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

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

January 2, 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 a gradient or Laplacian regularization, extending existing data for each experiment to the union of all receiver locations. Results using the method of simultaneous sources with the completed data are then compared to those obtained by a more general but slower random subset 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 where the use of many experiments is crucial for obtaining credible reconstructions in practical situations [22, 10, 15, 12, 18, 25, 30, 29, 24, 6, 4]. Extensive theory (e.g., [23, 3, 2]) also suggests that many well-placed

*Dept. of Computer Science, University of British Columbia, Vancouver, Canada farbod/kvdoel/ascher@cs.ubc.ca . This work was supported in part by NSERC Discovery Grant 84306.

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 $\mathbf{d}_i \in \mathbb{R}^l$, $i = 1, 2, \ldots, s$. 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.[1] 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 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

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

[1] 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 $s_n \le s$ potentially satisfying $s_n \ll s$; see, e.g., [27, 20, 14]. Several recent papers have proposed methods to control the size s_n [9, 13, 5, 26]. Let us now concentrate on one such iteration n, for which a specialized Gauss-Newton (GN) or LBFGS method may be employed. Still assuming $P_i = P \ \forall i$, We can write (4) using (5) and 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 its estimator in the nth iteration as

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

where $W = W_n = \left[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{s_n}\right]$ is an $s \times s_n$ matrix. Different methods of simultaneous sources are obtained by using different algorithms for this *model and data reduction* process. In [26] 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 DC-resistivity problem, their performance was roughly comparable (although for just estimating the misfit function by (6), only the stochastic methods work well).

A fourth method considered in [26] (following [9]) was called "Random Subset", 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.[2] 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 projection matrix 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

[2] The relative efficiency factor further increases if a less conservative criterion is used for algorithm termination.

instance, because such maps assume knowledge of the field u (see, e.g., (8) 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.

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 [31, 11].

Our approach is to approximate or interpolate the given data directly while taking advantage of prior knowledge as to how the fields $\mathbf{u}_i$ must behave. For this, we use a Tikhonov-type regularization in our approximation, penalizing the discretized L_2 integral norm of the gradient or Laplacian of the fields. We believe that this approach applies to a more general class of PDE-based inverse problems. Our approach is further described and theoretically justified in Section 3.

We emphasize that the full justification of data completion by any method or approach must ultimately involve comparisons to existing alternative methods. In Section 4 we therefore apply one of the algorithms in [26] to test problems that are similar to those considered in [26] (although with different receiver locations of course). The purpose is to investigate whether the simultaneous source algorithm based on completed data achieves results of similar quality at a cheaper price, as compared to the random subset method of [9]. Overall, encouraging results are obtained provided that the original data set is not too sparse, and we quantify this for the problem described in Section 2. Conclusions are offered in Section 5.

2 The inverse problem

For the forward problem we consider, following [26, 9, 17], a linear PDE of the form

$$\nabla \cdot (\sigma(\mathbf{x})\nabla u) = q(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{8a}$$

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

$$\frac{\partial u}{\partial n} = 0, \quad \mathbf{x} \in \partial\Omega. \tag{8b}$$

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. The PDE (8) is discretized on a staggered grid as described in [26][3].

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

$$\begin{aligned} \psi(\tau) = \psi(\tau; \theta, \alpha_1, \alpha_2) &= \alpha \tanh\left(\frac{\tau}{\alpha\theta}\right) + \frac{\alpha_1 + \alpha_2}{2}, \\ \alpha &= \frac{\alpha_2 - \alpha_1}{2}, \end{aligned} \tag{9}$$

consider one of two scenarios:

1. In practice, often there are reasonably tight bounds available, say $\sigma_{\min}$ and $\sigma_{\max}$, such that $\sigma_{\min} \leq \sigma(\mathbf{x}) \leq \sigma_{\max}$. Such information may be enforced using (9) by defining

$$\sigma(\mathbf{x}) = \psi(m(\mathbf{x})), \quad \text{with } \psi(\tau) = \psi(\tau; 1, \sigma_{\min}, \sigma_{\max}). \tag{10}$$

2. Occasionally it is reasonable to assume that the sought conductivity function $\sigma(\mathbf{x})$ takes only one of two values, σ_I or σ_{II}, at each $\mathbf{x}$. As in [7, 8, 9, 26] we use an approximate level set function representation, writing $\sigma(\mathbf{x}) = \lim_{h\to 0} \sigma(\mathbf{x}; h)$, where

$$\sigma(\mathbf{x}; h) = \psi(m(\mathbf{x}); h, \sigma_I, \sigma_{II}). \tag{11}$$

   The function $\psi(\tau; h)$ depends on the resolution, or grid width h. As $h \to 0$ the sought function $m(\mathbf{x})$ satisfying

$$\begin{aligned} \nabla \cdot (\psi(m(\mathbf{x}))\nabla u_i) &= q_i, \quad i = 1, \ldots, s, \\ \frac{\partial u_i}{\partial n}\Big|_{\partial\Omega} &= 0, \end{aligned} \tag{12}$$

   has bounded first derivatives, whereas $\sigma(\mathbf{x})$ is generally discontinuous.

