

# Source Term Estimation Using Convex Optimization

Yang Cheng

Department of Mechanical  
and Aerospace Engineering  
University at Buffalo  
Buffalo, New York, U.S.A.  
Email: cheng3@eng.buffalo.edu

Tarunraj Singh

Department of Mechanical  
and Aerospace Engineering  
University at Buffalo  
Buffalo, New York, U.S.A.  
Email: tsingh@eng.buffalo.edu

**Abstract**—A computationally efficient, grid-based estimation method is presented for multiple source identification from distributed sensors. Under the assumption that the sources are located on a grid over the region of interest, the solution to the multiple source identification problem, that is, the number, locations, and intensities of the sources, is represented by a large sparse vector (whose size is greater than that of the observation vector) and is obtained by solving a convex optimization problem using the  $\ell_1$  minimization method. The method can exactly and efficiently recover the true source parameters in the absence of source representation error and measurement noise and can efficiently identify the areas of the true sources with the clusters of grid points in the more realistic scenarios when the source locations do not coincide with the grid points and the sensor data are contaminated by noise.

**Keywords:** source identification, convex optimization,  $\ell_1$  minimization.

## I. INTRODUCTION

Detection and tracking of sources of material release or emission is an important problem in a range of applications. In order to determine the locations and intensities of the sources, typically, mathematical models about the source-sensor relation (and about the source/sensor motion) are presumed, data from an array of distributed sensors are collected and processed, and a set of (constant or time-varying) source parameters are inferred by fusing the models and the data. Both centralized and decentralized configurations can be used for data processing and fusion.

The source characteristics parameters include the locations and intensities (as well as the release times in certain settings) of the sources. When multiple sources are to be identified, the number of sources needs to be determined unless prior information about the number is available. Determining the appropriate number of sources usually involves exhaustively comparing the models of all possible numbers of sources and then choosing the best match under a certain model selection criterion, for example, the Bayes Information Criterion. Many model selection criteria exist [1], and the best choice of the model depends on the criterion used. All model selection criteria are a compromise between model complexity and model accuracy. Suppose there are  $K$  possible numbers of sources (sources that are too close to each other are treated as a single source). A source identification method based on exhaustive comparison solves the  $K$  individual source parameter

estimation problems under the  $K$  hypotheses, and then chooses the best hypothesis and the best estimate under it. Here, model complexity is given by the number of independent parameters in the model, proportional to the number of sources. For large  $K$ , the source identification problem is hard.

From the Bayes perspective, all the information about the source parameters is contained in the posterior probability distribution of the parameters conditional on the sensor data. However, obtaining the posterior distribution is difficult and sometimes impossible. Most source parameter estimation methods solve for a point estimate instead. Choosing a point estimate as the solution of the source identification problem is satisfactory when there exists a single point in the parameter space that is representative of the posterior distribution or the likelihood function. The condition is valid when the parameters are well observable. The optimal point estimate may be the maximum likelihood estimate, the maximum *a posteriori* estimate, or the posterior mean conditional on the sensor data. These point estimates do not differ much in information-rich applications.

Many methods have been used to solve the parameter estimation problem in source identification or source characterization, including the least-squares estimator [2], the genetic algorithm [3], [4], and the particle filter [5], among others. A drawback of the least-squares estimation methods or other gradient-based methods is that they often only converge to the local minimum close to the initial guess. The random sample-based methods are more likely to find the globally optimal solution. Some source identification problems involve large-scale dynamical models. Source identification methods for complex atmospheric dispersion models are reviewed in [6], including forward and backward modeling methods. Forward modeling methods include stochastic Monte Carlo or Markov Chain Monte Carlo sampling techniques; backward or inverse modeling methods include adjoint and tangent linear models, Kalman filters, and variational data assimilation [6]. Most existing methods assume the number of the sources is known or small.

This paper is concerned with multiple source identification where the number of sources is unknown and possibly large. A general grid-based formulation for multiple source identification and an estimation method using convex optimization are presented; a specific two-dimensional, multiple point source

identification problem is then studied. The background of the problem is stationary and mobile radioactive source identification using stationary or mobile sensor networks, which has been studied in [2], [5], [7]–[9], among others. Our proposed method is based on the observation that in many systems, the contributions of multiple sources to a (noiseless) sensor observation satisfies the linear superposition principle. This *linear* structure is explicitly exploited in our problem formulation. By discretizing a parameter subspace (in that specific problem of point source estimation, the subspace of the source location), we formulate the multiple source identification problem as a convex programming problem, which can be solved efficiently using open-source or proprietary software even when the size of the problem is large [10]. The proposed method handles model selection and parameter estimation at the same time and is efficient in identifying locations of significant releases. The proposed method is not optimal in the Bayes sense, however, because of the discretization of the parameter space as well as the cost function we choose to minimize, but many strategies exist to further improve the result.

The organization of the remainder of the article proceeds as follows. First, a generic observation model and a specific radiation sensor model are reviewed. Then, the source identification problem is formulated as a convex optimization problem. Finally, the numerical results of the convex program in four scenarios are presented, followed by the comparisons of the convex program with a gradient-based method and a simulated annealing method.

