Inverse Problems and Imaging
Volume 8, No. 1, 2014, 199–221
doi:10.3934/ipi.2014.8.199

HEAT SOURCE IDENTIFICATION BASED ON ℓ_1 CONSTRAINED MINIMIZATION

Yingying Li and Stanley Osher
University of California, Los Angeles
Los Angeles, CA 90095, USA

Richard Tsai
The University of Texas at Austin
Austin, TX 78712, USA

(Communicated by Zuowei Shen)

Abstract. We consider the inverse problem of finding sparse initial data from the sparsely sampled solutions of the heat equation. The initial data are assumed to be a sum of an unknown but finite number of Dirac delta functions at unknown locations. Point-wise values of the heat solution at only a few locations are used in an ℓ_1 constrained optimization to find the initial data. A concept of domain of effective sensing is introduced to speed up the already fast Bregman iterative algorithm for ℓ_1 optimization. Furthermore, an algorithm which successively adds new measurements at specially chosen locations is introduced. By comparing the solutions of the inverse problem obtained from different number of measurements, the algorithm decides where to add new measurements in order to improve the reconstruction of the sparse initial data.

1. **Introduction.** Heat source identification problems have important applications in many branches of engineering and science. For example, an accurate estimation of a pollutant source [7, 12] is a crucial environmental safeguard in cities with dense populations. Typically, a recovery of the unknown source is a reverse process in time. The major difficulty in establishing any numerical algorithm for approximating the solution is the severe ill-posedness of the problem. It appears that the mathematical analysis and numerical algorithms for inverse heat source problems are still very limited. For the kind of problem we consider in this paper, where we want to find the initial condition with known measurements in the future time, existing methods either need many measurements [5] or have stability issues [13]. In this paper, we treat the source identification problem as an optimization problem. Our goal is to invert the heat equation to get the sparse initial condition. In other words, the problem can be formulated as an ℓ_0 minimization problems with PDE constraints.

It is difficult to solve the ℓ_0 problem since it is a nonconvex and NP-hard problem. In compressed sensing [6], we can solve an ℓ_0 problems by solving its ℓ_1 relaxation

2010 *Mathematics Subject Classification.* Primary: 49N45, 65M32.
Key words and phrases. $L1$, source identification, heat equation.
Li and Osher are partially supported by NGA HM1582-07-C-0010P000003, NIH 5U54RR021813-05, and ARO MURI subcontract from Rice University and from University of South Carolina. Osher is partially supported also by NSF grants DMS 0914561 and 118971. Tsai is partially supported by a Moncrief Grand Challenge Award, NSF grants DMS-0914840 and DMS-0914465.

when the associated linear operator has the restricted isometry property (RIP) [4]. The heat operator does not satisfy RIP, but we can adopt the idea of substituting ℓ_0 with ℓ_1 for sparse optimization. We will show numerical evidence that indicates the effectiveness of this strategy. To solve a constrained ℓ_1 minimization problem we apply the Bregman iterative method [1, 18], which solves the constrained problem as a sequence of unconstrained subproblems. To solve these subproblems, we use the greedy coordinate descent method developed in [11], which was shown to be very efficient for sparse recovery.

Since the theory of compressive sensing does not apply to the heat operator, it is unclear if constrained ℓ_1 minimization provides a good solution to our problem. However, this also means that there is room for finding specialized measurement locations for better solutions to the inverse problem. Hence, in this paper we attempt to understand the following questions:

- Is ℓ_1-regularization adequate for inverse problems involving point sources?
- In which way can additional data improve the inversion?

In related work, the author [9] discussed optimal experimental design for ill-posed problems and suggested a numerical framework to efficiently achieve such a design in a statistical manner. In [2], the authors used reciprocity and maximum principle for discovery of a single point source in partially known environments. In [10], the authors considered point source discovery for the Helmholtz equation with partially known obstacles. There, the authors introduced an L_1 optimization algorithm for reconstructing incoming wave fronts at measurement locations, and an imaging functional to image point sources. In the same paper, the authors also proposed an algorithm to successively explore the partially known domain in order to discover the point sources. In [8], the authors applied and generalized the reciprocity algorithm of [2] for multiple point source discovery from measurements which comes from line integrals of solution to an atmospheric model. Finally, in [15], the authors generalized a reciprocity approach for multiple point source discovery for nonlinear systems of advection-diffusion-reaction equations.

The paper is organized as follows. In section 2, we give a more detailed introduction of the heat and related source identification problems. A useful stability estimate for a simple case is obtained in section 2.4. In section 3, we present our algorithm for solving the heat source identification problem and some methods for improving the efficiency. The performace of the algorithm is evaluated in the numerical experiments in section 3.5 in the case of two spatial dimensions. In section 4 we consider the successive sampling. Finally, section 5 summarizes and discusses future directions. The details of the proof of the stability estimate are given in appendix A.

The main contributions of this paper are:

- Using ℓ_1 minimization for heat source identification.
- Proving the stability estimate in terms of Wasserstein distance for a simple case.
- Introducing a successive sampling strategy.
- Proposing the ideas of exclusion region and support restriction for reducing the problem size.

2. Source problems.

2.1. 1D heat equation.

We consider first a simple case of a heat equation in 1D with periodic boundary conditions

$$\begin{cases} u_t(x,t) = \Delta u(x,t), & x \in (0,1),\ t > 0, \\ u(x,0) = u_0(x), & x \in [0,1], \\ u(0,t) = u(1,t), & t > 0. \end{cases} \tag{1}$$

The initial condition $u_0(x)$ is assumed to be sparse in the sense that

$$u_0(x) = \sum_{k=1}^{K} \alpha_k \delta(x - s_k), \tag{2}$$

where $\alpha_k > 0$ and $\delta(x - s_k)$ are Dirac δ-functions concentrated at location s_k.

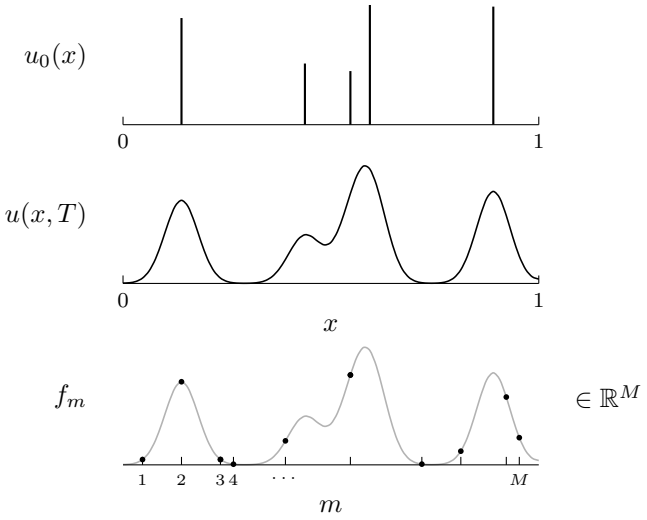


FIGURE 1. Heat source identification problem in 1D: given the samples $f_m = u(x_m, T)$, $m = 1, \ldots, M$, recover the sparse initial condition u_0.

The heat source identification problem that we consider is the following: if we observe (possibly noisy) measurements $f_m = u(x_m, T)$, $m = 1, \ldots, M$, then without knowing K, α_k, or s_k in advance, can we recover u_0?

We propose to recover the point sources via discretizing (1) and looking for sparse solutions of the discretized problem. We proceed as follows: partition $[0, 1]$ into N elements so that u_0 is approximated by vector $v \in \mathbb{R}^N$. Let G denote a linear solution operator of the discretized problem; i.e. $w = \mathsf{G}v$ solves the discretized problem and approximates the solution of (1). The discrete solution w is sampled by a linear operator $\mathsf{S} : \mathbb{R}^N \to \mathbb{R}^M$. For example, in this paper, we take the pointwise measurements of the solution:

$$f_m = w_{j(m)}, \quad m = 1, \ldots, M, \tag{3}$$

where $w_{j(m)}$ is the $j(m)$th component of the vector w which approximates $u(x_m, T)$. So we can write $\mathsf{S}w = \mathsf{S}\mathsf{G}v = f$ with $f = (f_1, \cdots, f_M)^T \in \mathbb{R}^M$.

Thus, we may pose the heat source identification problem in the optimization framework as

$$\arg\min_{v\in\mathbb{R}^N} \|v\|_0 \quad \text{subject to} \quad \mathsf{S}\mathsf{G}v = f. \tag{4}$$

Since this problem is known to be NP-hard, we ultimately replace it by the proposed recovery problem:

$$\arg\min_{v\in\mathbb{R}^N} \|v\|_1 \quad \text{subject to} \quad |\mathsf{S}\mathsf{G}v - f| < \epsilon, \tag{5}$$

where $\epsilon > 0$ is a small parameter which may be related to the level of noise or errors in the measurements.

While there is no continuum analogue of (4), one may also consider a continuum regularized L^1 optimization problem, an analogue of (5):

$$\arg\min_{v\in L_1(\Omega)} \|v\|_{L^1(\Omega)} \quad \text{subject to} \quad |\mathcal{S}\mathcal{G}v - f| < \epsilon, \tag{6}$$

where $\mathcal{G}$ is the solution operator of (1), which generates the solution at time T, starting from initial condition v, and $\mathcal{S}$ is the linear sampling operator at points x_m, $m = 1, \ldots, M$. Obviously, G and S are the discrete approximations of $\mathcal{G}$ and $\mathcal{S}$.

