Data completion and stochastic algorithms for PDE inversion problems with many measurements

Farbod Roosta-Khorasani, Kees van den Doel and Uri Ascher *

October 9, 2013

Abstract

Inverse problems involving systems of partial differential equations (PDEs) with many measurements or experiments can be very expensive to solve numerically. In a recent paper we examined dimensionality reduction methods, both stochastic and deterministic, to reduce this computational burden, assuming that all experiments share the same set of receivers.

In the present article we consider the more general and practically important case where receivers are not shared across experiments. We propose a data completion approach to alleviate this problem. This is done by means of an approximation using an appropriately restricted gradient or Laplacian regularization, extending existing data for each experiment to the union of all receiver locations. Results using the method of simultaneous sources (SS) with the completed data are then compared to those obtained by a more general but slower random subset (RS) method which requires no modifications.

1 Introduction

The reconstruction of distributed parameter functions, by fitting to measured data solution values of partial differential equation (PDE) systems in which they appear as material properties, can be very expensive to carry out. This is so especially in cases where there are many experiments, where just one evaluation of the forward operator can involve hundreds and thousands of PDE solves. And yet, there are several such problems of intense current interest in which the use of many experiments is crucial for obtaining credible reconstructions in practical situations [22, 10, 15, 12, 18, 25, 31, 30, 24, 7, 5]. Extensive theory (e.g., [23, 3, 2]) also suggests that many well-placed

experiments are often a practical must for obtaining credible reconstructions. Thus, methods to alleviate the resulting computational burden are highly sought after.

To be more specific, consider the problem of recovering a model $\mathbf{m} \in \mathbb{R}^{l_m}$, representing a discretization of a surface function $m(\mathbf{x})$ in 2D or 3D, from measurements $\mathbf{d}_i \in \mathbb{R}^l$, $i = 1, 2, \ldots, s$.[1] For each i, the data is predicted as a function of $\mathbf{m}$ by a forward operator $\mathbf{F}_i$, and the goal is to find (or infer) $\mathbf{m} = \mathbf{m}^*$ such that the misfit function

$$\phi(\mathbf{m}) = \sum_{i=1}^{s} \|\mathbf{F}_i(\mathbf{m}) - \mathbf{d}_i\|^2 \tag{1}$$

is roughly at a level commensurate with the noise.[2] The forward operator involves an approximate solution of a PDE system, which we write in discretized form as

$$A(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i, \quad i = 1, \ldots, s, \tag{2a}$$

where $\mathbf{u}_i \in \mathbb{R}^{l_u}$ is the ith field, $\mathbf{q}_i \in \mathbb{R}^{l_u}$ is the ith source, and A is a square matrix discretizing the PDE plus appropriate side conditions. Furthermore, there are given projection matrices P_i such that

$$\mathbf{F}_i(\mathbf{m}) = P_i\mathbf{u}_i = P_iA(\mathbf{m})^{-1}\mathbf{q}_i \tag{2b}$$

predicts the ith data set. Thus, evaluating $\mathbf{F}_i$ requires a PDE system solve, and evaluating the objective function $\phi(\mathbf{m})$ requires s PDE system solves.

For reducing the cost of evaluating (1), stochastic approximations are natural. Thus, introducing a random vector $\mathbf{w} = (w_1, \ldots, w_s)^T$ from a probability distribution satisfying

$$\mathbb{E}(\mathbf{w}\mathbf{w}^T) = I \tag{3}$$

(with $\mathbb{E}$ denoting the expected value with respect to $\mathbf{w}$ and I the $s \times s$ identity matrix), we can write (1) as

$$\phi(\mathbf{m}) = \mathbb{E}\left(\|\sum_{i=1}^{s} w_i(\mathbf{F}_i(\mathbf{m}) - \mathbf{d}_i)\|^2\right), \tag{4}$$

and approximate the expected value by a few samples $\mathbf{w}$ [1]. If, furthermore, the data sets in different experiments are measured at the same locations, i.e., $P_i = P \ \forall i$, then

$$\sum_{i=1}^{s} w_i\mathbf{F}_i = \sum_{i=1}^{s} w_iP_iA(m)^{-1}\mathbf{q}_i = PA(m)^{-1}\left(\sum_{i=1}^{s} w_i\mathbf{q}_i\right), \tag{5}$$

---

[1]For notational simplicity we make the nonessential assumption that l does not depend on the experiment i.

[2] Throughout this article we use the ℓ_2 vector norm unless otherwise specified.

which can be computed with a single PDE solve per realization of the weight vector $\mathbf{w}$, so a very effective procedure for approximating the objective function $\phi(\mathbf{m})$ is obtained [17].

Next, in an iterative process for reducing (1) sufficiently, consider approximating the expectation value at iteration n by random sampling from a set of s_n vectors $\mathbf{w}$, with $s_n \leq s$ potentially satisfying $s_n \ll s$; see, e.g., [28, 20, 14]. Several recent papers have proposed methods to control the size s_n [9, 13, 6, 27]. Let us now concentrate on one such iteration n, for which a specialized Gauss-Newton (GN) or L-BFGS method may be employed. We can write (1) using the Frobenius norm $\| \cdot \|_F$ as

$$
\begin{aligned}
\phi(\mathbf{m}) &= \|F(\mathbf{m}) - D\|_F^2, \\
F &= \left[\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_s\right] \in \mathbb{R}^{l \times s}, \quad D = \left[\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_s\right] \in \mathbb{R}^{l \times s},
\end{aligned}
\tag{6}
$$

and hence, an unbiased estimator of $\phi(\mathbf{m})$ in the nth iteration is

$$
\hat{\phi}(\mathbf{m}, W) = \frac{1}{s_n} \|(F(\mathbf{m}) - D)W\|_F^2,
\tag{7}
$$

where $W = W_n = \left[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{s_n}\right]$ is an $s \times s_n$ matrix with $\mathbf{w}_j$'s drawn from any distribution satisfying (3). For the case where $P_i = P \; \forall i$, different methods of simultaneous sources (SS) are obtained by using different algorithms for this *model and data reduction* process [4, 33, 26]. In [27] we have discussed and compared three such methods: (i) a Hutchinson random sampling, (ii) a Gaussian random sampling, and (iii) the deterministic truncated singular value decomposition (TSVD). We have found that, upon applying these methods to the famous EIT/DC-resistivity problem, their performance was roughly comparable (although for just estimating the misfit function by (7), only the stochastic methods work well).

A fourth, random subset (RS) method was considered in [27, 9], where a random subset of the original experiments is selected at each iteration n. This method does not require that the receivers be shared among different experiments. However, its performance was found to be generally worse than the methods of simultaneous sources, roughly by a factor between 1 and 4, and on average about 2.[3] This brings us to the quest of the present article, namely, to seek methods for the general case where P_i does depend on i, which are as efficient as the simultaneous sources methods. The tool employed for this is to "fill in missing data", thus replacing P_i, for each i, by a common projection matrix P to the union of all receiver locations, $i = 1, \ldots, s$.

The prospect of such *data completion*, like that of casting a set of false teeth based on a few genuine ones, is not necessarily appealing, but is often necessary for reasons of computational efficiency. Moreover, applied mathematicians do a virtual data completion automatically when considering a Dirichlet-to-Neumann map, for

---

[3] The relative efficiency factor further increases if a less conservative criterion is used for algorithm termination, see Section 4.

instance, because such maps assume knowledge of the field u (see, e.g., (11) below) or its normal derivative on the entire spatial domain boundary, or at least on a partial but continuous segment of it. Such knowledge of noiseless data at uncountably many locations is never the case in practice, where receivers are discretely located and some noise, including data measurement noise, is unavoidable. On the other hand, it can be argued that any practical data completion must inherently destroy some of the "integrity" of the statistical modeling underlying, for instance, the choice of iteration stopping criterion, because the resulting "generated noise" at the false points is not statistically independent of the genuine ones where data was collected.

Indeed, the problem of proper data completion is far from being a trivial one, and its inherent difficulties are often overlooked by practitioners. In this article we consider this problem in the context of the EIT/DC-resistivity problem (Section 2.3), with the sources and receivers for each data set located at the boundary $\partial\Omega$ of the domain on which the forward PDE is defined. Our data completion approach is to approximate or interpolate the given data directly in smooth segments of the boundary, while taking advantage of prior knowledge as to how the fields $\mathbf{u}_i$ must behave there.