## II. PROBLEM FORMULATION

### A. Generic Sensor Model

It is assumed that  $M$  sensors are distributed in a region and  $K$  stationary point sources are to be identified in the region. In the absence of sensor noise and other errors, a generic observation model at the  $j^{\text{th}}$  sensor is given by

$$b_j = \sum_{k=1}^K \alpha_k \cdot a_{jk}(x_k, y_k; x_j, y_j) + \alpha_0 \quad (1)$$

where  $b_j$  is the observation at the  $j^{\text{th}}$  sensor, and  $x_j$  and  $y_j$  are the known coordinates of the  $j^{\text{th}}$  sensor;  $\alpha_0$  is the mean of the background noise, which, for sake of simplicity, is assumed to be stationary in time and space;  $K$  is the unknown number of point sources,  $\alpha_k$  is the unknown intensity of the  $k^{\text{th}}$  point source, and  $x_k$  and  $y_k$  are the unknown coordinates of the  $k^{\text{th}}$  point source; and  $a_{jk}(x_k, y_k; x_j, y_j)$  is a known function describing the contribution of the  $k^{\text{th}}$  source to the observation of the  $j^{\text{th}}$  sensor. For non-point sources, additional source parameters, for example, the size and shape of the source, may be needed in  $a_{jk}$  as well. The observation model is linear in  $\alpha_i$  and  $\alpha_0$ , and this fact will be used in our problem formulation and estimation method, which can be applied to any observation model of the form of equation (1). The actual sensor data, denoted by  $\tilde{b}_j$ , are contaminated by internal sensor noise and background noise. The noise properties depend on the sensor characteristics and

the environment. Given the observation model and the sensor data, the ultimate objective of the point source term estimation problem is to estimate the source parameters  $(\alpha_k, x_k, y_k)$  and the number  $K$  of sources (as well as the mean background noise  $\alpha_0$ ) from  $b_j, j = 1, \dots, M$ . The mean background noise is observable for the above observation model. If either the number of sources or the set of all possible locations of the sources is known, the complexity of the estimation problem is greatly reduced.

### B. Radiation Sensor Model

Now a specific radiation sensor model in [5] is briefly reviewed with a focus on the measurement noise properties. The model in [5] is chosen in our study mainly for its simplicity. In this simplified model,  $a_{jk}(x_k, y_k; x_j, y_j)$  is defined by

$$a_{jk}(x_k, y_k; x_j, y_j) = \frac{1}{(x_k - x_j)^2 + (y_k - y_j)^2} \quad (2)$$

It captures the fall-off property that the instantaneous measurement is inversely proportional to the square of the distance from the point source to the sensor. More realistic models for radiation sensors are given in [2], [8], [9]. In practical applications, additional factors are needed in the right-hand side of the equation to account for attenuation by intervening material and source-dependent efficiency of the sensor.

In the presence of sensor noise, the measured quantity  $\tilde{b}_j$  is an integer obeying the Poisson distribution [11], whose mean is determined by both the true source contributions and the background. No possible nonlinear effect is considered in this study. The likelihood function of the  $j^{\text{th}}$  sensor is

$$L_j \triangleq P(\tilde{b}_j | b_j) = \frac{b_j^{\tilde{b}_j} e^{-b_j}}{\tilde{b}_j!} \quad (3)$$

where  $\tilde{b}_j!$  is the factorial of  $\tilde{b}_j$ ,  $e$  is the natural logarithm base, and  $b_j$  is defined by equations (1) and (2). The mean and variance of  $\tilde{b}_j$  are both  $b_j$ .

### C. Formulation of the Estimation Problem

Our proposed method begins with the discretization of the  $x_k - y_k$  space. The motivation is that if the source locations  $x_k$  and  $y_k$  can only take on a finite set of known values, because of the linear structure of the problem, it is easier to obtain the solution to the source term estimation problem. Now let us suppose  $(x_k, y_k)$  take on a total of  $N^S$  ( $K \ll N^S$  and  $M < N^S$  in general) possible values, denoted by  $(x_i^G, y_i^G), i = 1, \dots, N^S$ . One choice of  $(x_i^G, y_i^G)$  is a fine grid over the region of interest. Prior knowledge about the source distribution, if available, greatly reduces the number of grid points. With the introduction of the grid, the sources can now be fully represented by an  $N^S \times 1$  vector  $\mathbf{x}^S$ , of which the element  $x_i^S$  means that the intensity of a radiative “source” at  $(x_i^G, y_i^G)$  is  $x_i^S$ . By definition,  $x_i^S \geq 0$  and a lot of  $x_i^S$  are equal to zero. In other words,  $\mathbf{x}^S$  is a sparse vector. We say that an actual source is located at  $(x_i^G, y_i^G)$  only when  $x_i^S$  exceeds a certain threshold. The number of the elements

in  $\mathbf{x}^S$  exceeding the threshold is the number of point sources over the region.

Equation (1) of the noiseless model is rewritten in terms of  $x_i^G$  and  $y_i^G$  as

$$b_j = \sum_{i=1}^{N^S} x_i^S \cdot a_{ji}(x_i^G, y_i^G; x_j, y_j) + \alpha_0 \quad (4)$$

Note that the summation is from 1 to  $N^S$ . In matrix form, it is

$$\mathbf{b} = A^S \mathbf{x}^S + \mathbf{1}\alpha_0 = \begin{bmatrix} A^S & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{x}^S \\ \alpha_0 \end{bmatrix} = \mathbf{A}\mathbf{X} \quad (5)$$

with obvious definition of  $\mathbf{b}$ ,  $A^S$ ,  $\mathbf{x}^S$ ,  $A$ , and  $\mathbf{X}$ . The dimension of the unknown parameter vector  $\mathbf{X}$  is  $N = N^S + 1$ . The observation (or sensing) matrix  $A$  is determined by the sensors and the grid. For stationary sources and sensors, the matrix only needs to be computed once. Note that the size of  $A$  is  $M \times N$  with  $M < N$  in general and that the linear system is under-determined.

When the true sources are not located at any grid points, resulting from the use of a coarse grid or a small set of random samples in the location space, the corresponding  $A$  matrix will be denoted by  $\tilde{A}$  and the error in  $A$  will be called the source representation error. An upper bound of the representation error is determined by the resolution of the grid. The observation model in this case is described by

$$\mathbf{b} = \tilde{A}\mathbf{X} + \boldsymbol{\epsilon} \quad (6)$$

where  $\boldsymbol{\epsilon}$  is due to the representation error and is unknown.