Next, we use a modified GN method precisely as described in [26]. We omit repeating that description here.

[3] We are citing descriptions in our previous work to avoid repetition. A verbose version of this reference can be found at `http://www.cs.ubc.ca/~ascher/papers/rodoas.pdf`

3 Data completion

Let Γ_i denote an open subset of the boundary, $\Gamma_i \subset \partial\Omega$, where the receivers for the i^{th} experiments are placed. Our goal here is to extend the data for each experiment to the union $\Gamma = \bigcup \Gamma_i \subseteq \partial\Omega$, the common measurement domain. In practice, the conductivity $\sigma(\mathbf{x})$ in (8a) is often piecewise smooth with finite jump discontinuities. As such one is faced with two scenarios: (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 u on Γ. Hence, in this section we treat these cases separately and propose an appropriate data completion algorithm for each.

Consider the problem (8). In what follows we assume that Ω is a bounded open domain and $\partial\Omega$ is Lipschitz. Furthermore, we assume that there is 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 $0 < \alpha \leq 1$. The conductivity is also assumed to be distributed such that $\sigma \in C^2(\overline{\Omega}_j)$, $j = 1, \ldots, N$. 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 3.3.

Under these assumptions and for the Dirichlet problem with a $C^2(\partial\Omega)$ boundary condition, there is a constant γ, $0 < \gamma \leq 1$, 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, $0 < \beta \leq 1$. 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 [28]. So the results stated above would still hold for (8).

3.1 Discontinuities in conductivity are away from common measurement domain

Next, let $\Gamma \subset (\partial\Omega_j \cap \partial\Omega)$ 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 i^{th} experiment, we define our data completion function $v_i \in H^2(\Gamma)$ as the solution of a discretized version of the optimization problem

$$\min_v \frac{1}{2} \|v - u_i\|^2_{L_2(\Gamma_i)} + \lambda \|\Delta v\|^2_{L_2(\Gamma)}, \tag{13}$$

where λ is the regularization parameter: it can depend on the amount of noise in our data and on how much smoothness we would like to impose on the approximated forward problem solution. The discretization of the first term in (13) involves a restriction which selects those given data points that fall inside Γ_i, whereas that of the regularization functional involves a standard uniform mesh as specified in Section 4.

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. Note that the extrapolated points at the bottom left corner of the graph are also recovered to a pleasing accuracy.

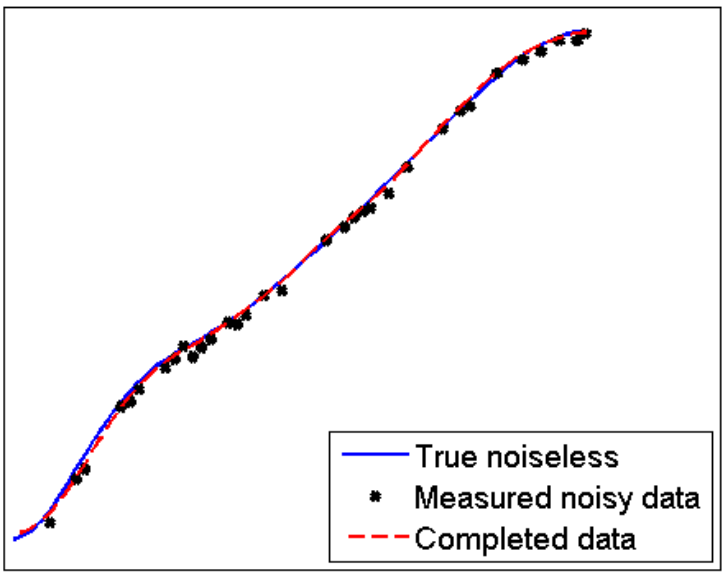


Figure 1: Completion using the regularization (13). This plot corresponds to an experiment taken from Example 3 where 50% of the data requires completion and 5% noise has been added.

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

This situation corresponds to the case in which Γ 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 $\{\Gamma \cap (\partial\Omega \cap \partial\Omega_j)^\circ \ , \ j \in \mathcal{J}\}$ forms a set of disjoint subsets of Γ such that $\overline{\Gamma} = \bigcup_{j\in\mathcal{J}} \overline{(\partial\Omega \cap \partial\Omega_j)^\circ}$, 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 is no longer necessarily in $H^2(\Gamma)$. Hence, the smoothing term in (13) is no longer valid, as $\|\Delta u\|_{L_2(\Gamma)}$ 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

$u \in C^{\beta}(\overline{\Omega})$. Proposition 1 below shows that the smoothness on Γ is not completely gone: we may lose one degree of regularity at worst.