We have tested several regularized approximations on the set of examples of Section 4, including several DCT, wavelet and curvelet approximations (for which we had hoped to leverage the recent advances in compressive sensing and sparse ℓ_1 methods) as well as straightforward piecewise linear data interpolation. However, the latter is well-known not to be robust against noise, while the former methods are not suitable in the present context, as they are not built to best take advantage of the known solution properties. The methods which proved winners in the experimentation ultimately use a Tikhonov-type regularization in the context of our approximation, penalizing the discretized L_2 integral norm of the gradient or Laplacian of the fields restricted to the boundary segment surface. They are further described and theoretically justified in Section 3, providing a rare instance where theory correctly predicts and closely justifies the best practical methods. We believe that this approach applies to a more general class of PDE-based inverse problems.

In Section 2 we describe the inverse problem and the algorithm variants used for its solution. Several aspects arise with the prospect of data completion: which data – the original or the completed – to use for carrying out the iteration, which data for controlling the iterative process, what stopping criterion to use, and more. These aspects are addressed in Section 2.1. The resulting algorithm, based on Algorithm 2 of [27], is given in Section 2.2. The specific EIT/DC resistivity inverse problem described in Section 2.3 then leads to the data completion methods developed and proved in Section 3.

In Section 4 we apply the algorithm variants developed in the two previous sections to solve test problems with different receiver locations. The purpose is to investigate whether the SS algorithms based on completed data achieve results of similar quality at a cheaper price, as compared to the RS method applied to the original data. Overall, very encouraging results are obtained even when the original data receiver

sets are rather sparse. Conclusions are offered in Section 5.

2 Stochastic algorithms for solving the inverse problem

The first two subsections below apply more generally than the third subsection. The latter settles on one application and leads naturally to Section 3.

Let us recall the acronyms for random subset (RS) and simultaneous sources (SS), used repeatedly in this section.

2.1 Algorithm variants

To compare the performance of our model recovery methods with completed data, $\tilde{D}$, against corresponding ones with the original data, D, we use the framework of Algorithm 2 in [27]. This algorithm consists of two stages within each GN iteration. The first stage produces a stabilized GN iterate, for which we use data denoted by $\hat{D}$. The second involves assessment of this iterate in terms of improvement and algorithm termination, using data $\bar{D}$. This second stage consists of evaluations of (7), in addition to (6). We consider three variants:

(i) $\hat{D} = D$, $\bar{D} = D$;

(ii) $\hat{D} = \tilde{D}$, $\bar{D} = \tilde{D}$;

(iii) $\hat{D} = \tilde{D}$, $\bar{D} = D$;

Note that only the RS method can be used in variant (i), whereas any of the SS methods as well as the RS method can be employed in variant (ii). In variant (iii) we can use a more accurate SS method for the stabilized GN stage and an RS method for the convergence checking stage, with the potential advantage that the evaluations of (7) do not use our "invented data". However, the disadvantage is that RS is potentially less suitable than Gaussian or Hutchinson precisely for tasks such as those in this second stage; see [26].

A major source of computational expense is the algorithm stopping criterion, which in [27] was taken to be

$$\phi(\mathbf{m}) \leq \rho, \tag{8}$$

for a specified tolerance ρ. In [27], we deliberately employed this criterion in order to be able to make fair comparisons among different methods. However, the evaluation of ϕ for this purpose is very expensive when s is large, and in practice ρ is hardly ever known in a rigid sense. In any case, this evaluation should be carried out as rarely as possible. In [27], we addressed this by proposing a safety check, called "uncertainty

check", which uses (7) as an unbiased estimator of $\phi(\mathbf{m})$ with a stochastic weight matrix $W = W^e$ which has far fewer columns than s, provided the columns of W^e are independent and satisfy (3). Thus, in the course of an iteration we can perform the relatively inexpensive uncertainty check whether

$$\hat{\phi}(\mathbf{m}, W^e) \leq \rho. \tag{9}$$

This is like the stopping criterion, but in expectation (with respect to W^e). If (9) is satisfied, it is an indication that (8) is likely to be satisfied as well, so we check the expensive (8) only then.

In the present article, we propose an alternative heuristic method of replacing (8) with another uncertainty check evaluation as in (9) with an independently drawn weight matrix $W^e \in \mathbb{R}^{s \times t_n}$, whose t_n columns have i.i.d. elements drawn from the Rademacher distribution (NB the Hutchinson estimator has smaller variance than Gaussian). The sample size t_n can be heuristically set as

$$t_n = \min\left(s, \max\left(t_0, s_n\right)\right), \tag{10}$$

where $t_0 > 1$ is some preset minimal sample size for this purpose. Thus, for each algorithm variant (i), (ii) or (iii), we consider two stopping criteria, namely,

(a) the hard (8), and

(b) the more relaxed (9)+(10).

When using the original data D in the second stage of our general algorithm, as in variants (i) and (iii) above, since the projection matrices P_i are not the same across experiments, one is restricted to the RS method as an unbiased estimator. However, when the completed data is used and we only have one P for all experiments, we can freely use the stochastic SS methods and leverage their rather better accuracy in order to estimate the true misfit $\phi(\mathbf{m})$. This is indeed an important advantage of data completion methods.

However, when using the completed data $\tilde{D}$ in the second stage of our general algorithm, as in variant (ii), an issue arises: when the data is completed, the given tolerance ρ loses its meaning and we need to take into account the effect of the additional data to calculate a new tolerance. Our proposed heuristic approach is to replace ρ with a new tolerance $\rho := (1 + c)\rho$, where c is the percentage of the data that needs to be completed expressed as a fraction. For example, if 30% of data is to be completed then we set $\rho := 1.3\rho$. Since the completed data after using (15) or (19) is smoothed and denoised, we only need to add a small fraction of the initial tolerance to get the new one, and in our experience, $1 + c$ is deemed to be a satisfactory factor. We experiment with this less rigid stopping criterion in Section 4.

2.2 General algorithm

Our general algorithm utilizes a stabilized Gauss-Newton (GN) method [9], where each iteration consists of two stages as described in Section 2.1. In addition to combining the elements described above, this algorithm also provides a schedule for selecting the sample size s_n in the nth stabilized GN iteration. In Algorithm 1, variants (i), (ii) and (iii), and criteria (a) and (b), are as specified in Section 2.1.

Algorithm 1 Solve inverse problem using variant (i), (ii) or (iii), cross validation, and stopping criterion (a) or (b)

Given: sources Q, measurements $\hat{D}$, measurements $\bar{D}$, stopping tolerance ρ, decrease factor $\kappa < 1$, and initial guess $\mathbf{m}_0$.

Initialize: $\mathbf{m} = \mathbf{m}_0$, $s_0 = 1$.

for $n = 0, 1, 2, \ldots$ until termination **do**

  - Choose a $W_n^f \in \mathbb{R}^{s \times s_n}$ stochastically from appropriate distribution.

  - **Fitting**: Perform one stabilized GN iteration, based on $\hat{D}$, with $W = W_n^f$.

  - Choose $W_n^c \in \mathbb{R}^{s \times s_n}$ and $W_n^e \in \mathbb{R}^{s \times s_n}$ stochastically from appropriate distribution.

  if $\hat{\phi}(\mathbf{m}_{n+1}, W_n^c) \leq \kappa \hat{\phi}(\mathbf{m}_n, W_n^c)$, based on $\bar{D}$, i.e., **Cross Validation** holds **then**

    - **Uncertainty Check**: Compute (7) based on $\bar{D}$ using $\mathbf{m}_{n+1}$ and W_n^e.

    if (9) holds **then**

      - **Stopping Criterion**:

      if Option (a) selected and (8) holds **then**

        terminate; otherwise set $s_{n+1} = s_n$.

      else

        Set $t_n = \min\left(s, \max\left(t_0, s_n\right)\right)$.

        Draw another $W_n^e \in \mathbb{R}^{s \times t_n}$ stochastically from appropriate distribution.