In the presence of measurement noise, the counterparts of equations (5) and (6) are

$$\tilde{\mathbf{b}} = \mathbf{A}\mathbf{X} + \boldsymbol{\nu} \quad (7)$$

and

$$\tilde{\mathbf{b}} = \tilde{A}\mathbf{X} + \boldsymbol{\epsilon} + \boldsymbol{\nu} \quad (8)$$

respectively, where  $\boldsymbol{\nu}$  is the unknown measurement noise vector. There is no approximation in the linear form of the observation models in equations (6), (7), and (8). The approximation lies in the assumed properties of  $\boldsymbol{\nu}$  and  $\boldsymbol{\epsilon}$ . For convenience, the mean and covariance of  $\boldsymbol{\nu}$  may be assumed to be  $\mathbf{0}$  and  $\tilde{R} \approx \text{diag}(\tilde{\mathbf{b}})$ , where  $\text{diag}(\tilde{\mathbf{b}})$  denotes the diagonal matrix whose major diagonal is given by  $\tilde{\mathbf{b}}$ .

The multiple source term estimation problem is now stated as follows: Given  $A$  and  $\mathbf{b}$ ,  $\tilde{A}$  and  $\mathbf{b}$ ,  $A$  and  $\tilde{\mathbf{b}}$ , or  $\tilde{A}$  and  $\tilde{\mathbf{b}}$ , determine the estimated parameter vector  $\hat{\mathbf{X}}$ , which provides a linear combination of the columns of  $A$  or  $\tilde{A}$  that is the most consistent with  $\mathbf{b}$  or  $\tilde{\mathbf{b}}$ . There is no advantage of this formulation if there is only one source. We have four scenarios:

- 1)  $\mathbf{b} = A\mathbf{X}$
- 2)  $\mathbf{b} = \tilde{A}\mathbf{X} + \boldsymbol{\epsilon}$
- 3)  $\tilde{\mathbf{b}} = A\mathbf{X} + \boldsymbol{\nu}$
- 4)  $\tilde{\mathbf{b}} = \tilde{A}\mathbf{X} + \boldsymbol{\epsilon} + \boldsymbol{\nu}$

The first is the ideal case and the last is the realistic case. The second and the third are also unrealistic because in practice

there is always measurement noise due to imperfectness of the sensor response and the background and there is no guarantee that the true sources are on the grid. Because  $M < N$ , we are faced with an under-determined estimation problem. Because  $K \ll N$ ,  $\mathbf{X}$  must be a sparse vector, that is, only a fraction of it is nonzero. In the next section, we will show how to formulate the optimization problems and solve for the large, sparse  $\hat{\mathbf{X}}$  in the four scenarios using the  $\ell_1$  minimization method [12].

### III. CONVEX OPTIMIZATION FOR SOURCE TERM ESTIMATION

#### A. Solution in the Ideal Scenario

The optimal estimate  $\hat{\mathbf{X}}$  is defined as the solution to the  $\ell_1$  minimization problem for the noiseless case in [12]:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|D\mathbf{X}\|_1 \quad \text{subject to } A\hat{\mathbf{X}} = \mathbf{b}, \mathbf{0} \leq \hat{\mathbf{X}} \leq \mathbf{X}^{UB} \quad (9)$$

where  $D$  is a diagonal matrix with the diagonal elements given by the 2-norms of the columns of  $A$ ,  $A\mathbf{X} = \hat{A}(D\hat{\mathbf{X}}) = (A D^{-1})(D\hat{\mathbf{X}})$ , with  $\hat{A} = A D^{-1}$  a matrix of unit (2-norm) column vectors. The 1-norm of a vector is defined by

$$\|\mathbf{X}\|_1 = \sum_{i=1}^N |X_i| \quad (10)$$

The 2-norm of a vector is defined by

$$\|\mathbf{X}\|_2 = \sqrt{\sum_{i=1}^N |X_i|^2} \quad (11)$$

This is called an  $\ell_1$  minimization problem because the cost is defined by the 1-norm of the vector. In the original  $\ell_1$  minimization problem in [12], the matrix  $A$  in the constraint  $A\mathbf{X} = \mathbf{b}$  is assumed to be a matrix of unit column vectors. That is why we introduced the normalized version of  $A$  as well as  $D$ . The constraint  $\mathbf{0} \leq \hat{\mathbf{X}} \leq \mathbf{X}^{UB}$  requires that all the elements of  $\hat{\mathbf{X}}$  be nonnegative and upper bounded by  $\mathbf{X}^{UB}$ . This constraint is not in the  $\ell_1$  minimization problem in [12] but is important in our source term estimation problem, as will be shown in the next section.

The solution of minimum 1-norm is preferred to the well-known least-square solution, defined by

$$\hat{\mathbf{X}}^{LS} = \arg \min_{\mathbf{X}} \|\mathbf{X}\|_2 \quad \text{subject to } A\hat{\mathbf{X}}^{LS} = \mathbf{b} \quad (12)$$

not only because the latter cannot guarantee nonnegative  $\hat{\mathbf{X}}^{LS}$  but because we know the truth is represented by a sparse vector and the solution of minimum 1-norm is sparser than the solution of the minimum 2-norm. Were the truth  $\mathbf{X}$  not a sparse vector, there would be no reason to prefer a sparse solution  $\hat{\mathbf{X}}$  of our problem. To see that the 1-norm is a better measure of sparsity than the 2-norm, consider two  $N$ -dimensional unit vectors  $[1, 0, \dots, 0]^T$  and  $[1/\sqrt{N}, \dots, 1/\sqrt{N}]^T$ . The former is sparse while the latter is not. The 2-norms of the two vectors are identical and the 1-norms of them are 1 and  $\sqrt{N}$ , respectively.

