, **b** structural information, **c** structural inversion using SI, **d** smooth inversion without structural information **e** structural inversion using JTV, **f** MI did not converge for this problem

similar structure. In CM we assume to have additional 129 noisy laboratory measurements of m and s. The relations between m and s are then approximated from these measurements by total least squares. Since all measurements contain noise, we obtain only an approximate relation that contains errors. The data used to obtain the m-s curve are plotted in Fig. 7. Here, we assume a linear relation $m = a\ s + b$ and compute a and b from the curves. As can be seen, for data that has only 5 % noise, we obtain a fairly accurate estimation while for the data polluted with 20 % noise, the estimation has substantial

errors. We then use the estimated a and b to jointly invert the DC and tomography problems. We observe that when the errors are significant, the quality of the joint inversion deteriorates significantly.

We record the mean square error (MSE) in the recovery for each of the methods in Table 1. For the problems solved here, SI was superior to other joint inversion modalities. This can be visibly observed in the results as well as in the MSE of the recovery.

From an optimization point of view, JTV was the most robust and (as theoretically guaranteed) converged to the same point independent of the starting model. We have found that, when using SI, care must be taken to start from a reasonable model and to add sufficient smoothing in areas where the gradient of the model at the current iteration is small. We have found that if this is not dealt with then SI can be highly unstable and may give erroneous results.

In Table 1, we also show the number of iterations that each method performed to achieve its minimum. The correspondence map is the least expensive method compared to any other inversion. This is not surprising as the map used here is not very nonlinear. As expected from the theory, JTV that is convex did better (in terms of convergence) and SI was the most expensive.

8 Conclusions and Future Work

In this paper, we presented four methods for joint inversion. These methods are based on either correspondence maps or structure. Three new techniques have been introduced in the paper. Mutual information (MI) which performed poorly in all our experiments, a structural inversion based on the dot product of the gradients, which is similar to the well used cross-gradient product, and a new methodology of joint total variation reconstruction. We have introduced the concept of model fusion, when a new model is fused to a known model. We also explored joint inversion when the two models are jointly inverted. For model fusion, using the structural information and joint total variation leads to a convex regularization which is a significant advantage over other possible techniques. In particular, the structural information measure leads to a quadratic problem that can be solved using standard software packages.