        Terminate if (9) holds using $\bar{D}$; otherwise set $s_{n+1} = s_n$.

      end if

    end if

  else

    - **Sample Size Increase**: for example, set $s_{n+1} = \min(2s_n, s)$.

  end if

end for

2.3 The EIT/DC resistivity inverse problem

For the forward problem we consider, following [17, 9, 27], a linear PDE of the form

$$\nabla \cdot (\sigma(\mathbf{x})\nabla u) = q(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{11a}$$

where σ is a given conductivity function which may be rough (e.g., discontinuous) but is bounded away from 0: there is a constant $\sigma_0 > 0$ such that $\sigma(\mathbf{x}) \geq \sigma_0, \ \forall \mathbf{x} \in \Omega$.

This elliptic PDE is subject to the homogeneous Neumann boundary conditions

$$\frac{\partial u}{\partial n} = 0, \quad \mathbf{x} \in \partial\Omega. \tag{11b}$$

In our numerical examples we take $\Omega \subset \mathbb{R}^d$ to be the unit square or unit cube, and the sources $\mathbf{q}$ to be the differences of δ-functions. Furthermore, the receivers (where data values are measured) lie in $\partial\Omega$, so in our data completion algorithms we approximate data along one of four edges in the 2D case or within one of six square faces in the 3D case.

For the inverse problem we introduce additional a priori information, when such is available, via a point-wise parameterization of $\sigma(\mathbf{x})$ in terms of $m(\mathbf{x})$. Define the transfer function

$$\psi(\tau) = \psi(\tau; \theta, \alpha_1, \alpha_2) = \alpha \tanh\left(\frac{\tau}{\alpha\theta}\right) + \frac{\alpha_1 + \alpha_2}{2}, \quad \alpha = \frac{\alpha_2 - \alpha_1}{2}. \tag{12}$$

If we know that the sought conductivity function $\sigma(\mathbf{x})$ takes only one of two values, σ_I or σ_{II}, at each $\mathbf{x}$, then we use an approximate level set function representation, writing $\sigma(\mathbf{x}) = \lim_{h\to 0} \sigma(\mathbf{x}; h)$, where

$$\sigma(\mathbf{x}; h) = \psi(m(\mathbf{x}); h, \sigma_I, \sigma_{II}). \tag{13}$$

The function ψ here depends on the resolution, or grid width h. More commonly, we may only know reasonably tight bounds, say $\sigma_{\min}$ and $\sigma_{\max}$, such that $\sigma_{\min} \leq \sigma(\mathbf{x}) \leq \sigma_{\max}$. Such information may be enforced using (12) by defining

$$\sigma(\mathbf{x}) = \psi(m(\mathbf{x})), \quad \text{with } \psi(\tau) = \psi(\tau; 1, \sigma_{\min}, \sigma_{\max}). \tag{14}$$

For details of this, as well as the PDE discretization and the *stabilized* GN iteration used, we refer to [9, 27] and references therein.

3   Data completion

Let $\Lambda_i \subset \partial\Omega$ denote the point set of receiver locations for the ith experiment. Our goal here is to extend the data for each experiment to the union $\Lambda = \bigcup_i \Lambda_i \subseteq \partial\Omega$, the common measurement domain. To achieve this, we choose a suitable boundary patch $\Gamma \subseteq \partial\Omega$, such that $\Lambda \subset \bar{\Gamma}$, where $\bar{\Gamma}$ denotes the closure of Γ with respect to the bounbdary subspace topology. For example, one can choose Γ to be the interior of the convex hull (on $\partial\Omega$) of Λ. We also assume that Γ can be selected such that it is a simply connected open set. For each experiment i, we then construct an extension function v_i on $\bar{\Gamma}$ which approximates the measured data on Λ_i. The extension method can be viewed as an inverse problem, and we select a regularization based on knowledge of the function space that v_i (which represents the restriction of potential

u_i to Γ) should live in. Once v_i is constructed, the extended data, $\tilde{\mathbf{d}}_i$, is obtained by restricting v_i to Λ, denoted in what follows by v_i^Λ. Specifically, for the receiver location $x_j \in \Lambda$, we set $[\tilde{\mathbf{d}}_i]_j = v_i(x_j)$, where $[\tilde{\mathbf{d}}_i]_j$ denotes the jth component of vector $\tilde{\mathbf{d}}_i$ corresponding to x_j. Below we show that the trace of potential u_i to the boundary is indeed continuous, thus point values of the extension function v_i make sense.

In practice, the conductivity $\sigma(\mathbf{x})$ in (11a) is often piecewise smooth with finite jump discontinuities. As such one is faced with two scenarios leading to two approximation methods for finding v_i: (a) the discontinuities are some distance away from Γ; and (b) the discontinuities extend all the way to Γ. These cases result in a different a priori smoothness of the field v_i on Γ. Hence, in this section we treat these cases separately and propose an appropriate data completion algorithm for each.

Consider the problem (11). In what follows we assume that Ω is a bounded open domain and $\partial\Omega$ is Lipschitz. Furthermore, we assume that σ is continuous on a finite number of disjoint subdomains, $\Omega_j \subset \Omega$, such that $\bigcup_{j=1}^N \overline{\Omega}_j = \overline{\Omega}$ and $\partial\Omega_j \cap \overline{\Omega} \in C^{2,\alpha}$, for some $0 < \alpha \le 1$, i.e., $\sigma \in C^2(\overline{\Omega}_j)$, $j = 1,\ldots,N$.[4] Moreover, assume that $q \in L_\infty(\Omega)$ and $q \in \mathrm{Lip}(\overline{\Omega}_j \cap \Omega)$, i.e., it is Lipschitz continuous in each subdomain; this assumption will be slightly weakened in Subsection 3.4.

Under these assumptions and for the Dirichlet problem with a $C^2(\partial\Omega)$ boundary condition, there is a constant γ, $0 < \gamma \le 1$, such that $u \in C^{2,\gamma}(\overline{\Omega}_j)$ [19, Theorem 4.1]. In [21, Corollary 7.3], it is also shown that the solution on the entire domain is Hölder continuous, i.e., $u \in C^\beta(\overline{\Omega})$ for some β, $0 < \beta \le 1$. Note that the mentioned theorems are stated for the Dirichlet problem, and in the present article we assume a homogeneous Neumann boundary condition. However, in this case we have infinite smoothness in the normal direction at the boundary, i.e., C^∞ Neumann condition, and no additional complications arise; see for example [29]. So the results stated above would still hold for (11).

3.1 Discontinuities in conductivity are away from common measurement domain

This scenario corresponds to the case where the boundary patch Γ can be chosen such that $\Gamma \subset (\partial\Omega_j \cap \partial\Omega)$ for some j. Then we can expect a rather smooth field at Γ; precisely, $u \in C^{2,\gamma}(\overline{\Gamma})$. Thus, u belongs to the Sobolev space $H^2(\Gamma)$, and we can impose this knowledge in our continuous completion formulation. For the ith experiment, we define our data completion function $v_i \in H^2(\Gamma) \cap C(\overline{\Gamma})$ as

$$v_i = \arg\min_v \ \frac{1}{2}\|v^{\Lambda_i} - \mathbf{d}_i\|_2^2 + \lambda\,\|\Delta_S v\|_{L_2(\Gamma)}^2\,, \tag{15}$$

where Δ_S is the Laplace-Beltrami operator for the Laplacian on the boundary surface and v^{Λ_i} is the restriction of the continuous function v to the point set Λ_i. The regularization parameter λ depends on the amount of noise in our data; see Section 3.3.

---

[4]$\overline{X}$ denotes the closure of X with respect to the appropriate topology.

We next discretize (15) using a mesh on Γ as specified in Section 4, and solve the resulting linear least squares problem using standard techniques.

Figure 1 shows an example of such data completion. The true field and the measured data correspond to an experiment described in Example 3 of Section 4. We only plot the profile of the field along the top boundary of the 2D domain. As can be observed, the approximation process imposes smoothness which results in an excellent completion of the missing data, despite the presence of noise at a fairly high level.
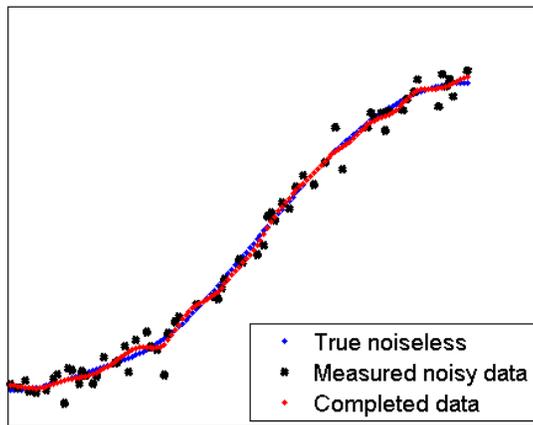
Figure 1: Completion using the regularization (15), for an experiment taken from Example 3 where 50% of the data requires completion and the noise level is 5%. Observe that even in the presence of significant noise, the data completion formulation (15) achieves a good quality reconstruction.