It should also be noted that although the solution of minimum 1-norm is sparser than that of minimum 2-norm, the solution of minimum 1-norm is not necessarily the sparsest. For example, if there are only two vectors  $[1, 1, 1, 0, \dots, 0]^T$  and  $[10, 0, \dots, 0]^T$  satisfying  $A\mathbf{X} = \mathbf{b}$ , the former, because its 1-norm is smaller, will be chosen as the solution of the  $\ell_1$  minimization problem, but the latter is actually sparser. However, finding out the sparsest solution is NP-hard [13]. An iterative reweighting scheme in [13] helps improve the sparsity of the solution, but there is no absolute guarantee that the sparsest solution can be found by using the iterative reweighting scheme. The basic idea of iterative reweighting is to minimize a cost in favor of a sparser solution than that of minimum 1-norm. Formally, the new cost is given by

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|[\text{diag}(\mathbf{X})]^{-1} \mathbf{X}\|_1 = \arg \min_{\mathbf{X}} \sum_{i=1}^N \frac{X_i}{|X_i|}$$

with  $0/0 \triangleq 0$  in the summation. The new cost corresponds to the 0-norm of  $\mathbf{X}$  or the total number of nonzero elements of  $\mathbf{X}$ . To avoid the ‘‘divide by zero’’ problem in practice,  $X_i/|X_i|$  is replaced by  $X_i/(|X_i| + \epsilon)$ , where  $\epsilon$  is a small positive number. The minimization problem is solved iteratively using the  $\ell_1$  minimization method, with the cost given by

$$\hat{\mathbf{X}}^{(l+1)} = \arg \min_{\mathbf{X}} \sum_{i=1}^N \frac{X_i}{|\hat{X}_i^{(l)}| + \epsilon}$$

where  $\hat{\mathbf{X}}^{(l+1)}$  is the optimal estimate in the  $(l+1)$ <sup>th</sup> iteration and  $\hat{X}_i^{(l)}$  is the  $i$ <sup>th</sup> element of the optimal estimate from the  $l$ <sup>th</sup> iteration.

### B. Solutions in Non-Ideal Scenarios

Only the formulations for the last scenario are given below. The formulations for the other two scenarios are different only in the data matrices. The normalized version of  $\hat{A}$  is used in the problem definition, given by  $\hat{A} = \tilde{A}\tilde{D}^{-1}$ , where  $\tilde{D}$  is the diagonal matrix of the 2-norms of the columns of  $\hat{A}$ .

$$\begin{aligned} \hat{\mathbf{X}} &= \arg \min_{\mathbf{X}} \|\tilde{D}\mathbf{X}\|_1 \\ \text{subject to } \|W(\hat{A}\hat{\mathbf{X}} - \tilde{\mathbf{b}})\|_2 &\leq \lambda_2, \quad \mathbf{0} \leq \hat{\mathbf{X}} \leq \mathbf{X}^{UB} \end{aligned} \quad (13)$$

where  $\lambda_2$  is a control parameter. A lower bound of  $\lambda_2$  is obtained by solving

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|W(\tilde{A}\mathbf{X} - \tilde{\mathbf{b}})\|_2 \quad \text{subject to } \mathbf{0} \leq \hat{\mathbf{X}} \leq \mathbf{X}^{UB} \quad (14)$$

The tuning parameter  $\lambda_2$  is determined by simulations. When the control parameters are too small, no feasible solution exists. When they are too large, the optimal solution may be nothing but the null vector. We have used 1.1 times the lower bound in the second formulation. For the weighting matrix  $W$ , we have used  $W = I_{M \times M}$  because of its simplicity. The optimal weighting matrix  $W$  requires knowledge about the noise as well as the error.

## IV. NUMERICAL RESULTS

### A. Simulation Setup

The region of interest is a rectangle of  $30 \times 30$ . A uniform  $20 \times 20$  grid is generated over the region. So,  $N^S = 400$  and  $N = N^S + 1 = 401$ . The numbers of the sensors that cover the region are  $M = 100$ ,  $M = 144$ , and  $M = 196$ . The sensors provide reasonably good coverage over the region. Good coverage may also be achieved by much fewer mobile sensors with known position and movement. Three patterns of the sensor locations are used in the numerical tests: random, uniform grid, and concentric circles. The interesting question about the different patterns is whether one is significantly better than the other two. No sensor is located at a grid point in order to avoid the singularity problem of the sensor model given by equation (2). No optimization of the sensor configuration is done. The matrix  $A$  is determined by the sensor locations and the grid. Its size is  $M \times N$ , i.e.,  $100 \times 401$ ,  $144 \times 401$ , or  $196 \times 401$ .

Two sensor noise levels are used. Suppose the true concentration is  $b_j$ . The two sensor noise variances used are  $0.1b_j$  and  $b_j$ . For the smaller variance, the noise distribution is chosen to be zero-mean Gaussian. For the larger variance, the distribution of the sensor measurement is the Poisson distribution with mean and variance  $b_j$ . The maximum number of the point sources in the region is 8, i.e.,  $1 \leq K \leq 8$ . The maximum number of sources is not limited by the  $\ell_1$  minimization method, but the observability of the source parameters. In scenario 1 and 3, in which there is no representation error, the true source locations are randomly selected from the grid points. In the other two scenarios, the true source locations are randomly chosen in the region and are not necessarily located at any grid points. The maximum source representation error is half the grid resolution, 0.79 in each direction. The intensity of a source is randomly chosen between 50 and 100. The upper bound, 100, is available to the source term estimation method. The background noise is set to 1.

We use CVX for MATLAB, a package for specifying and solving convex programs [14], [15], to solve the convex optimization problems. The exit conditions of the CVX solver include ‘‘Solved,’’ ‘‘Inaccurate/Solved,’’ ‘‘Unbounded,’’ ‘‘Inaccurate/Unbounded,’’ ‘‘Infeasible,’’ ‘‘Inaccurate/Infeasible,’’ ‘‘Failed,’’ and ‘‘Overdetermined.’’ We accept the solution when the condition is either ‘‘Solved’’ or ‘‘Inaccurate/Solved.’’