2.2. **General linear parabolic equations.** Our approach applies more generally to problems where u_0 is sparse and linearly related to the known measurements f_m. Let us consider on a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$ a parabolic problem of the form

$$\begin{cases} \partial_t u = \sum_{i,j} \partial_{x_i}\big(a_{i,j}(x)\partial_{x_j} u\big) + \sum_i b_i(x)\partial_{x_i} u + c(x)u + g(x,t) \\ u(x,0) = u_0(x) \end{cases} \tag{7}$$

with periodic or Neumann boundary conditions, and sparse initial condition of the form (2) with $s_k \in \Omega$. We suppose that a, b, c, and g satisfy appropriate conditions, so that $u(x,t)$ belongs to a suitable (linear) function space $\mathcal{F}$ on $\Omega \times [0,T]$. The solution is sampled by a linear operator $\mathcal{S} : \mathcal{F} \to \mathbb{R}^M$, for example given by

$$f_m = u(x_m, t_m), \qquad m = 1, \ldots, M. \tag{8}$$

Other interesting choices include sampling the derivative values $\partial_{x_i} u(x_m, t_m)$ or some weighted local averages $(\varphi * u)(x_m, t_m)$. Then, the heat source identification problem if to recover u_0 given $f = \mathcal{S}(u)$, assuming that all other information (Ω, a, b, c, g, $\mathcal{S}$, f) is known.

As in the previous section, we discretize (7) on a grid and formulate the corresponding ℓ_1 minimization problem for source recovery. With the same notation for the discrete solution operator and sampling operator, we pose:

$$\arg\min_{v\in\mathbb{R}^N} \|v\|_1 \quad \text{subject to} \quad |\mathsf{S}(\mathsf{G}v + u_p) - f| < \epsilon. \tag{9}$$

Here, u_p refers to a particular solution to the inhomogeneous equation, and it can be constructed quite easily through the application of Duhamel's principle. If we denote by G_t the solution operator of (7) with $g \equiv 0$, then Duhamel's principle gives the solution to (7) with any g as

$$u(x,t) = G_t u_0 + \int_0^t G_{t-s} g(x,s) ds. \tag{10}$$

Therefore, at the discrete level, we may take u_p to be an approximation of the second term above.

Finally, we remark that with a similar formulation, we can also solve source identification problems where g takes the form

$$g(x) = \sum_{k=1}^{K} \alpha_k \delta(x - s_k).$$

2.3. **Sparsity in a transformed domain.** Rather than considering u_0 itself as being sparse, we can also consider u_0 as being sparse when represented in some basis or frame. For example, a function like

$$u_0(x) = \sum_{k=1}^{K} c_k \cos(s_k x) + d_k \sin(s_k x)$$

has a sparse Fourier representation. If u_0 is piecewise smooth, it has an approximately sparse wavelet representation, see Figure 2. Let $\mathcal{R}$ be a linear operator (e.g., inverse Fourier or inverse wavelet transform) such that $u_0 = \mathcal{R}\hat{u}_0$ for some sparse function $\hat{u}_0$. Then (9) becomes

$$\arg\min_{\hat{u}_0 \in \ell_0} \|\hat{u}_0\|_0 \quad \text{subject to} \quad f = \mathcal{S}(\mathcal{G}\mathcal{R}\hat{u}_0), \tag{11}$$

where $\mathcal{G}$ is the solution operator of (7).

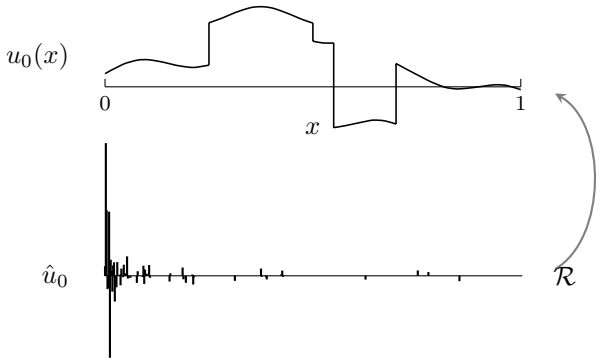


FIGURE 2. Our approach also applies when u_0 is sparse under a transformed representation. Here we show a piecewise smooth function and its Cohen-Daubechies-Feauveau 9/7 wavelet transform.

2.4. **Inequalities.** There is no rigorous proof to ensure that ℓ_1 minimization for inverting the (discretized) heat operator enhances the sparsity. However, the following inequalities, proven in the one dimensional, continuous setting (6), suggest that it is true under some limited conditions. In one dimensional space, if the true solution has only one spike (i.e. positively weighted Dirac delta function), the minimizer will be very close to the true solution under a Wasserstein distance. Wasserstein distance is designed to compare probability measures, and it comes from the theory of optimal transport. Under proper normalization, it may be used to compare sparse initial data that we consider in this paper. In one dimension, Wasserstein distance that we use can be defined and evaluated easily as follows. Suppose that

f and g are nonnegative functions on $[0,1]$ with $\int f\,dx = \int g\,dx$, and let F and G denote their primitives,

$$F(x) := \int_0^x f(t)\,dt, \quad G(x) := \int_0^x g(t)\,dt. \tag{12}$$

Then the (first) Wasserstein distance between f and g is

$$W_1(f,g) := \int_0^1 |F(x) - G(x)|\,dx. \tag{13}$$

The inequalities shown below imply a sense of stability of the solution's spike locations. Intuitively, two spikes that are close under the Wasserstein distance will also be close after the heat diffusion process. In our future work, we hope to verify this intuition for more general cases.

Theorem 2.1. *Suppose that $u(x) = \alpha\delta(x - s_1)$, where $\alpha > 0$. Let x_j denote the sampling locations, $j = 1, 2, \ldots, J$, and $f_j = (\mathcal{G}u)(x_j, T)$ denote the measurements taken at these locations. Suppose further that and $S = [x_2 - \sqrt{2T}, x_1 + \sqrt{2T}]$, $x_1 < s_1 < x_2$ and $x_2 - x_1 < \sqrt{2T}$. For any v of the form*

$$v(x) = \sum_j \beta_j \delta(x - \tilde{s}_j) \text{ and } \hat{f}_j = (\mathcal{G}v)(x_j, T). \tag{14}$$

satisfying $\beta_j > 0$, $\|v\|_1 \le \|u\|_1$, and $\|\hat{f} - f\|_\infty \le \epsilon$, there exist $C' > 0$ and $C'' > 0$ such that

$$1 \ge \frac{\sum_{j:\tilde{s}_j \in S} \beta_j}{\alpha} \ge 1 - C'\epsilon, \tag{15}$$

$$\frac{\sum_{j:\tilde{s}_j \in S} \beta_j |\tilde{s}_j - s_1|^2}{\alpha} \le C''\epsilon. \tag{16}$$

We present a proof of the above theorem in the appendix A.

We can derive some simple conclusion from the above theorem: when the true sparse solution has only one spike, the recovery obtained by ℓ_1 minimization should be close to the true solution. They are close in ℓ_1 norm and under Wasserstein distance after normalization.

Theorem 2.2. *Suppose $u^\star = \alpha\delta(x - s_1)$, $\mathsf{S}\mathsf{G}u^\star = f_0$, $\|f_0 - f\|_\infty \le \epsilon$, and*

$$v = \arg\min_u \|u\|_1 \text{ s.t. } \|\mathsf{S}\mathsf{G}u - f\|_\infty \le \epsilon. \tag{17}$$

Let x_j, $j = 1, 2, \ldots, J$ denote the sampling locations and suppose there are two samples x_1 and x_2 such that $x_1 < s_1 < x_2$ and $x_2 - x_1 < \sqrt{2T}$. Then

$$\|u^\star\|_1 - \|v\|_1 \le C_1\epsilon \tag{18}$$

and there are $S \subseteq [s_1 - \sqrt{2T}, s_1 + \sqrt{2T}]$ and $C_2 > 0$, such that

$$W_1\left(\tfrac{\|v_S\|_1}{\alpha} u^\star - v_S\right) \le C_2\sqrt{\epsilon}. \tag{19}$$

Proof. Since both $u^\star$ and v satisfy the constraint, but v is the minimizer, so $\|v\|_1 \le \|u^\star\|_1$, and

$$\|\mathsf{S}\mathsf{G}u^\star - \mathsf{S}\mathsf{G}v\|_\infty \le \|\mathsf{S}\mathsf{G}u^\star - f\|_\infty + \|f - \mathsf{S}\mathsf{G}v\|_\infty \le 2\epsilon. \tag{20}$$

Using Theorem 2.1, there are $S \subseteq [s_1 - \sqrt{2T}, s_1 + \sqrt{2T}]$ and $C > 0$, such that

$$\frac{\sum_{j:\tilde{s}_j \in S} \beta_j |\tilde{s}_j - s_1|}{\alpha} \le C\sqrt{2\epsilon}. \tag{21}$$

Denote $v|_S = v_S$, then the Wasserstein distance between $\frac{\|v_S\|_1}{\alpha}u^\star$ and v_S is

$$W_1\left(\frac{\|v_S\|_1}{\alpha}u^\star - v_S\right) = \frac{\sum_{\tilde{s}_j\in S}|\tilde{s}_j - s_1|}{\alpha} \le C\sqrt{2\epsilon}. \tag{22}$$

□

3. **Solving the ℓ_1 minimization problem.** While (4) is a natural way to pose the problem, it is hard to solve. There are two challenges in solving (4). First, the ℓ_0-norm is nonconvex, thus the existence and uniqueness of solutions are not guaranteed, and on a practical level, the nondifferentiability of the ℓ_0-norm precludes the use of gradient-based minimization methods. Second, inverting the matrix $\mathsf{A} = \mathsf{SG}$ is an ill-conditioned process since heat diffusion may make two different initial conditions appear increasingly similar over time, hence the solution is extremely sensitive with respect to perturbations of the measurements.