3.2 Discontinuities in conductivity extend all the way to common measurement domain

This situation corresponds to the case in which Γ can only be chosen such that it intersects more than just one of the $(\partial\Omega \cap \partial\Omega_j)$'s. More precisely, assume that there is an index set $\mathcal{J} \subseteq \{1, 2, \cdots N\}$ with $|\mathcal{J}| = K \geq 2$ such that $\{\overline{\Gamma \cap (\partial\Omega \cap \partial\Omega_j)^\circ} \; , \; j \in \mathcal{J}\}$ forms a set of disjoint subsets of Γ such that $\overline{\Gamma} = \bigcup_{j \in \mathcal{J}} \overline{\Gamma \cap (\partial\Omega \cap \partial\Omega_j)^\circ}$, where X° denotes the interior of the set X, and that the interior is with respect to the subspace topology on $\partial\Omega$. In such a case u, restricted to Γ, is no longer necessarily in $H^2(\Gamma)$. Hence, the smoothing term in (15) is no longer valid, as $\|\Delta_S u\|_{L_2(\Gamma)}$ might be undefined or infinite. However, as described above, we know that the solution is piecewise smooth and overall continuous, i.e., $u \in C^{2,\gamma}(\overline{\Omega}_j)$ and $u \in C^\beta(\overline{\Omega})$. The

following theorem shows that the smoothness on Γ is not completely gone: we may lose one degree of regularity at worst.

Theorem 1 *Let U and $\{U_j |\, j = 1, 2, \ldots, K\}$ be open and bounded sets such that the U_j are pairwise disjoint and $\overline{U} = \bigcup_{j=1}^{K} \overline{U}_j$. Further, let $u \in C(\overline{U}) \cap H^1(U_j)\ \forall j$. Then $u \in H^1(U)$.*

Proof It is easily seen that since $u \in C(\overline{U})$ and U is bounded, then $u \in L_2(U)$. Now, let $\phi \in C_0^\infty(U)$ be a test function and denote $\partial_i \equiv \frac{\partial}{\partial \mathbf{x}_i}$. Using the assumptions that the U_j's form a partition of U, u is continuous in $\overline{U}$, ϕ is compactly supported inside U, and the fact that the ∂U_j's have measure zero, we obtain

$$\int_U u \partial_i \phi = \int_{\overline{U}} u \partial_i \phi = \int_{\cup_{j=1}^K \overline{U}_j} u \partial_i \phi = \int_{(\cup_{j=1}^K U_j) \bigcup (\cup_{j=1}^K \partial U_j)} u \partial_i \phi$$

$$= \int_{\cup_{j=1}^K U_j} u \partial_i \phi = \sum_{j=1}^K \int_{U_j} u \partial_i \phi = \sum_{j=1}^K \int_{\partial U_j} u \phi \nu_i^j - \sum_{j=1}^K \int_{U_j} \partial_i u \phi,$$

where ν_i^j is the ith component of the outward unit surface normal to ∂U_j. Since $u \in H^1(U_j)\ \forall j$, the second part of the rightmost expression makes sense. Now, for two surfaces ∂U_m and ∂U_n such that $\partial U_m \cap \partial U_n \neq \emptyset$, we have $\nu_i^m(\mathbf{x}) = -\nu_i^n(\mathbf{x})\ \forall \mathbf{x} \in \partial U_m \cap \partial U_n$. This fact, and noting in addition that ϕ is compactly supported inside U, makes the first term in the right hand side vanish. We can now define the weak derivative of u with respect to $\mathbf{x}_i$ to be

$$v(\mathbf{x}) = \sum_{j=1}^K \partial_i u \mathcal{X}_{U_j}, \tag{16}$$

where $\mathcal{X}_{U_j}$ denotes the characteristic function of the set U_j. This yields

$$\int_U u \partial_i \phi = -\int_U v \phi. \tag{17}$$

Also

$$\|v\|_{L_2(U)} \leq \sum_{j=1}^K \|\partial_i u\|_{L_2(U_j)} < \infty, \tag{18}$$

and thus we conclude that $u \in H^1(U)$. $\blacksquare$

If the assumptions stated at the beginning of this section hold then we can expect a field $u \in H^1(\Gamma) \cap C(\bar{\Gamma})$. This is obtained by invoking Theorem 1 with $U = \Gamma$ and $U_j = \Gamma \cap (\partial \Omega \cap \partial \Omega_j)^\circ$ for all $j \in \mathcal{J}$.

Now we can formulate the data completion method as

$$v_i = \arg\min_v \; \frac{1}{2}\|v^{\Lambda_i} - \mathbf{d}_i\|_2^2 + \lambda \left\|\nabla_S v\right\|_{L_2(\Gamma)}^2, \tag{19}$$

where v^{Λ_i} and λ are as in Section 3.1.

Figure 2 shows an example of data completion using the formulation (19), depicting the profile of v_i along the top boundary. The field in this example is continuous and only piecewise smooth. The approximation process imposes less smoothness along the boundary as compared to (15), and this results in an excellent completion of the missing data, despite a nontrivial level of noise.
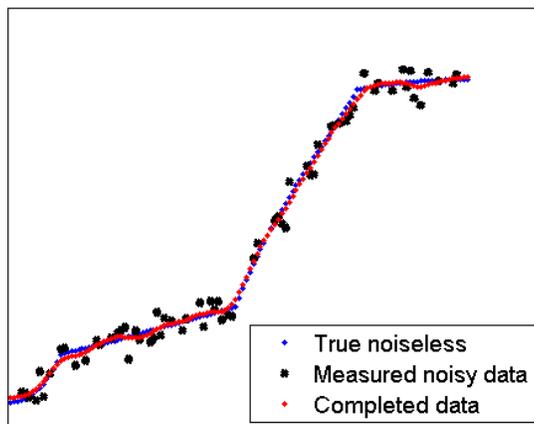
Figure 2: Completion using the regularization (19), for an experiment taken from Example 2 where 50% of the data requires completion and the noise level is 5%. Discontinuities in the conductivity extend to the measurement domain and their effect on the field profile along the boundary can be clearly observed. Despite the large amount of noise, data completion formulation (19) achieves a good reconstruction.

To carry out our data completion strategy, the problems (15) or (19) are discretized. This is followed by a straightforward linear least squares technique, which can be carried out very efficiently. Moreover, this is a preprocessing stage performed once, which is completed before the algorithm for solving the nonlinear inverse problem commences. Also, as the data completion for each experiment can be carried out independently of others, the preprocessing stage can be done in parallel if needed. Furthermore, the length of the vector of unknowns v_i is relatively small compared to those of u_i because only the boundary is involved. All in all the amount of work involved in the data completion step is dramatically less than one full evaluation of the misfit function (1).

3.3 Determining the regularization parameter

Let us write the discretization of (15) or (19) as

$$\min_{\mathbf{v}} \frac{1}{2}\|\hat{P}_i\mathbf{v} - \mathbf{d}_i\|_2^2 + \lambda\|L\mathbf{v}\|_2^2, \tag{20}$$

where L is the discretization of the surface gradient or Laplacian operator, $\mathbf{v}$ is a vector whose length is the size of the discretized Γ, $\hat{P}_i$ is the projection matrix from the discretization of Γ to Λ_i, and $\mathbf{d}_i$ is the ith original measurement vector.

Determining λ in this context is a textbook problem; see, e.g., [32]. Viewing it as a parameter, we have a linear least squares problem for $\mathbf{v}$ in (20), whose solution can be denoted $\mathbf{v}(\lambda)$. Now, in the simplest case, which we assume in our experiments, the noise level for the ith experiment, η_i, is known, so one can use the discrepancy principle to pick λ such that

$$\left\|\hat{P}_i\mathbf{v}(\lambda) - \mathbf{d}_i\right\|_2^2 \leq \eta_i. \tag{21}$$

Numerically, this is done by setting equality in (21) and solving the resulting nonlinear equation for λ using a standard root finding technique.