### B. Result of Scenario 1 ( $\mathbf{b} = A\mathbf{X}$ )

In the ideal scenario, the simulation results show that no pattern is better than the others. For all the three sensor patterns and all the three numbers of sensors, the optimal estimate  $\hat{\mathbf{X}}$  as the solution to the problem defined by equation (9) is identical to the truth  $\mathbf{X}$  within the solver precision. We believe that the constraint  $\mathbf{0} \leq \hat{\mathbf{X}} \leq \mathbf{X}^{UB}$  plays a role in the exact recovery of the sources. The same result cannot be obtained without the constraint  $\mathbf{0} \leq \hat{\mathbf{X}} \leq \mathbf{X}^{UB}$ . In fact, for any of the three sensor patterns, the solution to

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|D\mathbf{X}\|_1 \quad \text{subject to } A\hat{\mathbf{X}} = \mathbf{b} \quad (15)$$

or

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{X}\|_1 \quad \text{subject to } A\hat{\mathbf{X}} = \mathbf{b} \quad (16)$$

can sometimes be totally different from the truth in both the number of the sources and the values of the source parameters. It is also observed that the equality constraint  $A\hat{\mathbf{X}} = \mathbf{b}$  can be replaced by

$$\|A\hat{\mathbf{X}} - \mathbf{b}\|_2 \leq \lambda_\epsilon \quad (17)$$

with  $\lambda_\epsilon$  a small positive number, for example, in the order of  $10^{-12}$  or below. The resulting change in  $\hat{\mathbf{X}}$  due to this replacement is negligible.

### C. Result of Scenario 2 ( $\mathbf{b} = \tilde{A}\mathbf{X} + \epsilon$ )

The source representation error is unique to the grid-based method. Because of the source representation error, even in this noiseless scenario, the grid-based method cannot recover the true sources exactly, while a non-grid-based method still can.

There is no limitation in the distance between a sensor and a source in the simulations. As a result, a sensor may be very close to a source and the corresponding observation  $b_j$  may be extremely large:  $b_j \gg 100 \max_i A_{ji}$ , where 100 is the upper bound of the source intensity in the simulations. That means no source on the grid can compensate for the large residual at that sensor. When this happens, in the preprocessing of our methods,  $b_j$  is set to  $100 \max_i A_{ji}$  or a comparable value.

Now the result is presented of a hard case with 144 sensors on concentric circles and four sources. The sensors and the sources are shown in Figure 1. Note that two sources are close to the sensors. The resolution of the grid around the two sources are inadequate. This is the main problem due to source representation errors.

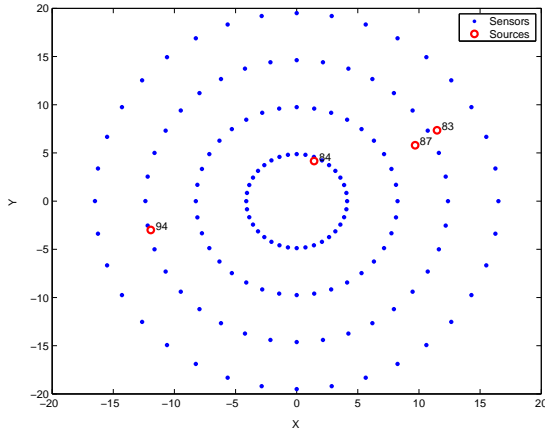


Figure 1. Sensors and Sources

The  $\ell_1$  minimization solution can satisfy  $\tilde{A}\hat{\mathbf{X}} = \mathbf{b}$  exactly only when the constraint  $\mathbf{0} \leq \hat{\mathbf{X}} \leq \mathbf{X}^{UB}$  is not imposed. The  $\ell_1$  minimization solution vector without the constraint is shown in Figure 2. This is a sparse solution that satisfies  $\tilde{A}\hat{\mathbf{X}} = \mathbf{b}$  exactly, but some elements of  $\hat{\mathbf{X}}$  are negative. The exact solution is thus meaningless. With the constraint

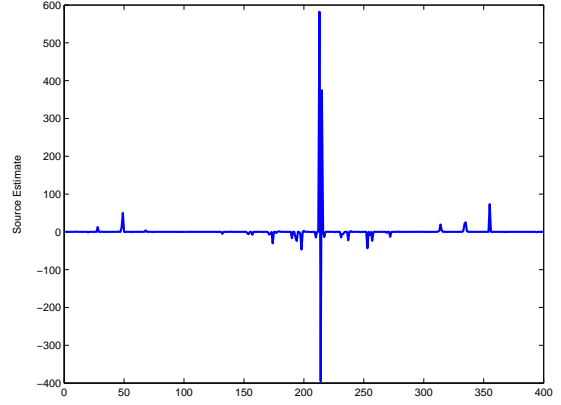


Figure 2. Source Estimation Solution Vector without Nonnegative Source Constraint

$\mathbf{0} \leq \hat{\mathbf{X}} \leq \mathbf{X}^{UB}$ , the minimum magnitude of the residual is approximately 297, which is the lower bound of  $\lambda_2$  in equation (13). The large residual results from the sensors close to the sources. The  $\ell_1$  minimization solution corresponding to  $\lambda_2 = \min\|\tilde{A}\hat{\mathbf{X}} - \mathbf{b}\|_2 \approx 297$  is given in Figures 3 and 4.