We attempt to overcome these challenges by replacing ℓ_0 by ℓ_1,

$$\arg\min_v \|v\|_1 \quad \text{subject to} \quad \mathsf{A}v - f = 0. \tag{23}$$

By convexity of the ℓ_1-norm, solutions of (23) exist. As demonstrated in the compressive sensing literature, the ℓ_1-norm tends to favor sparse solutions and makes for an effective approximation of ℓ_0. Furthermore, it has been shown that under some general conditions [4], ℓ_0 minimization and ℓ_1 minimization yield the same solution—though unfortunately, this theory does not apply to (23).

In the following sections, we discuss the solution of (23) using the Bregman iteration algorithm.

3.1. **Bregman iteration.** Bregman iterative techniques minimize the problems of the form

$$\arg\min_u J(u) \quad \text{subject to} \quad H(u) = 0 \tag{24}$$

or

$$\arg\min_u J(u) \quad \text{subject to} \quad H(u) \le \epsilon \tag{25}$$

where J is convex and H is convex and differentiable on a Hilbert space $\mathcal{H}$ when $\min_u H(u) = 0$.

Define the *Bregman distance* as

$$D_J^p(u,\tilde{u}) = J(u) - J(\tilde{u}) - \langle p, u-\tilde{u}\rangle_{\mathcal{H}}, \quad p \in \partial J(\tilde{u}). \tag{26}$$

Note that this is not a distance in the usual sense as it is not symmetric. The constrained minimization (24) is solved by the *Bregman iteration algorithm*:

$$\begin{cases} \text{Initialize: } u^0 = \mathbf{0}, p^0 = \mathbf{0} \\ \textbf{for } k = 0, 1, \dots \\ \quad u^{k+1} = \arg\min_u D_J^{p^k}(u, u^k) + \lambda H(u) \\ \quad p^{k+1} = p^k - \lambda \nabla H(u^{k+1}) \end{cases} \tag{27}$$

where λ is a positive parameter. For our application the objective is $J(u) = \|u\|_1$ and the constraint is $H(u;f) = \frac{1}{2}\|\mathsf{A}u - f\|_2^2$. In this case Bregman iteration algorithm takes the form

$$\begin{cases} \text{Initialize: } u^0 = \mathbf{0},\ p^0 = \mathbf{0} \\ \textbf{for } k = 0, 1, \dots \\ \quad u^{k+1} = \arg\min_u \|u\|_1 - \langle p^k, u\rangle + \frac{\lambda}{2}\|\mathsf{A}u - f\|_2^2 \\ \quad p^{k+1} = p^k - \lambda\mathsf{A}^*(\mathsf{A}u^{k+1} - f) \end{cases} \tag{28}$$

Equivalently, by refactoring $\langle p^k, u\rangle + \lambda\|\mathsf{A}u - f\|_2^2$, the sequence $\{p^k\}$ is concisely expressed as adding the residuals to f:

$$\begin{cases} \text{Initialize: } u^0 = \mathbf{0},\ f^0 = f \\ \textbf{for } k = 0, 1, \dots \\ \quad u^{k+1} = \arg\min_u \|u\|_1 + \frac{\lambda}{2}\|\mathsf{A}u - f^k\|_2^2 \\ \quad f^{k+1} = f^k + (f - \mathsf{A}u^{k+1}) \end{cases} \tag{29}$$

The Bregman iteration algorithm can be stopped for example when the $\|u^{k+1} - u^k\|$ is less than a chosen tolerance. Similarly, to solve the minimization problem with an inequality constraint like $\|\mathsf{A}u - f\|_2^2 \le \epsilon$, the algorithm should be stopped for the first k such that $\left\|\mathsf{A}u^k - f\right\|_2^2 \le \epsilon$.

The following general properties of the Bregman iterative algorithm are proved in [16]:

Theorem 3.1. (Bregman iteration properties)

1. *Monotonic decrease in H:*
$$H(u^{k+1}) \le H(u^{k+1}) + D_J^{p^k}(u^{k+1}, u^k) \le H(u^k);$$
2. *Convergence to the exact minimizer of H: If $\tilde{u}$ minimizes $H(\cdot)$ and $J(\tilde{u}) < \infty$, then $H(u^k) \le H(\tilde{u}) + J(\tilde{u})/k$;*
3. *Convergence with noisy data: Let $H(\cdot) = H(\cdot; f)$ and suppose $H(\tilde{u}; f) \le \epsilon$ and $H(\tilde{u}; g) = 0$; then $D_J^{p^{k+1}}(\tilde{u}, u^{k+1}) < D_J^{p^k}(\tilde{u}, u^k)$ as long as $H(u^{k+1}; f) > \epsilon$.*

3.2. **Shrinkage.** The Bregman iterative algorithm allows us to solve the constrained minimization problem (23) by solving a sequence of unconstrained problems,

$$\arg\min_u \|u\|_1 + \tfrac{\lambda}{2}\|\mathsf{A}u - f^k\|_2^2. \tag{30}$$

The one-dimensional subproblem has an efficient closed-form solution.

Consider the one-dimensional case where u is a scalar, then it is easy to solve the problem

$$u^\star = \arg\min_{u\in\mathbb{R}} |u| + \tfrac{\lambda}{2}(u - f)^2. \tag{31}$$

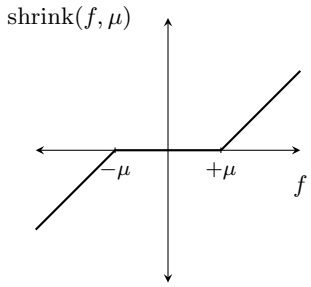
The solution to (31) is obtained by *shrinkage*, also known as soft thresholding [17]:

$$u^\star = \text{shrink}(f, \tfrac{1}{\lambda}) \equiv \text{sign}(f)\big(|f| - \tfrac{1}{\lambda}\big)^+. \tag{32}$$

The shrink operator is illustrated in Figure 3.

Similarly if u is constrained to be nonnegative, the scalar problem is

$$\arg\min_{u\ge 0} u + \tfrac{\lambda}{2}(u - f)^2,$$



$$\operatorname{shrink}(f,\mu) \equiv \operatorname{sign}(f)\big(|f|-\mu\big)^+ = \begin{cases} f+\mu & \text{if} \quad f \le -\mu, \\ 0 & \text{if} \quad -\mu \le f \le +\mu, \\ f-\mu & \text{if} \quad +\mu \le f. \end{cases}$$

FIGURE 3. Shrinkage operator.

and the minimizer is given by $u^\star = (f - \frac{1}{\lambda})^+$.

In the multidimensional case where u is a vector, the Bregman subproblem (30) is a lot more difficult to solve. In particular, we lose the explicit expression for the solution. Instead, we can apply the coordinate descent method developed in [11] to solve

$$\arg\min_u \|u\|_1 + \lambda\|\mathsf{A}u - f^k\|_2^2.$$

Since we ultimately seek a sparse solution, the process of finding the solution should give preference to sparsity. Instead of proceeding through all the coordinates, we choose only to update coordinates most likely to be the spikes and decrease the energy the most. Therefore, we choose a greedy coordinate algorithm which was introduced in [11].

Algorithm (*Greedy Coordinate Descent*):

Precompute: $w_j = \|a_j\|_2^2$;
  Normalization: $\mathsf{A}(\cdot,i) = \mathsf{A}(\cdot,i)/w_i$;
  Initialization: $u^0 = 0$, $\beta^0 = \mathsf{A}^* f$;

Iterate until converge:
  $\tilde{u} = \operatorname{shrink}(\beta^k, \frac{1}{2\lambda})$;
  $j = \arg\max_i |u_i^k - \tilde{u}_i|$,
  then $u_i^{k+1} = u_i^k$, $i \neq j$,
  $u_j^{k+1} = \tilde{u}_j$;
  $\beta^{k+1} = \beta^k - |u_j^k - \tilde{u}_j|(\mathsf{A}^*\mathsf{A})e_j$,
  $\beta_j^{k+1} = \beta_j^k$.

In the algorithm, the computation of $\tilde{u}$ and β is essential. To obtain $\tilde{u}$, the shrinkage formula with $O(N)$ complexity can be used; for efficiency, β should be updated recursively by adding the difference between two iterations. Every step in the loop has complexity $O(N)$, so combined with its preference for sparsity, this algorithm is very efficient for our problem.

3.3. **Support restriction.** We have two strategies to accelerate the solutions to the heat source identification problem. The first idea is to solve for u_0 only on its apparent support.