If the noise level is not known, one can use the generalized cross validation (GCV) method or the L-curve method; see [32]. We need not dwell on this longer here.

3.4 Point sources and boundaries with corners

In the numerical examples of Section 4, as in [9, 27], we use delta function combinations as the sources $q_i(\mathbf{x})$, and these are clearly not honest L_∞ functions. Moreover, our domains Ω are a square or a cube and as such they have corners.

However, the theory developed above, and the data completion methods that it generates, can be extended to our experimental setting because we have control over the experimental setup. The desired effect is obtained by simply separating the location of each source from any of the receivers, and avoiding domain corners altogether.

Thus, consider in (11a) a source function of the form

$$q(\mathbf{x}) = \hat{q}(\mathbf{x}) + \delta(\mathbf{x} - \mathbf{x}^*) - \delta(\mathbf{x} - \mathbf{x}^{**}),$$

where $\hat{q}$ satisfies the assumptions previously made on q. Then we select $\mathbf{x}^*$ and $\mathbf{x}^{**}$ such that there are two open balls $B(\mathbf{x}^*, r)$ and $B(\mathbf{x}^{**}, r)$ of radius $r > 0$ each and centered at $\mathbf{x}^*$ and $\mathbf{x}^{**}$, respectively, where (i) no domain corner belongs to $B(\mathbf{x}^*, r) \cup B(\mathbf{x}^{**}, r)$, and (ii) $(B(\mathbf{x}^*, r) \cup B(\mathbf{x}^{**}, r)) \cap \Gamma$ is empty. Now, in our elliptic PDE problem the lower smoothness effect of either a domain corner or a delta function is local! In particular, the contribution of the point source to the flux $\sigma\nabla u$ is the integral of $\delta(\mathbf{x} - \mathbf{x}^*) - \delta(\mathbf{x} - \mathbf{x}^{**})$, and this is smooth outside the union of the two balls.

4 Numerical experiments

The PDE problem used in our experiments is described in Section 2.3. For each experiment i there is a positive unit point source at $\mathbf{x}_1^i$ and a negative sink at $\mathbf{x}_2^i$, where $\mathbf{x}_1^i$ and $\mathbf{x}_2^i$ are two locations on the boundary $\partial\Omega$. Hence in (11a) we must consider sources of the form $q_i(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_1^i) - \delta(\mathbf{x} - \mathbf{x}_2^i)$, i.e., a difference of two δ-functions. For our experiments in 2D, when we place a source on the left boundary, the corresponding sink on the right boundary is placed in every possible combination. Hence, having p locations on the left boundary for the source would result in $s = p^2$ experiments. The receivers are located at the top and bottom boundaries. As such, the completion steps (15) or (19) are carried out separately for the top and bottom boundaries. No source or receiver is placed at the corners. In 3D we use an arrangement whereby four boreholes are located at the four edges of the cube, and source and sink pairs are put at opposing boreholes in every combination, except that there are no sources on the point of intersection of boreholes and the surface, i.e., at the top four corners, since these four nodes are part of the surface where data values are gathered.

In the sequel we generate data $\mathbf{d}_i$ by using a chosen true model (or ground truth) and a source-receiver configuration as described above. Since the field u from (11) is only determined up to a constant, only voltage differences are meaningful. Hence we subtract for each i the average of the boundary potential values from all field values at the locations where data is measured. As a result each row of the projection matrix P_i has zero sum. This is followed by peppering these values with additive Gaussian noise to create the data $\mathbf{d}_i$ used in our experiments. Specifically, for an additive noise of 2%, say, denoting the "clean data" $l \times s$ matrix by D^*, we reshape this matrix into a vector $\mathbf{d}^*$ of length sl, calculate the standard deviation $\mathtt{sd} = .02\|\mathbf{d}^*\|/\sqrt{sl}$, and define $D = D^* + \mathtt{sd} * \mathtt{randn(l,s)}$ using MATLAB's random generator function $\mathtt{randn}$.

For all of our numerical experiments, the "true field" is calculated on a grid that is twice as fine as the one used to reconstruct the model. For the 2D examples, the reconstruction is done on a uniform grid of size 129^2 with $s = 961$ experiments in the setup described above. For the 3D examples, we set $s = 512$ and employ a uniform grid of size 33^3, except for Example 3 where the grid size is 17^3.

Numerical examples are presented for both cases described in Sections 3.1 and 3.2. For all of our numerical examples except Examples 5 and 6, we use the transfer function (14) with $\sigma_{\max} = 1.2 \max \sigma(\mathbf{x})$, and $\sigma_{\min} = \frac{1}{1.2} \min \sigma(\mathbf{x})$. In the ensuing calculations we then "forget" what the exact $\sigma(\mathbf{x})$ is. Further, in the stabilized GN iteration we employ preconditioned conjugate gradient (PCG) inner iterations, setting the PCG iteration limit to $r = 20$, and the PCG tolerance to 10^{-3}. The initial guess is $\mathbf{m}_0 = \mathbf{0}$. Examples 5 and 6 are carried out using the level set method (13). Here we can set $r = 5$, significantly lower than above. The initial guess for the level set example is a cube with rounded corners inside Ω (see Figure 2 in [27]).

For Examples 1, 2, 3 and 5, in addition to displaying the log conductivities (i.e.,

$\log(\sigma)$) for each reconstruction, we also show the log-log plot of misfit on the entire data (i.e., $\|F(\mathbf{m}) - D\|_F$) vs. PDE count. A table of total PDE counts (not including what extra is required for the plots) for each method is displayed. In order to simulate the situation where sources do not share the same receivers, we first generate the data fully on the entire domain of measurement and then knock out at random some percentage of the generated data. This setting roughly corresponds to an EMG experiment with faulty receivers.

For each example, we use Algorithm 1 with one of the variants (i), (ii) or (iii) paired with one of the stopping criteria (a) or (b). For instance, when using variant (ii) with the soft stopping criterion (b), we denote the resulting algorithm by (ii, b). For the relaxed stopping rule (b) we (conservatively) set $t_0 = 100$ in (10). A computation using RS applied to the original data, using variant (i,x), is compared to one using SS applied to the completed data through variant (ii,x) or (iii,x), where x stands for a or b.

For convenience of cross reference, we gather all resulting seven algorithm comparisons and corresponding work counts in Table 1 below. For Examples 1, 2, 3 and 5, the corresponding entries of this table should be read together with the misfit plots for each example.

Example	Algorithm	Random Subset	Data Completion
1	$(i, a) \mid (iii, a)$	3,647	1,716
2	$(i, a) \mid (iii, a)$	6,279	1,754
3	$(i, a) \mid (iii, a)$	3,887	1,704
4	$(i, b) \mid (ii, b)$	4,004	579
5	$(i, a) \mid (iii, a)$	3,671	935
6	$(i, b) \mid (ii, b)$	1,016	390
7	$(i, b) \mid (ii, b)$	4,847	1,217

Table 1: Algorithm and work in terms of number of PDE solves, comparing RS against data completion using Gaussian SS.