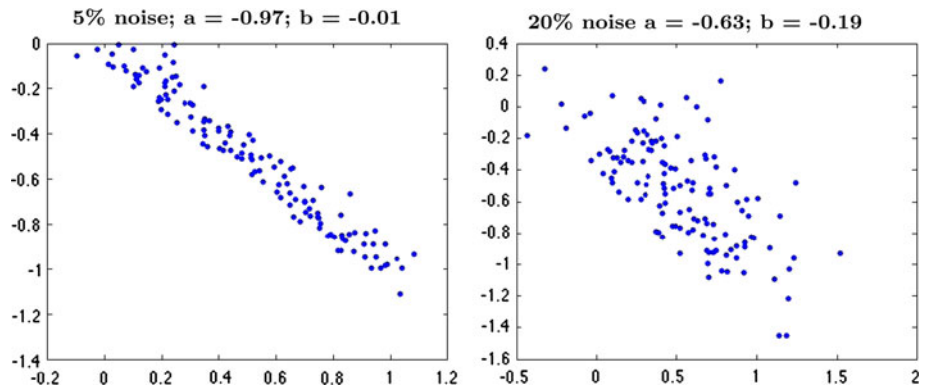

Fig. 7 Simulated laboratory s versus m data with various degrees of noise

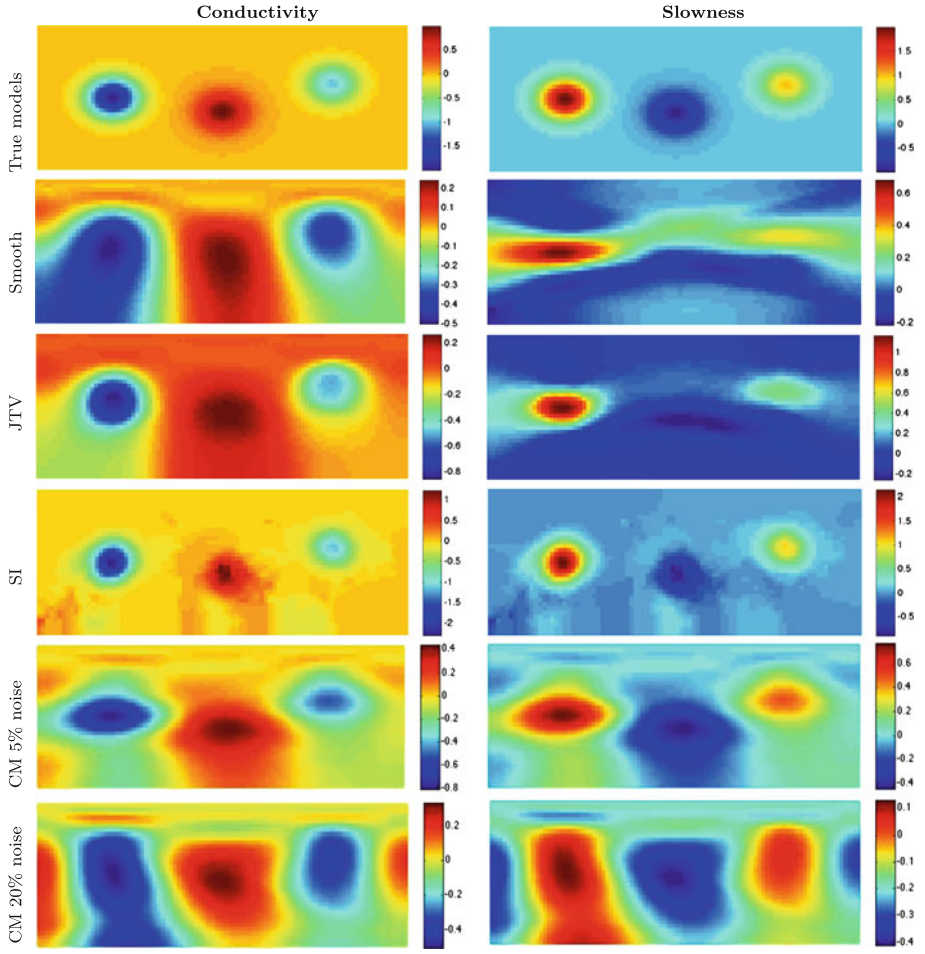

Fig. 8 True, smooth recovery, JTV recovery, SI recovery and CM recovery

Table 1 Number of iterations for each joint inversion technique

Method	Iterations	DC MSE (%)	Tomography MSE (%)
Smooth recovery	1	63	45
JTV	16	36	29
SI	45	28	22
CM 5 % noise	8	49	34
CM 20 % noise	7	61	41

The problem of joint inversion is significantly more complex. Correspondence maps are very much problem dependent and depend on available laboratory information and its extrapolation to field data sets. On the other hand, structure can be easily used whenever we believe that the structure of two models fully or partially align. We have explored two methods to incorporate structural information. We propose a new method, the joint total

variation, that uses a convex functional for joint inversion. This functional is the only convex functional known to us that can achieve this goal, and thus it has significant advantages over other regularization techniques, in particular for nonlinear problems.

We have also explored optimization techniques for the solution of the problems. For joint inversion, we advocate the use of block coordinate descent where one model is fixed and the other is fused to it, changing the models at each sub-iteration. This approach has a significant advantage as it can use existing optimization packages changing only the regularization term.

Finally, the techniques developed and discussed in this paper are applicable to a wide range of problems, and we expect that model fusion and joint inversion will be often used in practice whenever data is available.

Acknowledgments The authors would like to thank Adam Pidlisecky for the 2D DC code Pidlisecky and Knight (2008).

References

Archie G (1942) The electrical resistivity log as an aid in determining some reservoir characteristics. Petroleum Trans AIME 146:54–62

Ascher U (2010) Numerical methods for evolutionary PDE's. SIAM, Philadelphia

Ascher U, Haber E, Haung H (2005) On effective methods for implicit piecewise smooth surface recovery. SIAM J Sci Comput 28:339–397

Barton PJ (1986) The relationship between seismic velocity and density in the continental crust a useful constraint? Geophys J R Astronom Soc 87:195–20

van den Berg E, Friedlander M (2010) Joint-sparse recovery from multiple measurements. IEEE Trans Info Theory 56(5):2516–2527

Cardiff M, Kitanidis PK (2009) Bayesian inversion for facies detection: An extensible level set framework. Water Resour Res 45(10). doi:10.1029/2008WR007675

De Stefano M, Andreasi FG, Re S, Virgilio M, Snyder F (2011) Multiple-domain, simultaneous joint inversion of geophysical data with application to subsalt imaging. Geophysics 76(3):R69–R80. doi:10.1190/1.3554652

Gallardo L (2007) Multiple cross-gradient joint inversion for geospectral imaging. Geophys Res Lett 34:L19301

Gallardo L, Fontes S, Meju M, Buonora MP, de Lugao P (2012) Robust geophysical integration through structure-coupled joint inversion and multispectral fusion of seismic reflection, magnetotelluric, magnetic and gravity images: example from Santos Basin, offshore Brazil. Geophysics 77:B237–B251