Empirically, we observe that early iterations of Bregman iterations tend to produce blurry approximations of the solution and later iterations sharpen the initial approximation into spikes. Therefore, in solving for u^{k+1}, it is reasonable to expect

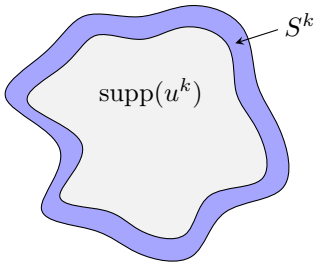


FIGURE 4. An illustration of a dilation of $\mathrm{supp}(u^k)$

that $\mathrm{supp}(u^{k+1})$ is similar to $\mathrm{supp}(u^k)$. Let S^k be a set containing $\mathrm{supp}(u^k)$ and solve for u^{k+1} with its support restricted to S^k,

$$u^{k+1} = \arg\min\{\|u\|_1 + \lambda \|\mathsf{A}u - f\|_2^2 : \mathrm{supp}(u) \subset S^k\}. \tag{33}$$

For example, S^k may be a morphological dilation of $\mathrm{supp}(u^k)$. It is important that S^k is strictly larger than $\mathrm{supp}(u^k)$ to prevent the iteration from getting trapped within an incorrect support. If we find a solution which is also the minimizer on its dilated support, then this solution is a local minimizer and a global minimizer due to the convexity.

The set S^k has to include $\mathrm{supp}(u^k)$ as a closed subset. In our numerical examples, we enlarge $\mathrm{supp}(u^k)$ by including all its connected neighbors in the discretized sense. That is, we increase $\mathrm{supp}(u^k)$ by one pixel in each direction. Then S^k is the smallest set including $\mathrm{supp}(u^k)$ as a closed subset in the discretized sense.

3.4. **Domain exclusion.** The second idea is to eliminate a region from consideration when a measurement is very small. Suppose that the stengths of the sources in u_0 are bounded from below by $\alpha_{\min} > 0$. Then, since A is nonnegative, this implies

$$\begin{aligned} f_m \equiv (\mathcal{G}u_0)(x_m, t_m) = \int G_{t_m}(x_m, y) u_0(y)\, dy &= \sum_{k=1}^{K} \alpha_k G_{t_m}(x_m, s_k) \\ &\geq \alpha_{\min} \sum_{k=1}^{K} G_{t_m}(x_m, s_k). \end{aligned} \tag{34}$$

Thus for a spike to exist at location s, we must have $f_m \geq \alpha_{\min} G_{t_m}(x_m, s)$ for all m. The contrapositive of this statement gives a way to identify regions of the domain that cannot have spikes:

$$\Omega_z \equiv \bigcup_{m=1}^{M} \{s \in \Omega : f_m < \alpha_{\min} G_{t_m}(x_m, s)\}. \tag{35}$$

Similarly, for noisy measurements $|f_m^{\text{exact}} - f_m^{\text{noisy}}| \leq \epsilon$ we have

$$\Omega_z \equiv \bigcup_{m=1}^{M} \{s \in \Omega : f_m^{\text{noisy}} + <_{\varepsilon} \alpha_{\min} G_{t_m}(x_m, s)\}. \tag{36}$$

Note that the validity of this strategy requires A to be nonnegative. Otherwise, cancelations could occur such that the bound (34) does not hold.

For the periodic boundary conditions with the point-value sampling $f_m = u(x_m, t_m)$, the exclusion condition simplifies to $f_m < \alpha_{\min} G_{t_m}(x_m - s)$, see Figure 5.

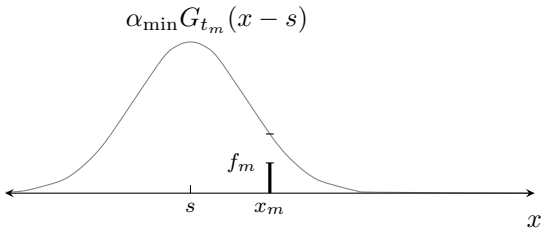


FIGURE 5. Domain exclusion for the case of periodic boundary conditions with point-value sampling: a small measurement $f_m < \alpha_{\min} G_{t_m}(x_m - s)$ implies that there cannot be a spike at $x = s$.

When support restriction and domain exclusion are added, the Bregman iterative algorithm (29) takes the following form.

$$
\left\{
\begin{array}{l}
\text{Use (35) to determine } \Omega_{\mathrm{z}} \\
\text{Initialize: } u^0 = \mathbf{0},\ f^0 = f, S^0 = \Omega \backslash \Omega_{\mathrm{z}} \\
\textbf{for } k = 0, 1, \ldots \\
\quad u^{k+1} = \arg\min\{\|u\|_1 + \frac{\lambda}{2}\|\mathsf{A}u - f^k\|_2^2 : \mathrm{supp}(u) \subset S^k\} \\
\quad f^{k+1} = f^k + \lambda \mathsf{A}^*(\mathsf{A}u - f^k) \\
\quad S^{k+1} = \big(B \oplus \mathrm{supp}(u^{k+1})\big) \backslash \Omega_{\mathrm{z}}
\end{array}
\right. \tag{37}
$$

Here $B\oplus$ denotes dilation by a structure element B. In the examples below, B is the 3×3 structure element

$$
B = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \tag{38}
$$

3.5. **Numerical examples.** In this section we show the numerical results for source identification using Bregman iterative algorithms applied to problems (6) or (9). The solutions are computed on a computer with Intel Dual-Core T9550 2.66GHz CPU and 4GB RAM.

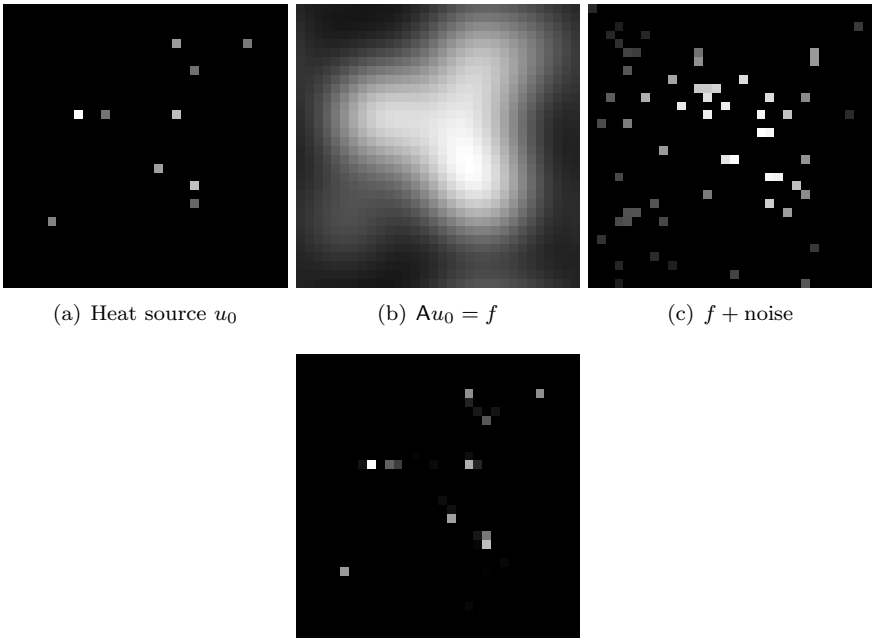
3.5.1. *Source identification for the heat equation.* In the first example we conside the heat equation

$$
\left\{
\begin{array}{ll}
u_t = u_{xx} + u_{yy}, & t > 0 \\
u_0 = \sum_k c_k \delta(x - x_k, y - y_k), & t = 0
\end{array}
\right. \tag{39}
$$

on the unit square $(0,1) \times (0,1)$ with periodic boundary conditions. We observe M point-value samples $f_m = u(x_m, y_m, T)$, $m = 1, \ldots, M$, of the solution at a final time $T = 0.01$. The heat source identification problem is to recover u_0 from these observations.

We discretize the domain $(0,1)\times(0,1)$ using a uniform $N \times N$ grid. The following approximation for the δ-function is used

$$
\delta(x - x_k, y - y_k) = \left\{
\begin{array}{ll}
N^2, & \text{if } (x,y) = (x_k, y_k) \\
0, & \text{otherwise.}
\end{array}
\right. \tag{40}
$$



(a) Heat source u_0 (b) $\mathsf{A}u_0 = f$ (c) f + noise

(d) recovered u_0

FIGURE 6. Recovery of the heat source u_0 from 60 randomly selected measurements with 1% noise on a 32×32 grid. Runtime: $24.6s$.

We introduce G to be an $N^2 \times N^2$ matrix such that $\mathsf{G}u_0$ is a finite difference approximation of $u(x, y, T)$,

$$(\mathsf{G}u_0)_k \approx u(x_k, y_k, T).$$

The constraint matrix A is formed by selecting the rows of G corresponding to the observation points (x_m, y_m). In other words, the observation vector f is a downsampled solution on the whole grid $f = \mathsf{S}(\mathsf{G}u_0)$, where $\mathsf{S} \in \mathbb{R}^{M \times N^2}$ is the downsampling operator.