Proposition 1 *Let U and $\{U_j |\ j = 1, 2, \cdots K\}$ 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 ∂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 i^{th} 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

$$v(\mathbf{x}) = \sum_{j=1}^{K}\partial_i u\mathcal{X}_{U_j}, \tag{14}$$

where $\mathcal{X}_{U_j}$ denotes the characteristic function of the set U_j. So we get

$$\int_U u\partial_i\phi = -\int_U v\phi. \tag{15}$$

We also have

$$\|v\|_{L_2(U)} \leq \sum_{j=1}^{K}\|\partial_i u\|_{L_2(U_j)} < \infty, \tag{16}$$

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

Returning to the problem on hand, we can see that if the assumptions stated at the beginning of this section hold, we can expect a field $u \in H^1(\Gamma)$. This is obtained by invoking Proposition 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 that of finding $v = v_i$ which solves the discretized version of the optimization problem

$$\min_v \frac{1}{2} \|v - u_i\|^2_{L_2(\Gamma_i)} + \lambda \|\nabla v\|^2_{L_2(\Gamma)}, \tag{17}$$

where λ is as in Section 3.1.

Figure 2 shows an example of such data completion using the formulation (17). The true solution and the measured data correspond to an experiment using the model in Example 4 of Section 4. We only plot the profile of v_i on the top boundary. Since the discontinuities in the conductivity σ extend to the measurement domain, we can see that the field is now only piecewise smooth yet continuous. The approximation process imposes less smoothness along the boundary as compared to (13), and this results in an excellent completion of the missing data. Note that the extrapolated points at the bottom left corner of the graph are also recovered to a pleasing accuracy.

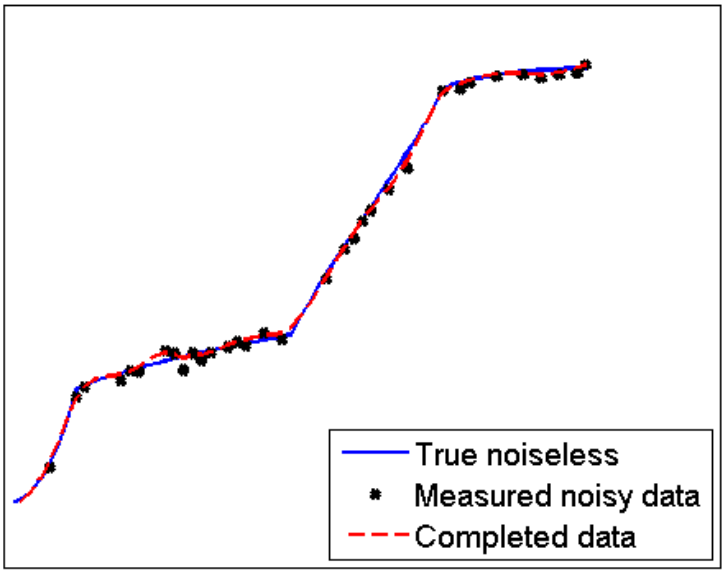


Figure 2: Completion using the regularization (17). This plot corresponds to an experiment taken from Example 4 where 50% of the data requires completion and 5% noise has been added. Discontinuities in the conductivity extend to the measurement domain and their effect on the profile of the field along the boundary can be clearly observed.

After solving the discretized (13) or (17) we have the complete discretized data, $\mathbf{v}_i$, on the entire measurement domain for all the experiments $i = 1, \ldots, s$. So, we can subsequently assume having $P = P_i \; \forall i$.

Regarding the cost of carrying out our data completion strategy, we note that strictly speaking there are s ℓ_2-optimization problems to be solved in (13) or (17). Fortunately, these problems can be easily and efficiently solved once discretized. Moreover, this is a preprocessing stage which is completed before the algorithm for solving the nonlinear inverse problem commences, so it is carried out once and for all, not anew at each GN iteration. 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 grid points are 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).

Now we can use one of the algorithms proposed in [26] in order to perform the inversion using the simultaneous sources. However, there is one difficulty which arises as a by-product of the data completion step: since the noise is also interpolated, the initial noise statistics is lost and there is no clear way to determine the noise tolerance which may be used to stop the iterative algorithm. One easy way around this is to evaluate the misfit for the stopping criterion on the initial, given data with the original noise level.

3.3 Point sources and boundaries with corners

In the numerical examples of Section 4, as in [9, 26], 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 (8a) 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 general setup of our numerical experiments is the same as that in [26]: we consider a unit square or unit cube for the domain Ω, and 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 (8a) 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, we place the corresponding sink on the right boundary 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. 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 (8) 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.

In the following numerical examples, we compare the performance of Algorithm 2 in [26] using the completed data against the same algorithm using the Random Subset method defined there. As advocated in [26], we use Gaussian distribution for the mixing weights. In both cases, we use Random Subset in the uncertainty check step, and the stopping criterion is verified using the original uncompleted data.