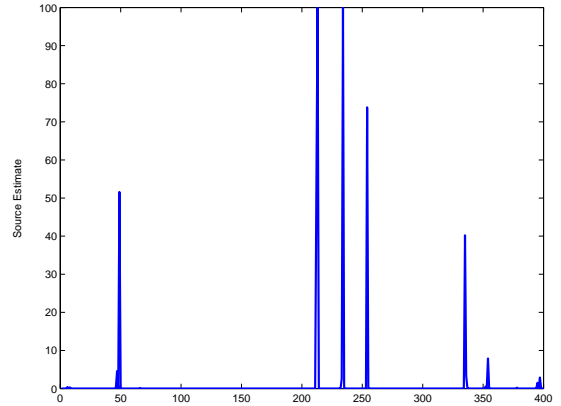


Figure 3. Source Estimation Solution Vector with Nonnegative Source Constraint

It can be seen that with the constraint, all the elements are nonnegative but there are many more than four sources. It is generally true that more sources are needed to account for the perturbation in  $\tilde{A}$  and minimize the residual error. Increasing  $\lambda_2$  by ten percent and applying the iterative reweighting scheme leads to much fewer sources, as shown in Figure 5. From Figure 5, it can be seen that although owing to the significant source representation error, the true sources cannot be exactly recovered, the clusters are informative about the true source locations. A gradient-based search in the local regions covered by the clusters (assuming one or two sources in each local region) can significantly improve the accuracy of the result. Using a finer grid for that region can also improve the result.

The estimated source intensity is inaccurate in this example. For instance, the intensity of the leftmost source is 94 while

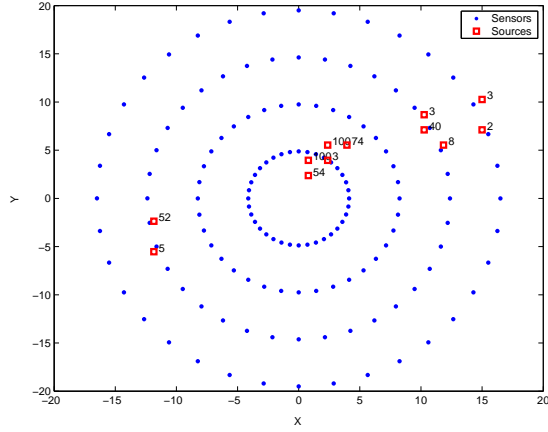


Figure 4. Estimated Sources (source threshold = 1)

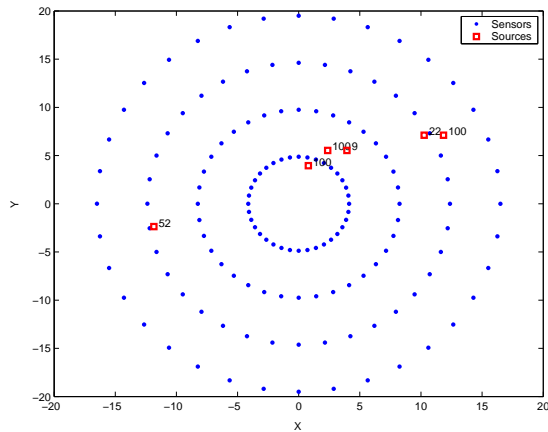


Figure 5. Estimated Sources (source threshold = 1, with iterative reweighting)

the intensity of the estimated source there is only 52. This discrepancy is mainly due to the fact that a sensor is close to both the true and the estimated sources. The distance between the sensor and the leftmost true source is approximately 0.52, while the distance between the sensor and the estimated source is 0.39. The square of the ratio is approximately 0.55, in good agreement with the ratio of 94 to 52. As a general rule, when there is a sensor that is very close to a source but is not as close to the neighboring grid points, large discrepancy in the source intensity between the truth and the estimate is inevitable. When the sensor is moved away from the source, the source intensity estimate becomes less sensitive to the variation in the location of the source.

*D. Result of Scenarios 3 and 4* ( $\tilde{\mathbf{b}} = \mathbf{A}\mathbf{X} + \boldsymbol{\nu}$  and  $\tilde{\mathbf{b}} = \tilde{\mathbf{A}}\mathbf{X} + \boldsymbol{\epsilon} + \boldsymbol{\nu}$ )

In the presence of sensor noise, no estimation or optimization method can exactly recover the truth. One hundred sensors are not sufficient to provide accurate estimates in certain hard cases. Of the three sensor patterns, the pattern of concentric circles is overall worse than the other two because it provides a less uniform coverage of the whole region. The area close

to the boundary of the rectangular region are not covered adequately.

When the sensor noise variance is  $0.1b_j$  and the source locations can be well approximated by the grid points (i.e., the source representation error is not so severe as in the example of the last section), the estimation result remains accurate, comparable to the results when the source representation error is the only error source. When the sensor noise variance increases to  $b_j$ , it is common to have an estimate of more sources than the maximum number of true sources. That is because more sources are needed to interpret the fluctuations and deviations of the measurements. That statement is true for other optimization methods as well. A typical example of eight sources and 196 sensors is given. The iterative reweighting method has been used to reduce the number of sources. The true sources are given in Figure 6. The estimated sources are plotted in Figures 7 and 8. In the former, all sources with intensities greater than one are plotted; in the latter, only the sources with intensities greater than 20 are plotted. Many small false sources are introduced by the large measurement noise. As in the source representation error only case, the source intensities are not accurately estimated. In addition, the two leftmost sources become one. However, the estimation results are informative about the source locations.

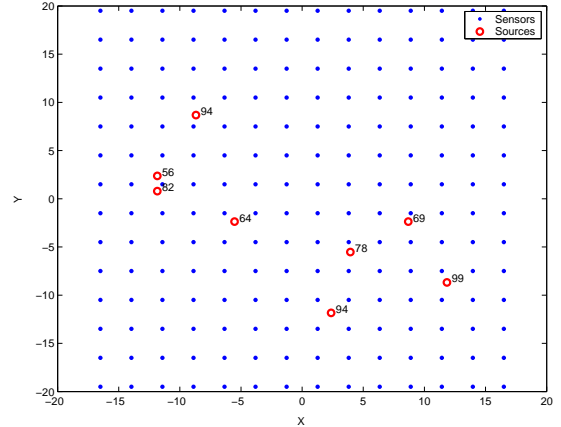


Figure 6. True Sources in Scenario 4

### E. Comparison with Other Methods

The proposed convex optimization-based source term estimation method is compared with a gradient-based, constrained local optimization method in the MATLAB Optimization Toolbox, `fmincon`, and a simple simulated annealing method. The large-scale algorithm of `fmincon` is a subspace trust-region method and is based on an interior-reflective Newton method; the medium-scale algorithm of `fmincon` uses a sequential quadratic programming method. The simulated annealing algorithm directly searches the parameter space by a sequence of random walk. All the methods are iterative.