Figure (6) shows an experiment using the Bregman iteration algorithm from the previous section to recover the sparse u_0 by solving the problem:

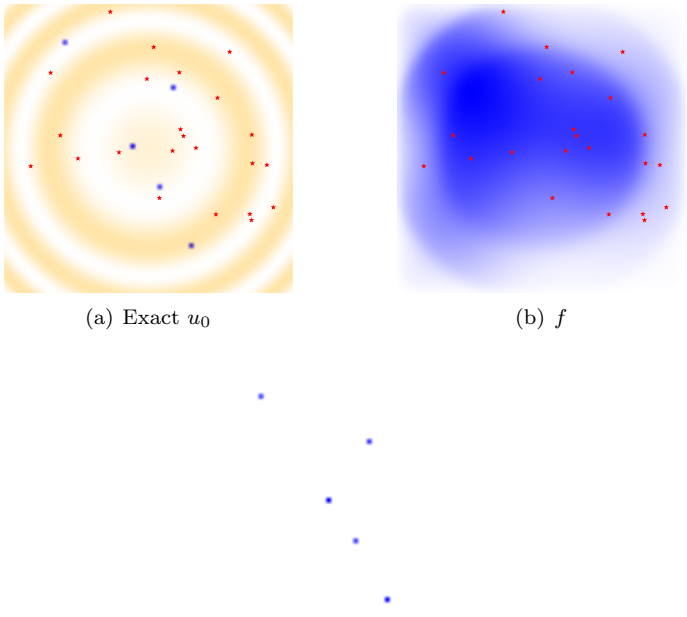
$$\min_u \|u\|_1 \qquad \text{subject to } \mathsf{A}u = f, \tag{41}$$

where $\mathsf{A} = \mathsf{SG}$.

3.5.2. *Source identification with spatially varying conductivity.* In case of a spatially varying thermal conductivity $a(x, y)$, we consider a parabolic equation

$$\begin{cases} u_t = \operatorname{div}\big(a(x,y)\nabla u\big) & (x,y) \in (0,1), t > 0, \\ u = \sum_k c_k \delta(x - x_k, y - y_k) & t = 0, \end{cases}$$

with Neumann boundary conditions. Similarly to the case of an ordinary heat equation, we sample u at time $T = 0.01$ and try to recover the initial condition using compressed sensing. In Figures 7 and 8 we show the results of source recovery for smooth and piecewise constant thermal conductivities respectively.



(a) Exact u_0 (b) f

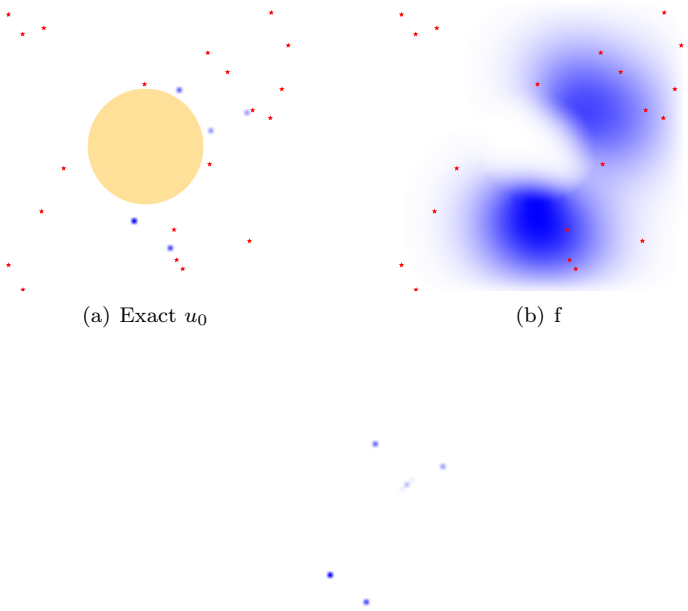
(c) Recovered u_0

FIGURE 7. Source recovery with a smooth spatially varying thermal conductivity. Left: distribution of $a(x, y)$ (shades of orange); sampling locations are red stars, heat source locations are blue dots. Middle: heat distribution at time T (shades of blue). Right: recovered source. Runtime: 14.4s.

4. **Successive sampling.** In the previous sections we developed a way to solve a heat source identification problem from a fixed set of observations. Therefore, we considered the random sampling scenario, as it is better suited for the compressed sensing setting. However, taking random observations may not be the best strategy for the heat source identification. Random sampling works well for compressed sensing if some incoherence is present in the operator $\mathcal{G}$. In our case the coherence in $\mathcal{G}$ is strong, because of the smoothness of the solution of the heat equation. This suggests that more structured observations may work better than random sampling. For example, if we happened to know a region of very low heat distribution, then it is certain that it is impossible to have strong heat sources there. When we are choosing our sample locations, we may want to concentrate in the strong heat distribution area or explore unsampled areas. Therefore, if we have a chance to pick the next sample location, we should consider the existing information instead of picking a random location.

Here we consider solving the source identification problem in an adaptive or online kind of approach according to the following procedure. We want to come up with a better sampling strategy than random sampling. Since we want the adopt the existing information for picking the next sampling location, the whole process for solving our problem is the following:

1. Solve the heat source identification problem with k samples;



(a) Exact u_0 (b) f

(c) Recovered u_0

FIGURE 8. Source recovery with a piecewise constant spatially varying thermal conductivity. Left: distribution of $a(x, y)$ (shades of orange); sampling locations are red stars, heat source locations are blue dots. Middle: heat distribution at time T (shades of blue). Right: recovered source. Runtime: 15.8s.

2. Use the solution u^k to select a $(k+1)$th sample;
3. Iterate.

Let us give a mathematical statement of this problem: Let $X_k = (x_1, x_2, \ldots, x_k)$, $T_k = (t_1, t_2, \ldots, t_k)$ and the measurements $f_j = f(x_j, t_j)$, $j = 1, 2, \cdots, k$. We denote $F_k = f = (f_1, f_2, \ldots, f_k)^T$, and $A_k : \mathbb{R}^N \to \mathbb{R}^k$, satisfies $A_k u = F_k$. We denote the solution from k measurements by u^k.

Suppose the spike amplitudes are bounded from below by $\alpha_{\min} > 0$, which is a plausible assumption since we can treat small spikes as noise but not real heat sources. Define the *covering region of* x as the set

$$\mathcal{C}(x) = \{y \in \Omega : \mathcal{G}(\alpha_{\min}\delta_y)(x) \geq \text{threshold}\}.$$

This set describes an effective domain of dependence of $u(x, T)$. We define a way to measure to what extend a point x is covered by samples x_j,

$$V(x) = \mathcal{G}\big(\textstyle\sum_j \delta(x - x_j)\big).$$

It is equivalent to placing a single heat source on all sample locations as an initial condition, then computing the total heat distribution. The bigger $V(x)$ is, the more information is available at x.

To choose the next sample location, there are two competing objectives: to refine locally or to explore further. Our approach is to prioritize local refinement.

Local refinement is needed since we want to improve the resolution if we discover a possible heat source cluster. Also, we want to explore further to enlarge the effective coverage as a necessary stopping condition.

The local resolution can be improved as follows. If u^k varies significantly from u^{k-1}, then we conclude that both u^k and u^{k-1} are not close to the true solution. Thus, we need more information to identify the heat source inside the existing covering region. So we choose the next sampling location x_{k+1} by comparing the difference between the two solution u^k and u^{k-1}, and picking the location where they differ the most. We define the $(k+1)$th sampling location x_{k+1} as

$$x_{k+1} = \arg \max_{x \,:\, x \notin B_r(x_j)} |G_\sigma * u^k - G_\sigma * u^{k-1}|,$$

where G_σ is a Gaussian with variance σ. The role of G_σ is to act as a smoother, and we typically choose σ to be small. The balls $B_r(x_j)$ with center x_j and radii r are introduced to exclude small regions around the existing samples.

In the other case, we are satisfied with the heat sources found inside the existing covering region. We then want to discover heat sources outside of the existing covering region. Therefore, we sample outside the covering region by selecting a point where V has minimal magnitude. The $(k+1)$th sample location x_{k+1} is

$$x_{k+1} = \arg \min_{x \,:\, x \notin B_r(x_j)} |V(x)|.$$

Compared to a random sampling, these two criteria approach the heat source faster and without wasting the samples in the region which cannot contain heat sources.

4.1. **Numerical experiments.** For the recovery with successive sampling we consider the 2D heat flow over a unit square with periodic boundaries and an initial condition u_0 shown in Figure 9. The sources in this example are grouped into two relatively well separated clusters. The cluster in the upper left is formed by three nearby point sources with variable strengths, and the cluster in the lower right has two point sources. The difficulty of the source recovery in this case lies in the need to not only resolve the neighboring sources, but also in detecting the clusters. The measurements are taken from $f = u(\cdot, T)$ which is also shown in Figure 9.

In Figures 10 and 11 we show the recovered sources for a number of steps of the successive sampling algorithm. The sources are recovered from the noiseless data. The difference between the experiment in Figures 10 and 11 is in the sampling locations chosen initially. However, we observe that for both choices our procedure recovers the source successfully. We see that in both simulations, the algorithm automatically determines measurement locations that surround the two clusters of sources.

In Figure 12 we show the result of successive sampling with 0.1% Gaussian noise added. The recovery process is sensitive to higher level noise. This is illustrated in Figure 13, where a failed attempt to recover the source is shown.

We conclude by showing in Figure 14 the comparison between the proposed successive sampling strategy, the non-negative least squares and ℓ_1 minimization approaches. With the same number of random samples, the solutions of least squares and ℓ_1 minimization are not as accurate as the solution obtained with the successive sampling approach.