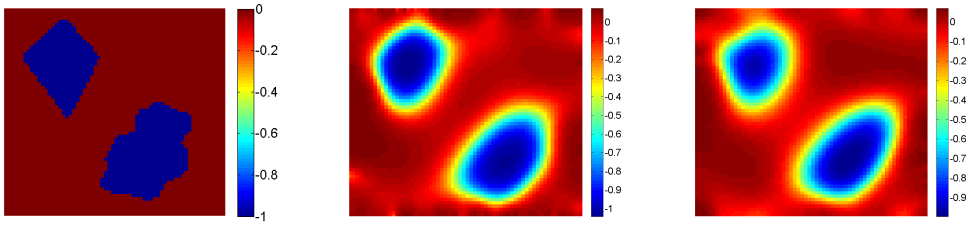
For all of our numerical experiments, the "true field" is calculated on a grid that is twice as fine as the one used to reconstruct the model. For the 2D examples, the reconstruction is done on a uniform grid of size 64^2 with $s = 961$ experiments in the setup described above. For our 3D examples, the reconstruction is done on a uniform grid of size 17^3.

Numerical examples are presented for both cases described in Sections 3.1 and 3.2. For all of our numerical examples except the last one, we use the transfer function (10) 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, we set the PCG iteration limit to $r = 20$, and the PCG tolerance to 10^{-3}. The initial guess is $\mathbf{m}_0 = \mathbf{0}$. Our last example is done using the level set method (11). Here we can set $r = 5$, significantly lower than above. The initial guess for the level set example is also as in [26].

In addition to displaying the log conductivities (i.e., $\log(\sigma)$) for each reconstruc-

tion, we also show the log-log plot of misfit on the entire data (i.e., $\|F(\mathbf{m}) - D\|_F$) vs. PDE count, and the relative ℓ_2 error (i.e., $\frac{\|\sigma^* - \sigma\|}{\|\sigma^*\|}$) 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. In each of the numerical examples corresponding to the scenarios explained in Sections 3.1 and 3.2, by "Completed Data" we mean using Algorithm 2 in [26] with completed data employing regularizations (13) and (17), respectively.

Example 1 *In this example, we place two target objects of conductivity* $\sigma_I = 0.1$ *in a background of conductivity* $\sigma_{II} = 1$*, and* 3% *noise is added to the data. Also,* 20% *of the data requires completion. The discontinuities in the conductivity are some distance away from the common measurement domain so we can use* (13) *to complete the data.*



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

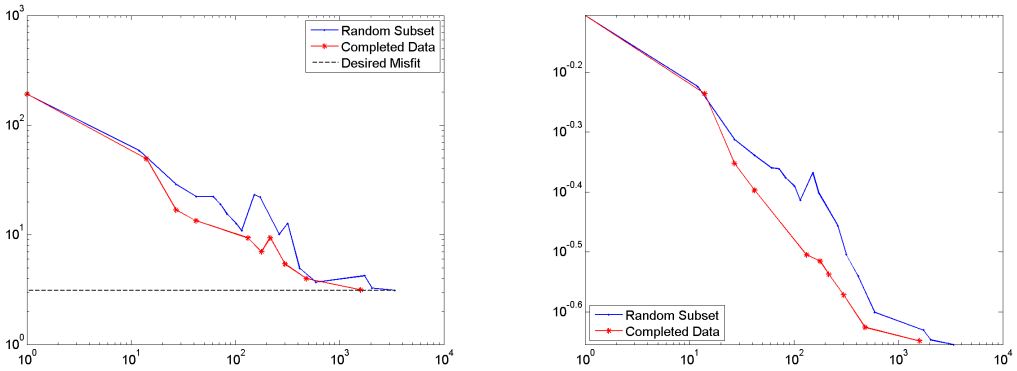
Figure 3: Example 1 – reconstructed log conductivity with 20% data missing and 3% noise: (a) true solution, (b) Random Subset, (c) Data Completion. Regularization (13) has been used to complete the data.

Method	Random Subset	Data Completion
Work	3,367	1,597

Table 1: Work in terms of number of PDE solves for Example 1.

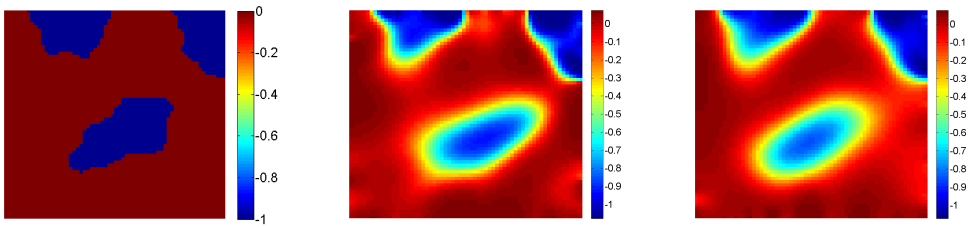
Figures 3 and 4, as well as Table 1, reflect superiority of the simultaneous source method combined with data completion over the Random Subset method.

Example 2 *This is the same as Example 1, except that the discontinuities in the conductivity are touching the measurement domain. Therefore, we use the regularization* (17) *to complete the data.*



(a) Data misfit (b) Relative ℓ_2 error

Figure 4: Data misfit and relative ℓ_2 error vs. PDE count for Example 1.