Example 1 *In this example, we place two target objects of conductivity $\sigma_I = 1$ in a background of conductivity $\sigma_{II} = 0.1$, and 5% noise is added to the data as described above. Also, 25% of the data requires completion. The discontinuities in the conductivity are touching the measurement domain, so we use (19) to complete the data. The hard stopping criterion (a) is employed, and iteration control is done using the original data, i.e., variants (i, a) and (iii, a) are compared: see the first entry of Table 1 and Figure 6(a).*

The corresponding reconstructions are depicted in Figure 3. It can be seen that roughly the same quality reconstruction is obtained using the data completion method at less than half the price.
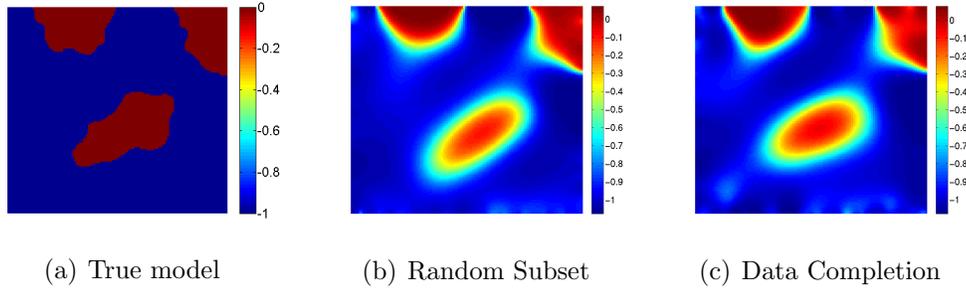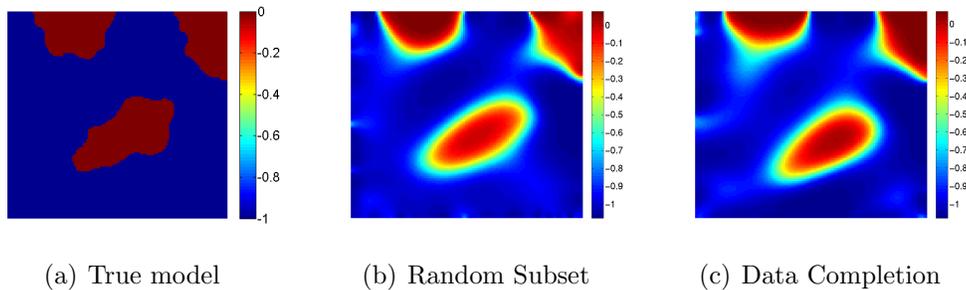
(a) True model (b) Random Subset (c) Data Completion

Figure 3: Example 1 – reconstructed log conductivity with 25% data missing and 5% noise. Regularization (19) has been used to complete the data.

Example 2 *This example is the same as Example 1, except that 50% of the data is missing and requires completion. The same algorithm variants as in Example 1 are compared. The reconstructions are depicted in Figure 4, and comparative computational results are recorded in Table 1 and Figure 6(b).*

(a) True model (b) Random Subset (c) Data Completion

Figure 4: Example 2 – reconstructed log conductivity with 50% data missing and 5% noise. Regularization (19) has been used to complete the data.

Similar observations to those in Example 1 generally apply here as well, despite the smaller amount of original data.

Example 3 *This is the same as Example 2 in terms of noise and the amount of missing data, except that the discontinuities in the conductivity are some distance away from the common measurement domain, so we use (15) to complete the data. The same algorithm variants as in the previous two examples are compared, thus isolating the effect of a smoother data approximant.*

Results are recorded in Figure 5, the third entry of Table 1 and Figure 6(c).

Figures 3, 4 and 5 in conjunction with Figure 6 as well as Table 1, reflect superiority of the SS method combined with data completion over the RS method with the original data. From the first three entries of Table 1, we see that the SS reconstruction with completed data can be done more efficiently by a factor of more than two.

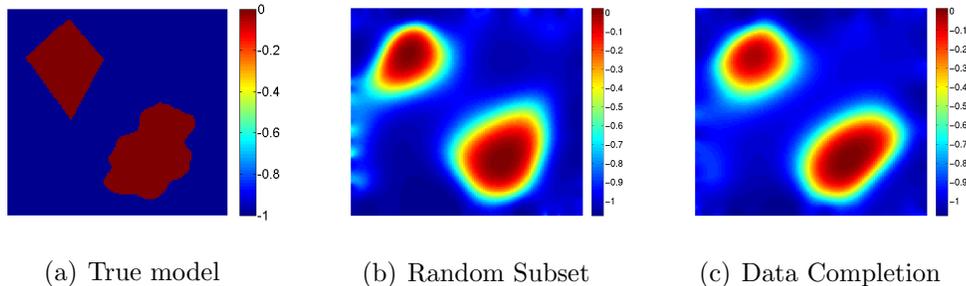(a) True model   (b) Random Subset   (c) Data Completion

Figure 5: Example 3 – reconstructed log conductivity with 50% data missing and 5% noise. Regularization (15) has been used to complete the data.

The quality of reconstruction is also very good. Note that the graph of the misfit for Data Completion lies mostly under that of Random Subset. This means that, given a fixed number of PDE solves, we obtain a lower (thus better) misfit for the former than for the latter.

Next, we consider examples in 3D.

Example 4 *In this example, the discontinuities in the true, piecewise constant conductivity extend all the way to the common measurement domain, see Figure 7. We therefore use (19) to complete the data. The target object has the conductivity $\sigma_I = 1$ in a background with conductivity $\sigma_{II} = 0.1$. We add 2% noise and knock out 50% of the data. Furthermore, we consider the relaxed stopping criterion (b). With the original data (hence using RS), the variant (i, b) is employed, and this is compared against the variant (ii, b) with SS applied to the completed data. For the latter case, the stopping tolerance is adjusted as discussed in Section 2.1.*

Reconstruction results are depicted in Figure 8, and work estimates are gathered in the 4th entry of Table 1. It can be seen that the results using data completion, obtained at about 1/7th the cost, are comparable to those obtained with RS applied to the original data.

Example 5 *The underlying model in this example is the same as that in Example 4 except that, since we intend to plot the misfit on the entire data at every GN iteration, we decrease the reconstruction mesh resolution to 17^3. Also, 30% of the data requires completion, and we use the level set transfer function (13) to reconstruct the model. With the original data, we use the variant (i, a), while the variant (iii, a) is used with the completed data. The reconstruction results are recorded in Figure 9, and performance indicators appear in Figure 10 as well as Table 1.*

The algorithm proposed here produces a better reconstruction than RS on the original data. A relative efficiency observation can be made from Table 1, where a factor of roughly 4 is revealed.
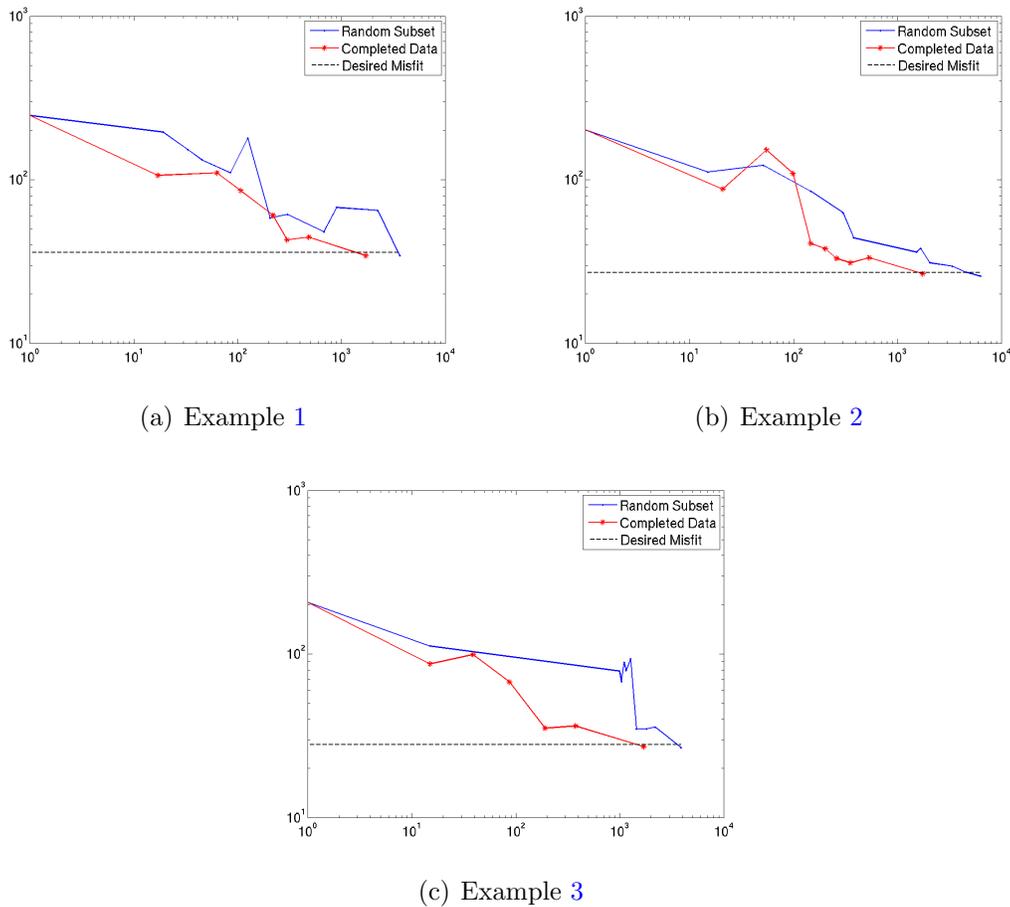
(a) Example 1

(b) Example 2

(c) Example 3

Figure 6: Data misfit vs. PDE count for Examples 1, 2 and 3.