5. **Conclusion.** The heat source identification problem can be solved by ℓ_1 minimization. For two dimensions, numerical experiments suggest that we can recover

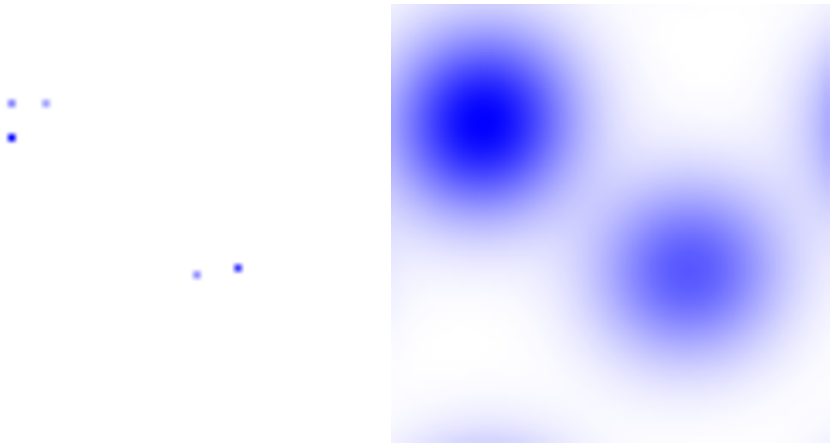


FIGURE 9. The source used in testing of the successive sampling approach. Left: initial condition u_0; right: heat distribution at time T.

the sparse initial condition by using 4 times more measurements than the number of total initial spikes. If we can solve in a successive manner, then we can use even fewer measurements. As for the stability of our method, as the noise increases, we need more measurements to obtain accurate solutions. In the future, we want to work on the error estimation and theoretical analysis. We are also interested in more general equations and high dimensional problems.

Acknowledgments. We thank Alexander Mamonov for helpful conversations and for his careful proof reading of the manuscript.

Appendix A. **Proof of Theorem 2.1.** In the following, we shall denote the heat kernel at time T as g for convenience; that is,

$$g(x) = \frac{1}{\sqrt{4\pi T}} \exp\left(-\frac{x^2}{4T}\right).$$

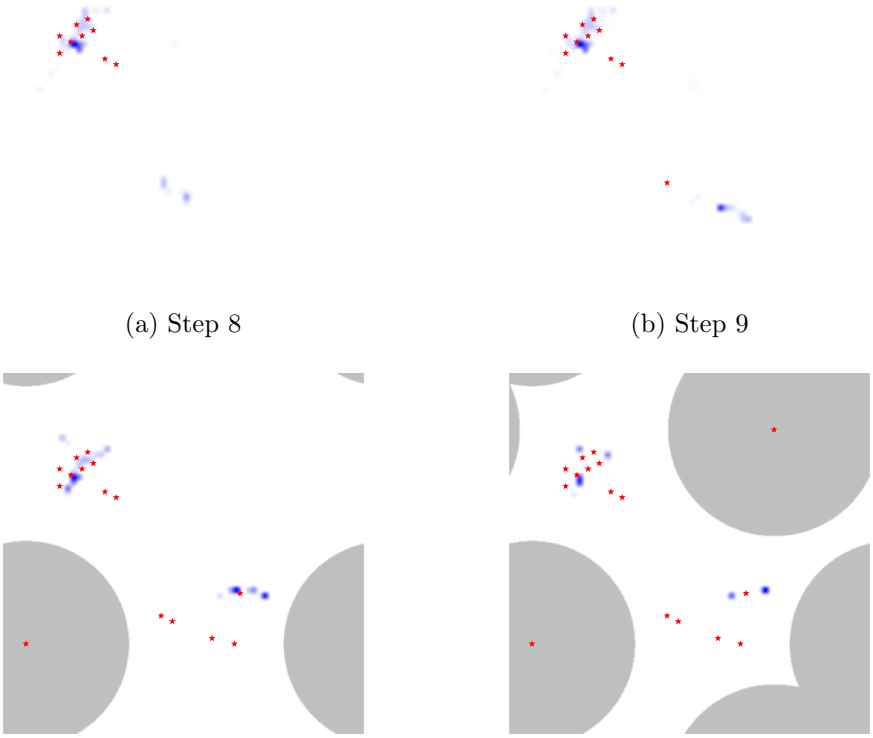
Lemma A.1. *Assume that $x_1 < s_1 < x_2$ and $x_2 - x_1 < \sqrt{2T}$.*

1. *The function $W(x) = -g'(x_2 - s_1)g(x - x_1) - g'(s_1 - x_1)g(x_2 - x)$ has only one maximum at $x = s_1$.*
2. *$W(s_1) - W(x) \geq C|s_1 - x|^2$ for $x_2 - \sqrt{2T} \leq x \leq x_1 + \sqrt{2T}$ for some constant $C > 0$.*
3. *If $x > x_1 + \sqrt{2T}$, then $W(s_1) - W(x) > W(s_1) - W(x_1 + \sqrt{2T})$.*
4. *If $x < x_2 - \sqrt{2T}$, then $W(s_1) - W(x) > W(s_1) - W(x_2 - \sqrt{2T})$.*

Proof. We first observe that s_1 is a critical point of W since $W'(s_1) = 0$. Since $g'(s_1 - x_1) < 0$ and $g'(x_2 - s_1) < 0$ for $x_1 < s_1 < x_2$, and $g''(x - x_1) < 0$ and $g''(x_2 - x) < 0$ if $x_2 - \sqrt{2T} < x < x_1 + \sqrt{2T}$, we have that

$$W''(s_1) = -g'(x_2 - s_1)g''(s_1 - x_1) - g'(s_1 - x_1)g''(x_2 - s_1) < 0$$

Thus, W reaches a local maximum at s_1. We will show that there are no other maxima.



(a) Step 8 (b) Step 9

(c) Step 14 (d) Step 15

Figure 10. Source recovery with successive sampling. The $(k+1)^{th}$ measurement is added in step k. The estimate of the source term is in blue, the exclusion region is in gray, sample locations are shown as red stars.

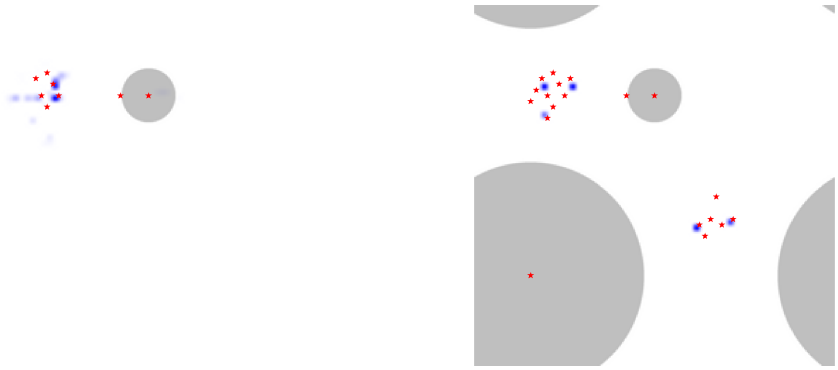
Performing Taylor expansion of W around s_1, we have

$$W(x) = W(s_1) + W''(\theta)(x-s_1)^2/2, \text{ where } \theta \in [x, s_1] \tag{42}$$

Denote $C = \inf_{\theta\in[x_2-\sqrt{2T},x_1+\sqrt{2T}]}\{-W''(\theta)/2\}$, then $C \geq 0$. Now we want to prove $C > 0$. Since $g'(x) = -(\frac{x}{2T})g(x)$ and $g''(x) = -\frac{1}{2T}(1-\frac{x^2}{2T})g(x)$, we can write

$$\begin{aligned} W''(x) = &\frac{((x-x_1)^2-2T)(x_2-s_1)}{8T^3}g(x-x_1)g(x_2-s_1) \\ &+ \frac{(s_1-x_1)((x_2-x)^2-2T)}{8T^3}g(x_2-x)g(s_1-x_1). \end{aligned}$$

Denoting $G_1 = \min\{g(\sqrt{2T}), g(x_2-x_1), g(x_1+\sqrt{2T}-x_2))\} > 0$, then for $x \in [x_2-\sqrt{2T}, x_1+\sqrt{2T}]$ and $x_1 < s_1 < x_2$, we have $g(x-x_1) \geq G_1$, $g(x_2-x) \geq G_1$,



(a) Step 7 (b) Step 17

FIGURE 11. Source recovery with successive sampling. Starting sampling locations differ from those in the previous example. The $(k+1)^{th}$ measurement is added in step k. The estimate of the source term is in blue, the exclusion region is in gray, sample locations are shown as red stars.

and with $G_2 = \min\{\frac{s_1-x_1}{4T^2}, \frac{x_2-s_1}{4T^2}\} > 0$ we obtain

$$\begin{aligned} -W''(x) &\geq G_1^2 G_2 \left(1 - \frac{(x-x_1)^2}{2T}\right) + G_1^2 G_2 \left(1 - \frac{(x_2-x)^2}{2T}\right) \\ &= G_1^2 G_2 \left(2 - \frac{(x_2-x)^2 + (x-x_1)^2}{2T}\right). \end{aligned}$$

Since $x \in [x_2 - \sqrt{2T}, x_1 + \sqrt{2T}]$, $(x_2-x)^2 + (x-x_1)^2 \leq 2T + (x_1 + \sqrt{2T} - x_2)^2$, hence

$$\begin{aligned} -W''(x) &\geq G_1^2 G_2 \left(2 - \frac{(x_2-x)^2 + (x-x_1)^2}{2T}\right) \geq G_1^2 G_2 \left(1 - \frac{(x_1 + \sqrt{2T} - x_2)^2}{2T}\right) \\ &> 0. \end{aligned}$$

This implies that $C > 0$. Combined with (42) this gives statement 2 of the Lemma.