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

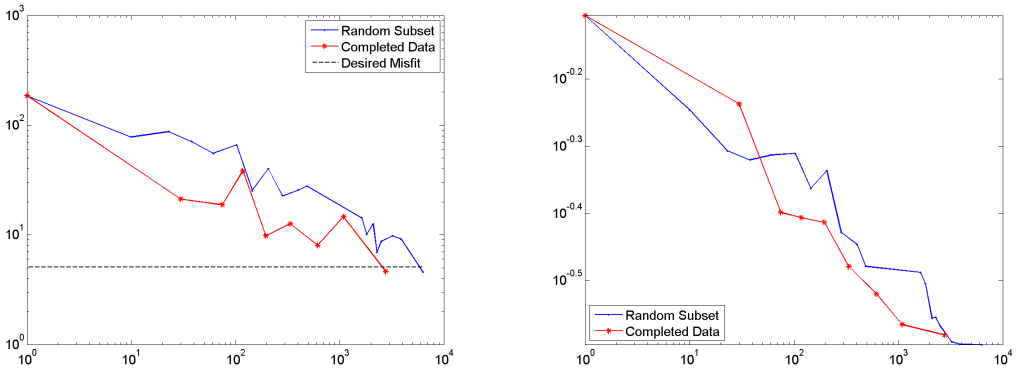
Figure 5: Example 2 – reconstructed log conductivity with 20% data missing and 3% noise: (a) true solution, (b) Random Subset, (c) Data Completion. Regularization (17) has been used to complete the data.

Method	Random Subset	Data Completion
Work	6,302	2,769

Table 2: Work in terms of number of PDE solves for Example 2.

From Tables 1 and 2, we see that the reconstruction with completed data and the simultaneous sources method can be done more efficiently (by a factor of more than two), compared to using the Random Subset method. The quality of reconstruction is also very good. Note that the graph of the misfit for Data Completion lies entirely under that of the Random Subset method. This means that, given a fixed number of PDE solves, we obtain a lower (thus better) misfit for Data Completion as opposed to Random Subset.

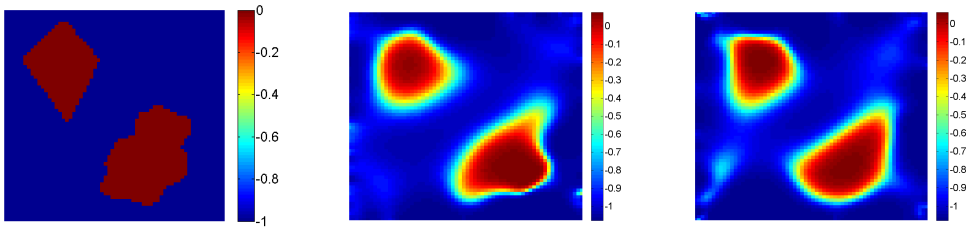
Example 3 *For this example, we merely swap the conductivities of Example 1, and increase the added noise to* 5%. *The proportion of the data that requires completion*



(a) Data misfit (b) Relative ℓ_2 error

Figure 6: Data misfit and relative ℓ_2 error vs. PDE count for Example 2.

is increased to 50%. *We again use the regularization* (13) *to complete the data. The results are gathered in Table 3 and Figures 7 and 8.*



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

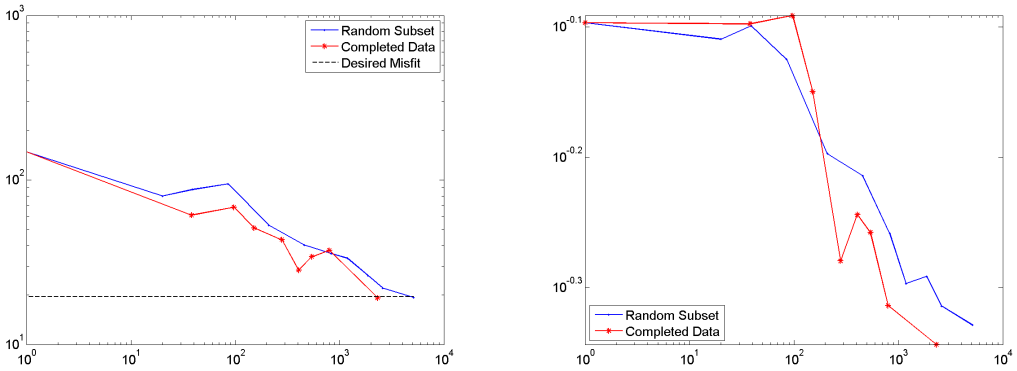
Figure 7: Example 3 – reconstructed log conductivity with 50% data missing and 5% noise: (a) true solution, (b) Random Subset, (c) Data Completion. Formulation (13) has been used to complete the data.

Method	Random Subset	Data Completion
Work	5,139	2,320

Table 3: Work in terms of number of PDE solves for Example 3.