Example 6 *This is exactly the same as Example 4, except that we use the level set transfer function* (13) *to reconstruct the model. The same variants of Algorithm 1 as in Example 4 are employed.*

It is evident from Figure 11 that employing the level set formulation allows a significantly better quality reconstruction than in Example 4. This is expected, as much stronger assumptions on the true model are utilized. It was shown in [8, 27] that using level set functions can greatly reduce the total amount of work, and this is observed here as well.

It can be clearly seen from the results of Examples 4, 5 and 6 that Algorithm 1 does a great job recovering the model using the completed data plus the SS method as compared to RS with the original data. This is so both in terms of total work and the quality of the recovered model. Note that for all reconstructions, the conductive object placed deeper than the ones closer to the surface is not recovered well. This is due to the fact that we only measure on the surface and the information coming
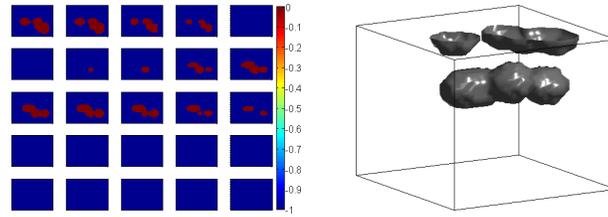
Figure 7: True Model for Example 4.

(a) RS slices
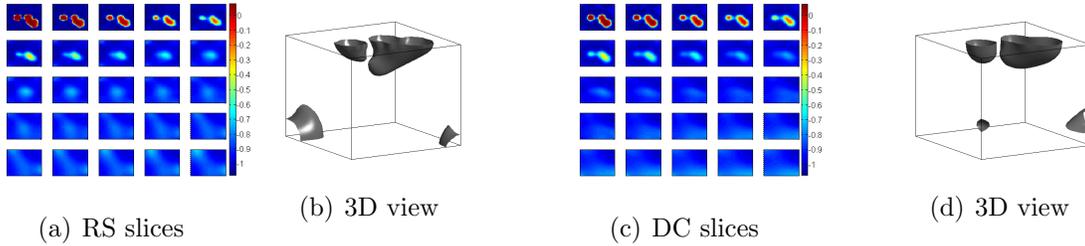
(b) 3D view

(c) DC slices

(d) 3D view

Figure 8: Example 4 – reconstructed log conductivity for the 3D model with (a,b) Random Subset, (c,d) Data Completion for the case of 2% noise and 50% of data missing. Regularization (19) has been used to complete the data.

from this deep conductive object is majorized by that coming from the objects closer to the surface.

Example 7 *In this 3D example, we examine the performance of our data completion approach for more severe cases of missing data. For this example, we place a target object of conductivity $\sigma_I = 1$ in a background with conductivity $\sigma_{II} = 0.1$, see Figure 12, and 2% noise is added to the "exact" data. Then we knock out 70% of the data and use (15) to complete it. The algorithm variants employed are the same as in Examples 4 and 6.*

Results are gathered in Figures 13 as well as Table 1. The data completion plus simultaneous sources algorithm again does well, with an efficiency factor ≈ 4.

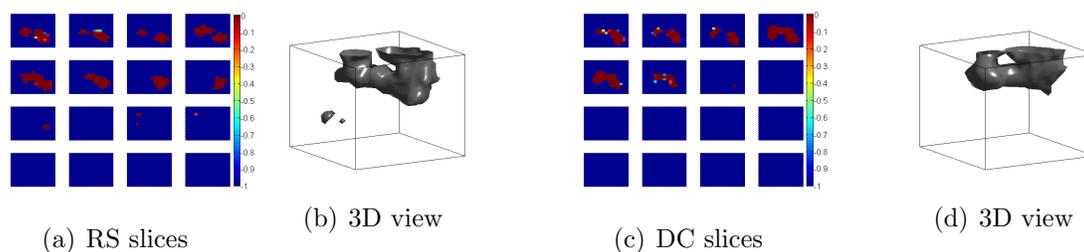(a) RS slices   (b) 3D view   (c) DC slices   (d) 3D view

Figure 9: Example 5 – reconstructed log conductivity for the 3D model using the level set method with (a,b) Random Subset, (c,d) Data Completion for the case of 2% noise and 30% of data missing. Regularization (19) has been used to complete the data.
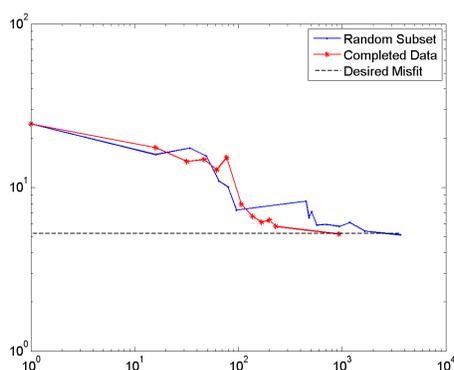

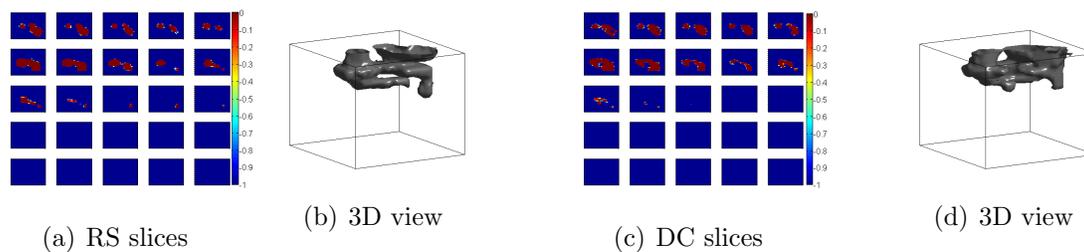Figure 10: Data misfit vs. PDE count for Example 5.

(a) RS slices   (b) 3D view   (c) DC slices   (d) 3D view

Figure 11: Example 6 – reconstructed log conductivity for the 3D model using the level set method with (a,b) Random Subset, (c,d) Data Completion for the case of 2% noise and 50% of data missing. Regularization (19) has been used to complete the data.

5   Conclusions and further comments

This paper is a sequel to [27] in which we studied the case where sources share the same receivers. Here we have focused on the very practical case where sources do not share the same receivers yet are distributed in a particular manner, and have proposed a new approach based on appropriately regularized data completion. Our data completion methods are motivated by theory in Sobolev spaces regarding the properties of weak solutions along the domain boundary. The resulting completed data allows an efficient use of the methods developed in [27] as well as utilization of a relaxed stopping criterion. Our approach shows great success in cases of moderate data completion, say up to 60-70%. In such cases we have demonstrated that, utilizing some variant of Algorithm 1, an execution speedup factor of at least 2 and often much more can be achieved while obtaining excellent reconstructions.
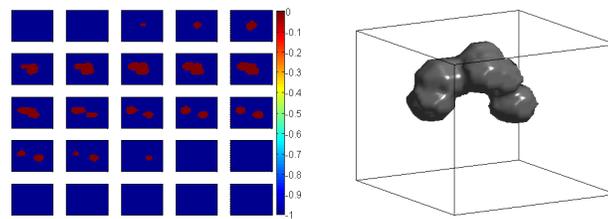

Figure 12: True Model for Example 7.

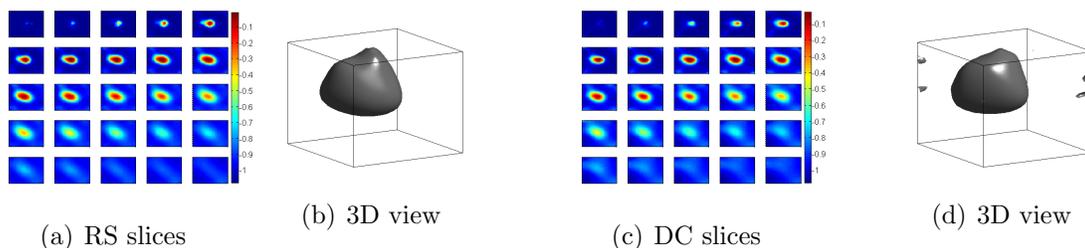(a) RS slices

(b) 3D view

(c) DC slices

(d) 3D view

Figure 13: Example 7 – reconstructed log conductivity for the 3D model with (a,b) Random Subset, (c,d) Data Completion for the case of 2% noise and 70% data missing. Regularization (15) has been used to complete the data.