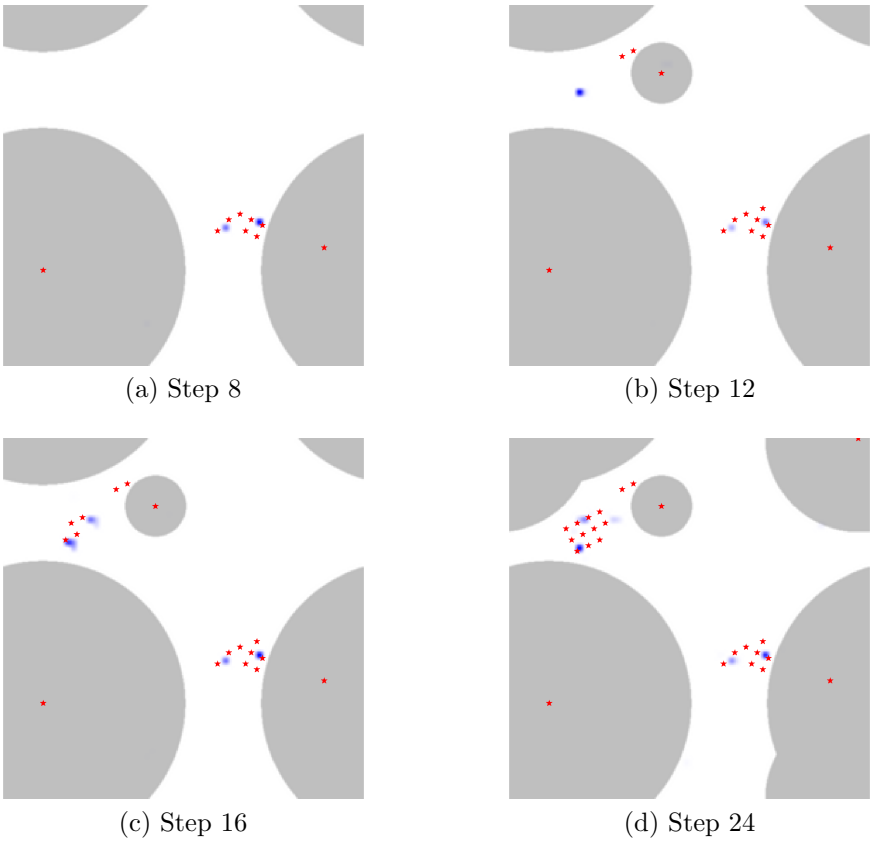
Finally, we consider statements 3 and 4. For $x > x_1 + \sqrt{2T}$, since $g(x)$ is decreasing as $x > 0$,

$$\begin{aligned} g(x-x_1) &< g(x_1 + \sqrt{2T} - x_1) = g(\sqrt{2T}), \\ g(x_2 - x) &= g(x - x_2) < g(x_1 + \sqrt{2T} - x_2). \end{aligned}$$

Therefore, $W(x) < W(x_1 + \sqrt{2T})$ for $x > x_1 + \sqrt{2T}$. In the same way, we can prove that $W(x) < W(x_2 - \sqrt{2T})$ for $x < x_2 - \sqrt{2T}$. □

Theorem A.1. *Suppose that $u(x) = \alpha\delta(x - s_1)$, where $\alpha > 0$. Let x_j denote the sampling locations, $j = 1, 2, \ldots, J$, and $f_j = (\mathcal{G}u)(x_j)$ denote the measurements taken from these locations. Suppose further that and $S = [x_2 - \sqrt{2T}, x_1 + \sqrt{2T}]$, $x_1 < s_1 < x_2$ and $x_2 - x_1 < \sqrt{2T}$. For any v of the form*

$$v(x) = \sum_j \beta_j \delta(x - \tilde{s}_j) \text{ and } \hat{f}(x_j) = (\mathcal{G}v)(x_j). \tag{43}$$



(a) Step 8 (b) Step 12

(c) Step 16 (d) Step 24

FIGURE 12. Successful source recovery with successive sampling in the presence of noise in the measurements. The $(k+1)^{th}$ measurement is added in step k. The estimate of the source term is in blue, the exclusion region is in gray, sample locations are shown as red stars.

satisfying $\beta_j > 0$, $\|v\|_1 \leq \|u\|_1$, *and* $\|\hat{f} - f\|_\infty \leq \epsilon$, *there exist* $C' > 0$ *and* $C'' > 0$ *such that*

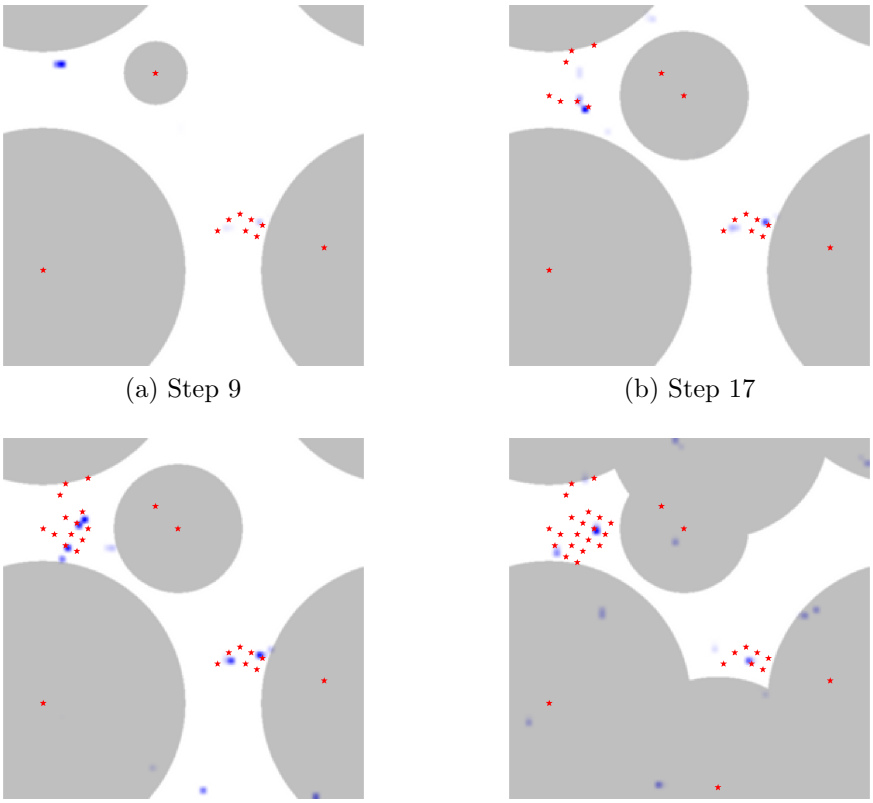
$$1 - C'\epsilon \leq \frac{\sum_{j:\tilde{s}_j \in S} \beta_j}{\alpha} \leq 1 \tag{44}$$

$$\frac{\sum_{j:\tilde{s}_j \in S} \beta_j |\tilde{s}_j - s_1|^2}{\alpha} \leq C''\epsilon. \tag{45}$$

Proof. The solution to the heat equation with initial data $\sum \delta(x-\sigma_j)$ is $\sum g(\sigma_j - x)$. Therefore, we have

$$\alpha W(s_1) = -g'(x_2 - s_1)f(x_1) - g'(s_1 - x_1)f(x_2),$$

$$\begin{aligned}\sum \beta_j W(\tilde{s}_j) &= \sum \beta_j(-g'(x_2 - s_1)g(\tilde{s}_j - x_1) - g'(s_1 - x_1)g(x_2 - \tilde{s}_j)) \\ &= -g'(x_2 - s_1)\hat{f}(x_1) - g'(s_1 - x_1)\hat{f}(x_2).\end{aligned}$$



(a) Step 9 (b) Step 17

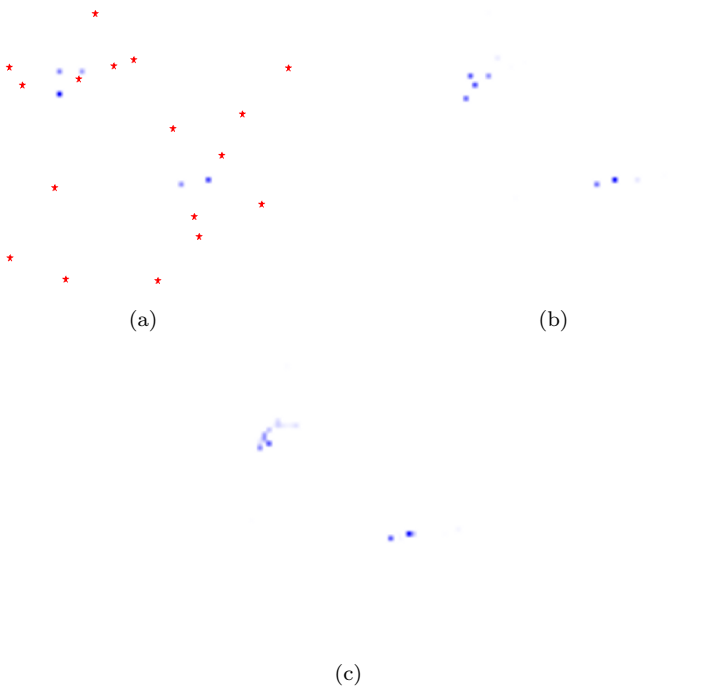
(c) Step 23 (d) Step 31

FIGURE 13. Failed source recovery attempt with successive sampling in the presence of noise in the measurements. The $(k+1)^{th}$ measurement is added in step k. The estimate of the source term is in blue, the exclusion region is in gray, sample locations are shown as red stars.