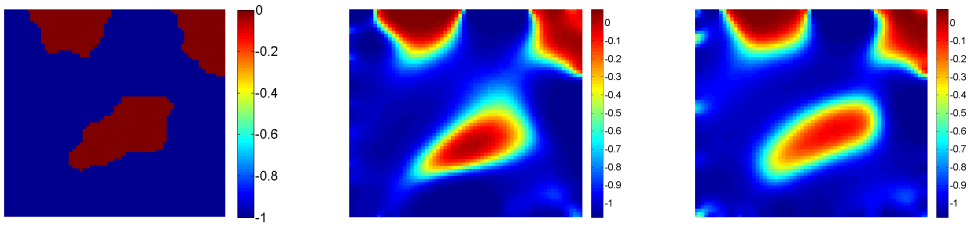
Similar observations to those for the previous examples hold here, too.

Example 4 *For this example, we use the same setup as the previous example, but with the discontinuities of the conductivity now touching the common measurement domain. We use* (17) *to complete the data. The results are gathered in Table 4 and Figures 9 and 10.*



(a) Data misfit (b) Relative ℓ_2 error

Figure 8: Data misfit and relative ℓ_2 error vs. PDE count for Example 3.



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

Figure 9: Example 4 – reconstructed log conductivity with 50% data missing and 5% noise: (a) true solution, (b) Random Subset, (c) Data Completion. Formulation (17) has been used to complete the data.

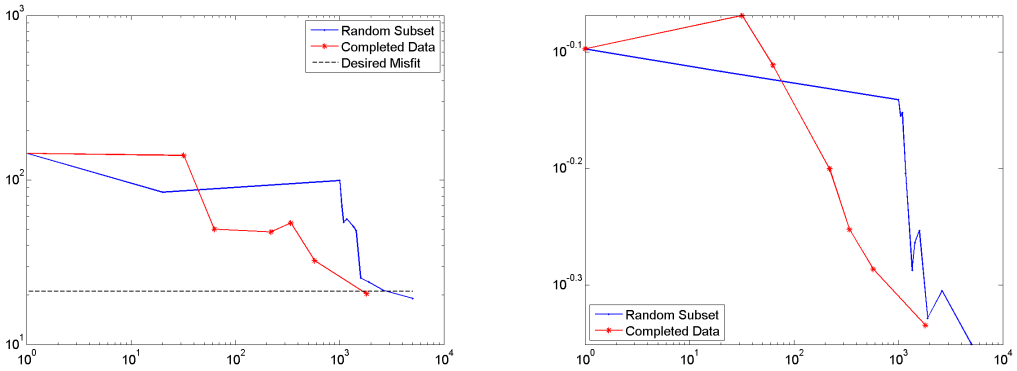
Method	Random Subset	Data Completion
Work	5,025	1,818

Table 4: Work in terms of number of PDE solves for Example 4.

Based on Tables 3 and 4, we can make the same observations as in Examples 1 and 2 in terms of total work required: a similar speedup factor of more than 2 in the number of PDE solves is observed. The quality of reconstruction is again very good even in the case where half of the measurement domain needs to be completed.

Example 5 *In this 3D example, we place a target object of conductivity* $\sigma_I = 1$ *in a background with conductivity* $\sigma_{II} = 0.1$*, see Figure 11, and* 2% *noise is added to the "exact" data. Then we knock out* 50% *of the data and use* (13) *to complete the data.*

Results are gathered in Figures 12 and 13 as well as Table 5. The data completion plus simultaneous sources algorithm again does well, with an efficiency factor ≈ 2.75.



(a) Data misfit (b) Relative ℓ_2 error

Figure 10: Data misfit and relative ℓ_2 error vs. PDE count for Example 4.

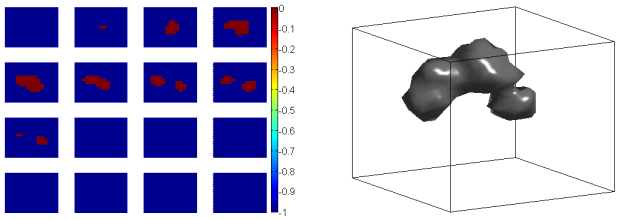


Figure 11: True Model for Example 5.

Method	Random Subset	Data Completion
Work	11,620	4,243

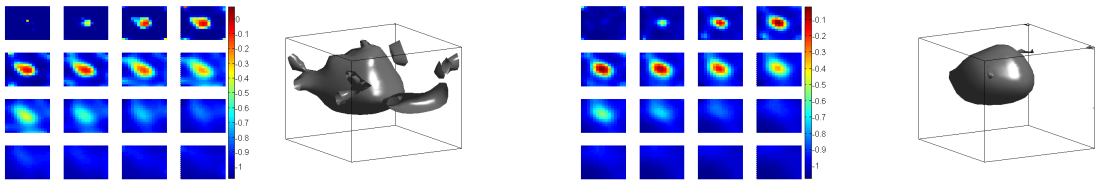
Table 5: Work in terms of number of PDE solves for Example 5.