It needs to be emphasized that a blind employment of some interpolation/approximation method would not take into account available a priori information about the sought signal. In contrast, the method developed in this paper, while being very simple, is in fact built upon such a priori information, and is theoretically justified.

Note that with the methods of Section 3 we have also replaced the original data with new, approximate data. Alternatively we could keep the original data, and

just add the missing data sampled from v_i at appropriate locations. The potential advantage of doing this is that fewer changes are made to the original problem, so it would seem plausible that the data extension will produce results that are close to the more expensive inversion without using the simultanous sources method, at least when there are only a few missing receivers. However, we found in practice that this method yields similar or worse reconstructions for moderate or large amounts of missing data as compared to the methods of Section 3.

For severe cases of missing data, say 80% or more, we do not recommend data completion in the present context as a safe approach. With so much completion the bias in the completed field could overwhelm the given observed data, and the recovered model may not be correct. In such cases, one can use the RS method applied to the original data. A good initial guess for this method may still be obtained with the SS method applied to the completed data. Thus, one can always start with the most daring variant (ii, b) of Algorithm 1, and add a more conservative run of variant (i, b) on top if necessary.

If the forward problem is very diffusive and has a strong smoothing effect, as is the case for the DC-resistivity and EIT problems, then data completion can be attempted using a (hopefully) good guess of the sought model $\mathbf{m}$ by solving the forward problem and evaluating the solution wherever necessary [16]. The rationale here is that even relatively large changes in $m(\mathbf{x})$ produce only small changes in the fields $u_i(\mathbf{x})$. However, such a prior might prove dominant, hence risky, and the data produced in this way, unlike the original data, no longer have natural high frequency noise components. Indeed, a potential advantage of this approach is in using the difference between the original measured data and the calculated prior field at the same locations for estimating the noise level ϵ for a subsequent application of the Morozov discrepancy principle [32, 11].

The specific data completion techniques proposed in this paper have been justified and used in our model DC resistivity (or EIT) problem. However, the overall idea can be extended to other PDE based inverse problems as well by studying the properties of the solution of the forward problem. One first needs to see what the PDE solutions are expected to behave like on the measurement domain, for example on a portion of the boundary, and then imposing this prior knowledge in the form of an appropriate regularizer in the data completion formulation. Following that, the rest can be similar to our approach here. Investigating such extensions to other PDE models is a subject for future studies.

Acknowledgments

References

[1] D. Achlioptas. Database-friendly random projections. In *ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 01*, volume 20, pages 274–281, 2001.

[2] A. Alessandrini and S. Vessella. Lipschitz stability for the inverse conductivity problem. *Adv. Appl. Math.*, 35:207–241, 2005.

[3] K. Astala and L. Paivarinta. Calderon inverse conductivity problem in the plane. *Annals of Math.*, 163:265–299, 2006.

[4] H. Avron and S. Toledo. Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix. *JACM*, 58(2), 2011. Article 8.

[5] L. Borcea, J. G. Berryman, and G. C. Papanicolaou. High-contrast impedance tomography. *Inverse Problems*, 12:835–858, 1996.

[6] R. Byrd, G. Chin, W. Neveitt, and J. Nocedal. On the use of stochastic hessian information in optimization methods for machine learning. *SIAM J. Optimization*, 21(3):977–995, 2011.

[7] M. Cheney, D. Isaacson, and J. C. Newell. Electrical impedance tomography. *SIAM Review*, 41:85–101, 1999.

[8] K. van den Doel and U. Ascher. On level set regularization for highly ill-posed distributed parameter estimation problems. *J. Comp. Phys.*, 216:707–723, 2006.

[9] K. van den Doel and U. Ascher. Adaptive and stochastic algorithms for EIT and DC resistivity problems with piecewise constant solutions and many measurements. *SIAM J. Scient. Comput.*, 34:DOI: 10.1137/110826692, 2012.

[10] O. Dorn, E. L. Miller, and C. M. Rappaport. A shape reconstruction method for electromagnetic tomography using adjoint fields and level sets. *Inverse Problems*, 16, 2000. 1119-1156.

[11] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer, Dordrecht, 1996.

[12] A. Fichtner. *Full Seismic Waveform Modeling and Inversion*. Springer, 2011.

[13] M. Friedlander and M. Schmidt. Hybrid deterministic-stochastic methods for data fitting. *SIAM J. Scient. Comput.*, 34(3), 2012.

[14] S. Geisser. *Predictive Inference*. New York: Chapman and Hall, 1993.

[15] E. Haber, U. Ascher, and D. Oldenburg. Inversion of 3D electromagnetic data in frequency and time domain using an inexact all-at-once approach. *Geophysics*, 69:1216–1228, 2004.

[16] E. Haber and M. Chung. Simultaneous source for non-uniform data variance andmissing data. 2012. submitted.

[17] E. Haber, M. Chung, and F. Herrmann. An effective method for parameter estimation with PDE constraints with multiple right-hand sides. *SIAM J. Optimization*, 22:739–757, 2012.

[18] F. Herrmann, Y. Erlangga, and T. Lin. Compressive simultaneous full-waveform simulation. *Geophysics*, 74:A35, 2009.

[19] V. Isakov. *Inverse Problems for Partial Differential Equations*. Springer, 2006.

[20] A. Juditsky, G. Lan, A. Nemirovski, and A. Shapiro. Stochastic approximation approach to stochastic programming. *SIAM J. Optimization*, 19(4):1574–1609, 2009.

[21] Yan Yan Li, Michael Vogelius, and Communicated R. V. Kohn. Gradient estimates for solutions to divergence form elliptic equations with discontinuous coefficients. *Arch. Rational Mech. Anal*, 153:91–151, 2000.

[22] G. A. Newman and D. L. Alumbaugh. Frequency-domain modelling of airborne electromagnetic responses using staggered finite differences. *Geophys. Prospecting*, 43:1021–1042, 1995.

[23] L. Paivarinta, A. Panchenko, and G. Uhlmann. Complex geometrical optics solutions for Lipschitz conductivities. *Rev. Mat. Iberoamericana*, 19:57–72, 2003.

[24] A. Pidlisecky, E. Haber, and R. Knight. RESINVM3D: A MATLAB 3D Resistivity Inversion Package. *Geophysics*, 72(2):H1–H10, 2007.

[25] J. Rohmberg, R. Neelamani, C. Krohn, J. Krebs, M. Deffenbaugh, and J. Anderson. Efficient seismic forward modeling and acquisition using simultaneous random sources and sparsity. *Geophysics*, 75(6):WB15–WB27, 2010.

[26] F. Roosta-Khorasani and U. Ascher. Improved bounds on sample size for implicit matrix trace estimators. 2013. submitted.

[27] F. Roosta-Khorasani, K. van den Doel, and U. Ascher. Stochastic algorithms for inverse problems involving pdes and many measurements. *SIAM J. SISC*, 2013. accepted.

[28] A. Shapiro, D. Dentcheva, and D. Ruszczynski. *Lectures on Stochastic Programming: Modeling and Theory*. Piladelphia: SIAM, 2009.

[29] S. Shkoller. *Lecture Notes on Partial Differential Equations.* Department of Mathematics, University of California, Davis, June 2012.

[30] N. C. Smith and K. Vozoff. Two dimensional DC resistivity inversion for dipole dipole data. *IEEE Trans. on geoscience and remote sensing*, GE 22:21–28, 1984.

[31] T. van Leeuwen, S. Aravkin, and F. Herrmann. Seismic waveform inversion by stochastic optimization. *Hindawi Intl. J. Geophysics*, 2011:doi:10.1155/2011/689041, 2012.

[32] C. Vogel. *Computational methods for inverse problem.* SIAM, Philadelphia, 2002.

[33] J. Young and D. Ridzal. An application of random projection to parameter estimation in partial differential equations. *SIAM J. SISC*, 34:A2344–A2365, 2012.