$$\implies \alpha W(s_1) - \sum \beta_j W(\tilde{s}_j) \leq (-g'(x_2 - s_1) - g'(s_1 - x_1))\epsilon = \tilde{C}\epsilon. \tag{46}$$

Suppose $\tilde{s}_1, \ldots \tilde{s}_l \leq x_2 - \sqrt{T}$; $\tilde{s}_{l+1}, \ldots, \tilde{s}_k \in [x_2 - \sqrt{T}, x_1 + \sqrt{T}]$ and $\tilde{s}_{k+1}, \ldots, \tilde{s}_m \geq x_1 + \sqrt{T}$, then

$$\begin{aligned}
\alpha W(s_1) - \sum \beta_j W(\tilde{s}_j) &\geq (\alpha - \sum \beta_j) W(s_1) + \sum \beta_j (W(s_1) - W(\tilde{s}_j)) \\
&\geq (\alpha - \sum \beta_j) W(s_1) + \sum_{j=1}^{l} \beta_j \{(W(s_1) - W(x_2 - \sqrt{T}))\} + \\
&\quad \sum_{j=l+1}^{k} \beta_j M(s_1 - \tilde{s}_j)^2 + \sum_{j=k+1}^{m} \beta_j \{(W(s_1) - W(x_1 + \sqrt{T}))\}
\end{aligned}$$



(a) (b)

(c)

FIGURE 14. Comparison of the successive sampling to random sampling approaches. (a) The sources found by the proposed successive algorithm match the exact initial condition u_0; (b) The sources found by a least square algorithm using measurements from randomly chosen locations; (c) The source found by an ℓ_1 minimization algorithm using measurements from randomly chosen locations. Same number of measurements is used for all three examples.

Denote $C = \min(W(s_1) - W(x_2 - \sqrt{T}), W(s_1) - W(x_1 + \sqrt{T}))$.

$$\begin{aligned}\alpha W(s_1) - \sum \beta_j W(\tilde{s}_j) \geq & (\alpha - \sum \beta_j) W(s_1) + \sum_{j=1}^{l} \beta_j (W(s_1) - W(x_2 - \sqrt{T})) + \\ & \sum_{j=l+1}^{k} \beta_j M(s_1 - \tilde{s}_j)^2 + \sum_{j=k+1}^{m} \beta_j (W(s_1) - W(x_1 + \sqrt{T})) \\ \geq & (\alpha - \sum \beta_j) W(s_1) + C\Big\{\sum_{j=1}^{l} \beta_j + \sum_{j=k+1}^{m} \beta_j\Big\} \\ & + M \sum_{j=l+1}^{k} \beta_j (s_1 - \tilde{s}_j)^2.\end{aligned}$$

Therefore

$$(\alpha - \sum \beta_j)W(s_1) + C\{\sum_{j=1}^{l} \beta_j + \sum_{j=k+1}^{m} \beta_j\} + M \sum_{j=l+1}^{k} \beta_j (s_1 - \tilde{s}_j)^2 \leq \alpha \tilde{C}\epsilon. \tag{47}$$

We know that $\alpha - \sum \beta_j \geq 0$, so

$$(\alpha - \sum \beta_j)W(s_1) + C_3\{\sum_{j=1}^{l} \beta_j + \sum_{j=k+1}^{m} \beta_j\} + M \sum_{j=l+1}^{k} \beta_j (s_1 - \tilde{s}_j)^2 \leq \alpha \tilde{C}\epsilon, \tag{48}$$

$$\sum \beta_j \geq \alpha(1 - \frac{\tilde{C}}{W(s_1)}\epsilon). \tag{49}$$

Moreover,

$$\frac{C_3\{\sum_{j=1}^{l} \beta_j + \sum_{j=k+1}^{m} \beta_j\} + M \sum_{j=l+1}^{k} \beta_j (s_1 - \tilde{s}_j)^2}{\sum \beta_j} \leq \frac{1}{\frac{1}{\tilde{C}} - \frac{1}{W(s_1)}\epsilon}\epsilon. \tag{50}$$

Therefore,

$$\frac{\sum_{j=l+1}^{k} \beta_j}{\sum \beta_j} \geq 1 - \frac{1}{\frac{C_3}{\tilde{C}} - \frac{C_3}{W(s_1)}\epsilon}\epsilon, \tag{51}$$

and hence we have

$$\begin{aligned} \frac{\sum_{j=l+1}^{k} \beta_j}{\alpha} &\geq (1 - \frac{1}{\frac{C_3}{\tilde{C}} - \frac{C_3}{W(s_1)}\epsilon}\epsilon)(1 - \frac{\tilde{C}}{W(s_1)}\epsilon) \\ &\geq 1 - [\frac{1}{\frac{C_3}{\tilde{C}} - \frac{C_3}{W(s_1)}\epsilon} + \frac{\tilde{C}}{W(s_1)}]\epsilon \end{aligned}$$

and

$$\frac{\sum_{j=l+1}^{k} \beta_j (s_1 - \tilde{s}_j)^2}{\alpha} \leq \frac{\tilde{C}}{M}\epsilon. \tag{52}$$

□

REFERENCES

[1] L. Bregman, The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex optimization, *USSR Computational Mathematics and Mathematical Physics*, **7** (1967), 620–631.

[2] M. Burger, Y. Landa, N. Tanushev and R. Tsai, Discovering point sources in unknown environments, in *WAFR 2008: The Eighth International Workshop on the Algorithmic Foundations of Robotics*, **57** 2010, 663–678.

[3] J. Cai, S. Osher and Z. Shen, Convergence of the linearized Bregman iteration for ℓ_1-norm minimization, *Math. Comp.*, **78** (2009), 2127–2136.

[4] E. J. Candès and T. Tao, Decoding by linear programming, *IEEE Transactions on Information Theory*, **51**(12) (2005).

[5] Y. Cheng and T. Singh, Source term estimation using convex optimization, *The Eleventh International Conference on Information Fusion*, Cologne, Germany, (2008).

[6] D. Donoho, Compressed sensing, *IEEE Transactions on Information Theory*, **52** (2006), 1289–1306.

[7] A. El Badia, T. Ha Duong and A. Hamdi, Identification of a point source in a linear advection-dispersion-reaction equation: Application to a pollution source problem, *Inverse Problems*, **21** (2005), 1121–1136.

[8] B. Farmer, C. Hall and S. Esedoglu, Source identification from line integral measurements and simple atmospheric models, *Inverse Probl. Imaging*, **7** (2013), 471-C490.

[9] E. Haber, Numerical methods for optimal experimental design of large-scale ill-posed problems, *Inverse Problems*, **24** (2008).
[10] Y. Landa, N. Tanushev and R. Tsai, Discovery of point sources in the Helmholtz equation posed in unknown domains with obstacles, *Comm. in Math. Sci.*, **9** (2011), 903–928.
[11] Y. Li and S. Osher, Coordinate descent optimization for L1 minimization with application to compressed sensing; A greedy algorithm, *Inverse Problems and Imaging*, **3** (2009), 487C-503.
[12] G. Li, Y. Tan, J. Cheng and X. Wang, Determining magnitude of groundwater pollution sources by data compatibility analysis, *Inverse Problem in Science and Engineering*, **14** (2006), 287–300.
[13] L. Ling and T. Takeuchi, Point sources identification problems for heat equations, *Communications in Computational Physics*, **5** (2009), 897–913.
[14] L. Ling, M. Yamamoto, Y. Hon and T. Takeuchi, Identification of source locations in two-dimensional heat equations, *Inverse Problems*, **22** (2006), 1289–1305.
[15] A.V. Mamonov and Y.-H. R. Tsai, Point source identification in non-linear advection-diffusion-reaction systems, *Inverse Problems*, **29** (2013).
[16] S. Osher, M. Burger, D. Goldfarb, J. Xu and W. Yin, An iterative regularization method for total variation-based image restoration, *MMS*, **4** (2005), 460–489.
[17] Z. Wen, W. Yin, D. Goldfarb and Y. Zhang, A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization and continuation, *SIAM J. Scientific Computing*, **32** (2010), 1832–1857.
[18] W. Yin, S. Osher, D. Goldfarb and J. Darbon, Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing, *SIAM J. Imaging Sciences*, (2008), 143–168.

Received January 2011; revised November 2012.

E-mail address: yingyingli1985@gmail.com
E-mail address: sjo@math.ucla.edu
E-mail address: ytsai@ices.utexas.edu