Example 6 *Here the discontinuities in the conductivity extend all the way to the common measurement domain, see Figure 14, and we need to use (17) 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 30% of the data.*

Method	Random Subset	Data Completion
Work	9,093	1,268

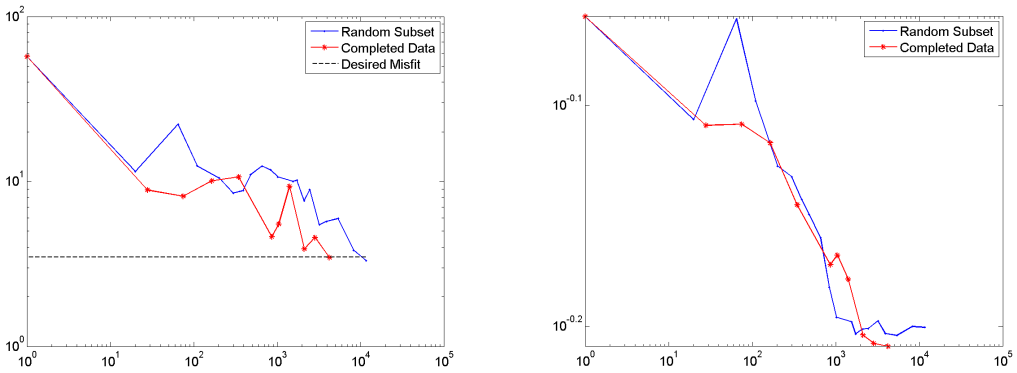
Table 6: Work in terms of number of PDE solves for Example 6.

It can clearly be seen from the results of Examples 5 and 6 that Algorithm 2 of [26] does a great job recovering the model using the completed data plus the



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

Figure 12: Example 5 – reconstructed log conductivity for the 3D model with (a,b) Random Subset, (c,d) Data Completion for the case of 2% noise and 50% data missing. Regularization (13) has been used to complete the data.



(a) Data Misfit (b) Relative ℓ_2 Error

Figure 13: Data misfit and relative ℓ_2 error vs. PDE count for Example 5.

simultaneous sources method as compared to Random Subset with uncompleted data. This is so both in terms of total work and the quality of the recovered model. One possible explanation for this is that after completing the data using the appropriate formulation, we inject some added information into the whole data set. If this extra information is valid then we are in fact heading for a better reconstruction.

Example 7 *This is the same as Example 6 except that we use the level set transfer function (11) to reconstruct the model. The results are recorded in Figures 17 and 18 as well as Table 7.*

Method	Random Subset	Data Completion
Work	3,671	935

Table 7: Work in terms of number of PDE solves for Example 7.

It is evident from Figure 17 that employing the level set formulation allows a

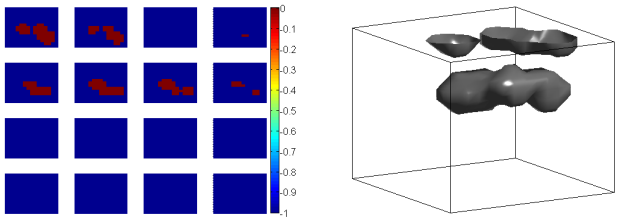
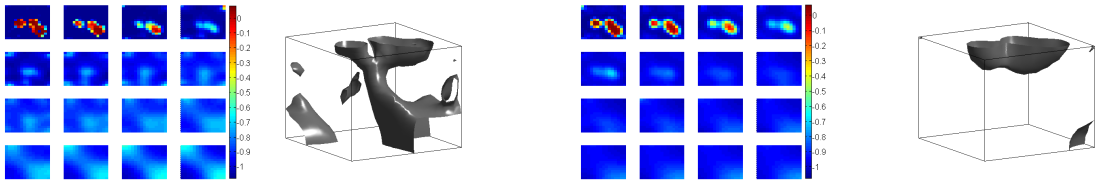
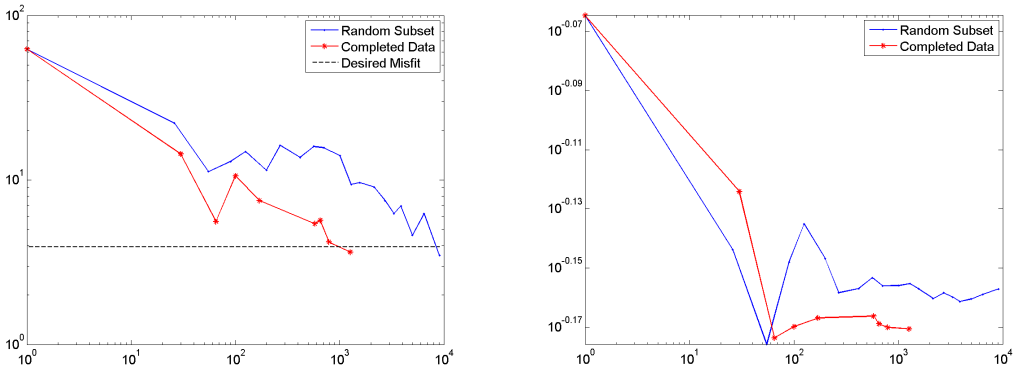


Figure 14: True Model for Example 6.