The design space as well as the complexity of the proposed convex method is determined mainly by the grid size  $N$ , not

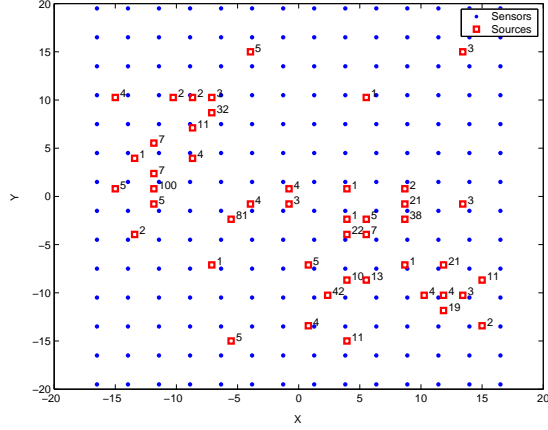


Figure 7. Estimated Sources in Scenario 4 (source threshold = 1, with iterative reweighting)

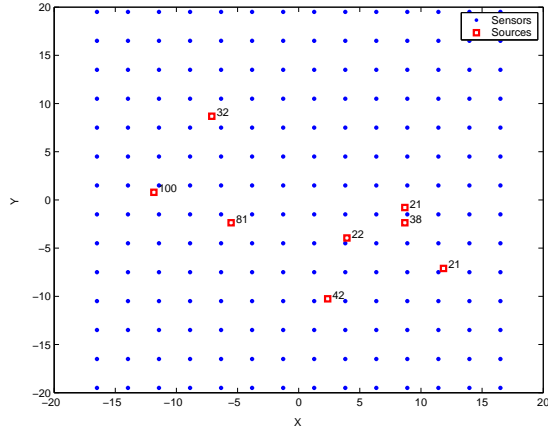


Figure 8. Estimated Sources in Scenario 4 (source threshold = 20, with iterative reweighting)

the number  $K$  of sources. In the previous numerical study, the dimension of the design space of the proposed method is 401. The solution of the proposed method determines the number of sources and the values of the source parameters at the same time.

The design space of the other two methods is determined by the possible number of sources (8 in the previous numerical study) and the number of independent parameters per model, given by  $3K + 1$ , where  $K$  is the assumed number of sources in the model. In the previous numerical study, the maximum number of sources is eight. Since the number of sources is unknown, in the worst case, eight solutions need to be solved for by `fmincon` and the simulated annealing method, with each solution vector being of the size of  $3K + 1$ ,  $K = 1, \dots, 8$ .

The proposed method takes advantage of the linear structure of the source term estimation problems. The other two do not. The accuracy of the proposed method depends on the scenarios as well as the resolution of the grid. It can perfectly recover the truth in the first scenario. Its accuracy in the number, locations, and intensities of the sources in the other

three scenarios mainly depends on how severe the effect of the source representation error is. The gradient-based method does not always converge to the global minimum. It can exactly recover the truth in the first and second scenarios only when the local minimum it arrives at happens to be the global minimum, for which the probability is not high. Because of this local minima problem, the accuracy of the gradient-based method is the worst; the result of the gradient-based method is of very limited use without good initial guess. The simulated annealing method can exactly recover the truth within its precision in the first and second scenarios. It also yields the most accurate estimation results in the other scenarios because its accuracy is not limited by the resolution of the grid and in principle it can converge to the global minimum when there is no limitation in the maximum number of iterations.

Table I  
EXECUTION TIME FOR `FMINCON` (SEC)

	$M = 100$	$M = 144$	$M = 196$
$K = 1$	0.02	0.05	0.29
$K = 2$	0.06	0.40	0.61
$K = 4$	0.25	2.78	3.42
$K = 6$	0.52	5.37	6.70
$K = 8$	1.09	17.40	25.30

Table II  
EXECUTION TIME FOR SIMULATED ANNEALING (SEC)

	$M = 100$	$M = 144$	$M = 196$
$K = 1$	116.80	147.69	241.26
$K = 2$	423.49	588.04	865.11
$K = 4$	1746	2289	3205
$K = 6$	3562	5096	6951
$K = 8$	64421	11302	16167

Table III  
EXECUTION TIMES FOR  $\ell_1$  MINIMIZATION (SEC)

	$M = 100$	$M = 144$	$M = 196$
$N = 401$	32.50	37.30	41.40
$N = 901$	57.23	80.87	114.50
$N = 1601$	82.46	122.37	209.14

The execution times of the three methods are compared in the MATLAB environment under Windows XP Professional on a DELL desktop computer of dual Intel Pentium CPU with the processing rate of 2.4 GHz and of 4 GB RAM. The most interesting question is how well the methods scale with the size of the problem. The same data sets are processed by the three methods. The MATLAB timing function `CPUTIME` is used to get the average execution times of the methods. It should be pointed out that the execution time of a method depends on the termination criterion and parameter tuning of the method, the shape of the cost and the constraints, the difficulty of the test problem for the method, the efficiency

of the MATLAB code, and whether pre-compiled functions are used in the code (the pre-compiled functions are executed much faster than the script functions), and so on. The execution times of the simulated annealing method and `fmincon` also heavily depend on the initial guess.