Gallardo LA, Meju MA (2003) Characterization of heterogeneous near-surface materials by joint 2d inversion of dc resistivity and seismic data. Geophys Res Lett 30(13):1658–1664

Gallardo LA, Meju MA (2004) Joint two-dimensional dc resistivity and seismic traveltime inversion with cross-gradients constraints. J Geophys Res 109B:3311–3315

Gallardo LA, Meju MA (2011) Structure-coupled multiphysics imaging in geophysical sciences. Rev Geophys 49(1). doi:10.1029/2010RG000330

Golub G, Heath M, Wahba G (1979) Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21:215–223

Haber E, Modersitzki J (2006) Intensity gradient based registration and fusion of multi-modal images. In: Proceeding of MICCAI, pp 323–330

Haber E, Oldenburg D (1997) Joint inversion a structural approach. Inverse Prob 13:63–67

Haber E, Oldenburg D (2000) A GCV based methods for nonlinear inverse problem. Comput Geosci 4, n1

Hansen PC (1997) Rank-deficient and discrete ill-posed problems. SIAM, Philadelphia

Hu W, Abubakar A, Habashy TM (November-December 2009) Joint electromagnetic and seismic inversion using structural constraints. Geophysics 74(6):R99–R109

Jackson D (1979) The use of a priori data to resolve non-uniqueness in linear inversion. Geophys J R Astronom Soc 57:137–157

Jilinski P, Fontes S, Gallardo L (2010) Joint interpretation of maps using gradient directions, cross and dot-product values to determine correlations between bathymetric and gravity anomaly maps. SEG Expand Abstr 29:1226–1229

Jilinski P, Fontes S, Meju A (2012) Estimating optimum density for regional Bouguer reduction by morphological correlation of gravity and bathymetric maps: examples from offshore southeastern Brazil. Geo Marine Lett 1–7

Jones AG, Fishwick S, Evans RL, Team TS (2009) Correlation of lithospheric velocity and electrical conductivity for Southern Africa. In: 11th SAGA biennial technical meeting and exhibition, Swaziland 16–18 September, pp 428 – 434

Lelivre PG, Oldenburg DW (2009) A comprehensive study of including structural orientation information in geophysical inversions. Geophys J Int 178(2):623–637. doi:10.1111/j.1365-246X.2009.04188.x

Linde N, Tryggvason A, Peterson JE, Hubbard S (2008) Joint inversion of crosshole radar and seismic traveltimes acquired at the south oyster bacterial transport site. Geophysics 73(4):G29–G37. doi:10.1190/1.2937467

Modersitzki J (2009) FAIR: flexible algorithms for image registration. SIAM, Philadelphia

Moorkamp M, Jones A, Eaton D (2007) Joint inversion of teleseismic receiver functions and magnetotelluric data using a genetic algorithm: are seismic velocities and electrical conductivities compatible? Geophys Res Lett 34:L16311

Moorkamp M, Jones A, Fishwick S (2011) Joint inversion of receiver functions, surface wave dispersion and magnetotelluric data. J Geophys Res Solid Earth 115:5–6

Nocedal J, Wright S (1999) Numerical optimization. Springer, New York

Parker RL (1994) Geophysical inverse theory. Princeton University Press, Princeton, NJ

Pidlisecky A, Knight R (2008) A rapid 2.5 forward modelling algorithm for electrical resistivity modelling. Comput Geosci 34(12):1645–1654

Pluim J, Maintz J, Viergever M (1999) Mutual-information-based registration of medical images: a survey. IEEE Trans Med Imaging 22:986–1004

Roux E, Moorkamp M, Jones A, Bischoff M, Endrun B, Levedev S, Meier T (2011) Joint inversion of long-period magnetotelluric data and surface-wave dispersion curves for anisotropic structure: application to data from Central Germany. Geophys Res Lett 38:5–6

Silverman BW (1986) Density estimation. Chapman and Hall, London

Tikhonov A, Arsenin V (1977) Methods for solving ill-posed problems. Wiley, New York

Tryggvason A, Linde N (2006) Local earthquake (le) tomography with joint inversion for p- and s-wave velocities using structural constraints. Geophys Res Lett L07303

Van Huffel S, Vandewalle J (1991) The total least squares problem: computational aspects and analysis. Frontiers in applied mathematics, Society for Industrial and Applied Mathematics

Viola PA (1995) Alignment by maximization of mutual information. PhD thesis, Massachusetts Institute of Technology

Vogel C (2001) Computational methods for inverse problem. SIAM, Philadelphia

Zhang J, Morgan FD (1996) Joint seismic and electrical tomography. In: Proceedings of EEGS symposium on applications of geophysics to engineering and environmental problems keystone, Colorado, pp 391–396