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

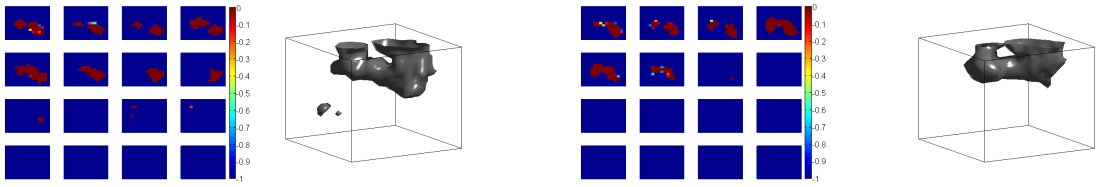
Figure 15: Example 6 – reconstructed log conductivity for the 3D model with (a,b) Random Subset, (c,d) Data Completion for the case of 2% noise and 30% of data missing. Regularization (17) has been used to complete the data.



(a) Data Misfit (b) Relative ℓ_2 Error

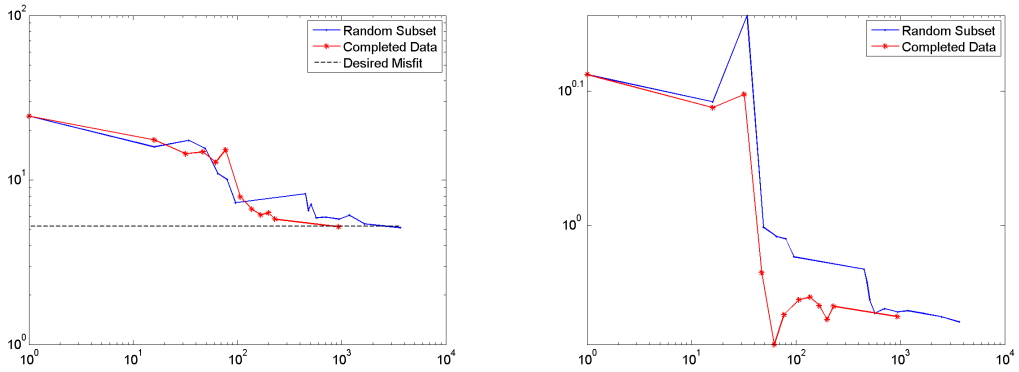
Figure 16: Data misfit and relative ℓ_2 error vs. PDE count for Example 6.

significantly better quality reconstruction than in Example 6. This is expected, as much stronger assumptions on the true model are employed. The algorithm proposed here produces a better reconstruction than the Random Subset one on the uncompleted data. The relative efficiency observation can be made from Table 7, where a factor of roughly 4 is obtained. It was shown in [7] and [26] that using level set functions can greatly reduce the total amount of work, and this is observed here as well.



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

Figure 17: Example 7 – reconstructed log conductivity for the 3D model using the level set transfer function (11) with (a,b) Random Subset, (c,d) Data Completion for the case of 2% noise and 30% of data missing. Regularization (17) has been used to complete the data.



(a) Data Misfit (b) Relative ℓ_2 Error

Figure 18: Data misfit and relative ℓ_2 error vs. PDE count for Example 7.

5 Conclusions and future work

This paper is a sequel to [26] 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 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 on the domain boundary. The resulting completed data allows an efficient use of the methods in [26]. Our approach shows great success in cases of mild to moderate data completion, say up to 50%. In such cases, we have demonstrated that, utilizing Algorithm 2 of [26], an execution speedup factor of 2 and possibly much more can be achieved while obtaining excellent reconstructions.

Let us caution that for severe cases of missing data, say 70% 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 simply use the Random Subset approach as described in [9]: although this method is slower, the reconstructed model can be more reliable.

We have tested several other regularized approximations on the set of examples of Section 4, including data interpolation using piecewise linear functions, and several DCT, wavelet and curvelet approximations. The first of these has the advantage of being very simple, while for the others we had hoped to leverage the recent advances in compressive sensing and sparse ℓ_1 methods. However, of all these, the methods using the formulations (13) and (17) were found to perform best in the context described in Sections 3 and 4.

The specific data completion techniques proposed in this paper were 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. Study of such extensions to other PDE models could be also another subject for future studies.

Acknowledgments

The authors would like to thank Drs. Adriano De Cezaro and Eldad Haber for several fruitful discussions.

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

[5] 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.

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

[7] 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.

[8] K. van den Doel and U. Ascher. Dynamic level set regularization for large distributed parameter estimation problems. *Inverse Problems*, 23:1271–1288, 2007.

[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; 2nd edition, 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, K. van den Doel, and U. Ascher. Stochastic algorithms for inverse problems involving pdes and many measurements. 2012. submitted.

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

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

[29] 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.

[30] 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.

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