The timing results are given in Table I, II, and III. The execution time of the  $\ell_1$  minimization method is mainly determined by the grid size. Hence, the execution times of the method (with the computation of the lower bound of  $\lambda_2$  and five iterations of iterative reweighting) are given for different  $M$  and  $N$ . The simulated annealing method and `fmincon` are not based on grid; therefore, the timing results for them are presented for different  $M$  and  $K$ . Note that the execution time for `fmincon` or the simulated annealing algorithm for a specific  $K$  in the tables is the time used to solve the parameter estimation problem under a single hypothesis that the number of sources is  $K$ . The times for the two methods in the tables are less than the times they take to solve a complete problem with unknown number of sources.

It can be seen that the execution time for `fmincon` is the shortest. However, that is only the execution time for one initial guess, which may not lead to any accurate or meaningful estimate. If `fmincon` is used to search the whole region with tens or hundreds of initial guesses, the total execution time of `fmincon` will easily exceed that of the  $\ell_1$  minimization method. The simulated annealing method is much slower than the other two. To solve a parameter estimation problem of 24 parameters (eight sources), the simulated annealing method we used takes more than four hours on average. The execution times of `fmincon` and the simulated annealing method increase dramatically with the number of sources, probably because the difficulty associated with high-dimensional space. The execution time of the  $\ell_1$  minimization method is below four minutes and does not scale fast with the grid size. Compared with the simulated annealing method and `fmincon`, the  $\ell_1$  minimization method has nice balance between accuracy and complexity for multiple source term estimation problems.

## V. CONCLUSIONS

Multiple source term estimation is complex when the number and locations of sources is unknown. A grid-based method is presented that determines the number of sources and the source characteristics simultaneously and efficiently. The high efficiency comes from the fact that the observation model can be made formally linear by a discretization procedure and that the optimization problem so formulated is convex and the solution vector is sparse. Using the computationally efficient  $\ell_1$  minimization method, we can exactly recover the true source characteristics from the distributed sensors in the region in ideal scenarios. In the presence of source representation error due to discretization and the measurement noise, the local areas of the sources can still be identified with the clusters of grid points. The accuracy is limited by the resolution of the grid. The accuracy of the solution of the proposed method can be further improved by re-solving the problem with finer

grids in the identified areas of sources. The accuracy can also be improved by a local search using a gradient-based method. The complexity of the proposed method is mainly determined by the size of the grid, but not the number of sources. The maximum number of sources that can be identified is not limited by the proposed method, but the amount of information contained in the sensors. Finally, we note that this source term estimation method can be extended to the case of multiple moving sources with modest modification when perfect or accurate knowledge about the source movement is available.

## ACKNOWLEDGMENT

This work was supported by the Defense Threat Reduction Agency (DTRA) under Contract No. W911NF-06-C-0162. The authors gratefully acknowledge the support and constructive suggestions of Dr. John Hannan of DTRA. The authors are also grateful to Dr. Zane W. Bell of Oak Ridge National Laboratory for helpful suggestions.

## REFERENCES

- [1] J. B. Kadane and N. A. Lazar, "Methods and criteria for model selection," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 279–290, 2004.
- [2] J. W. Howse, L. O. Ticknor, and K. R. Muske, "Least squares estimation techniques for position tracking of radioactive sources," *Automatica*, vol. 37, pp. 1727–1737, 2001.
- [3] C. Allen, G. Young, and S. Haupt, "Improving pollutant source characterization by optimizing meteorological data with a genetic algorithm," *Atmospheric Environment*, vol. 41, pp. 2283–2289, 2006.
- [4] C. Allen, S. Haupt, and G. Young, "Source characterization with a genetic algorithm-coupled dispersion-backward model incorporating scipuff," *Journal of Applied Meteorology and Climatology*, vol. 46, pp. 273–287, 2007.
- [5] M. Morelande, B. Ristic, and A. Gunatilaka, "Detection and parameter estimation of multiple radioactive sources," in *The 10th International Conference on Information Fusion*, Quebec City, Quebec, 2007.
- [6] K. S. Rao, "Source estimation methods for atmospheric dispersion," *Atmospheric Environment*, vol. 41, pp. 6964–6973, 2007.
- [7] S. M. Brennan, A. M. Mielke, and D. C. Torney, "Radioactive source detection by sensor networks," *IEEE Transactions on Nuclear Science*, vol. 52, no. 3, pp. 813–819, 2005.
- [8] J. Daniel L. Stephens and A. J. Peurrung, "Detection of moving radioactive sources using sensor networks," *IEEE Transactions on Nuclear Science*, vol. 51, no. 5, pp. 2273–2278, 2004.
- [9] R. J. Nemzek, J. S. Dreicer, D. C. Torney, and T. T. Warnock, "Distributed sensor networks for detection of mobile radioactive sources," *IEEE Transactions on Nuclear Science*, vol. 51, no. 4, pp. 1693–1700, 2004.
- [10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004, available at <http://www.stanford.edu/~boyd/cvxbook.html>.
- [11] G. F. Knoll, *Radiation Detection and Measurement*, 2nd ed. New York, NY: John Wiley & Sons, 1988, ch. 3.
- [12] E. J. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, Madrid, Spain, 2006.
- [13] E. J. Candès, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted  $\ell_1$  minimization," preprint.
- [14] M. Grant and S. Boyd, "CVX: MATLAB software for disciplined convex programming," (web page and software). <http://stanford.edu/~boyd/cvx>.
- [15] —, "Graph implementations for nonsmooth convex programs," to appear in *Recent Advances in Learning and Control* (a tribute to M. Vidyasagar), V. Blondel, S. Boyd, and H. Kimura, editors. Springer, 2008. [http://stanford.edu/~boyd/graph\\_dcp.html](http://stanford.edu/~boyd/graph_dcp.html).