A Comparison of Traditional and New Inverse Modelling Techniques for Source Term Identification in the Atmosphere

Lukáš Adam and Martin Branda

Institute of Information Theory and Automation, The Czech Academy of Sciences

branda@utia.cas.cz

The Czech Academy of Sciences

1. Introduction

Inverse modelling plays an important role in identifying the amount of harmful substances released into atmosphere during major incidents such as power plant accidents or volcano eruptions. Another possible application of the inverse modelling lies in the monitoring of CO_2 emission limits where only measurements at certain places are available and the task is to estimate the total releases at given locations.

Assume that vector $(x_1, \dots, x_n) \in \mathbb{R}^n$ stands for **unknown parameters** in spatial–time domain and consider a mapping $M : \mathbb{R}^n \to \mathbb{R}^m$ such that $M(x)$ is the predicted measurement vector at given points. Having a vector $(y_1, \dots, y_m) \in \mathbb{R}^m$ of **known measurements**, we would like to find x such that the following quantity

$$\|M(x) - y\|$$

is minimized, where $\|\cdot\|$ denotes the Euclidean norm. In atmospheric modelling, mapping M is often linear, represented by a matrix $M \in \mathbb{R}^{m\times n}$. This leads to an optimization problem

$$\underset{x}{\text{minimize}}\ \|Mx - y\|. \tag{1}$$

Matrix M is often denoted as **source–receptor sensitivity matrix** . The reason for this is that element m_{ij} of matrix M represents the sensitivity of the measurement at point j to the release at point i.

There are two main approaches to finding a solution to (1). The first one is a deterministic approach and makes use directly of formula (1) and tries to solve it by optimization techniques. Since the problem is often ill–conditioned, various regularizations are used to make the problem more tractable. This approach usually results in the necessity of solving a constrained quadratic program. The second approach is a stochastic one and instead of solving (1), it assumes that

$$y = Mx + \varepsilon, \tag{2}$$

where ε is a random vector. Provided that ε has normal distribution with independent components having zero mean and the same variance, then applying the maximal likelihood estimate reduces precisely to solving (1) which can be seen as ordinary least squares.

Our **first goal** is to propose a modification of the deterministic approach for solving problem (1) by adding nondiagonal weighting matrix W to obtain (8), see below. This approach is closely connected with Bayesian modelling, where the weighting matrix enters as a covariance matrix of the measurements. However, we base the weighting matrix purely on the topology of the measurement points.

The **second goal** is to show a new approach of dealing with ill–conditioned sensitivity matrix M. To the best of our knowledge, the usual approach is either to use some regularization or to ignore certain measurements, which reduces the number of rows in M. This, however, may lead to a suboptimal solution when the solution of the reduced problem is not a solution to the original problem. We try to prevent this behaviour and suggest to look for a sparse solution x, which means that x should contain as many zeros as possible. This problem may be formulated as a multiobjective optimization: we try to minimize the measurement error $\|Mx - y\|$ and at the same time, we try to minimize the number of nonzeros, which is denoted by $\|x\|_0$.

2. Spatial and temporal locations weighting

It could be advantageous not to compare Mx and y componentwise but to take into account their spatial and temporal locations and compare the sum on a neighborhood of every component. We assume that for every measurement y_j we know additional data $z_j = (z_j^x, z_j^y, z_j^t)$, where pair (z_j^x, z_j^y) represents the **longitude and latitude** of a measurement point and z_j^t the **measurement time**. We would like to define the distance between the measurement points in an easily tractable manner.

2. Spatial and temporal locations weighting – cont.

First, we define the space and time distances as follows

$$d_S(z_i, z_j) := \begin{cases} \exp(-\alpha_S\|(z_i^x, z_i^y) - (z_j^x, z_j^y)\|) \\ \quad \text{if } \|(z_i^x, z_i^y) - (z_j^x, z_j^y)\| & \le s_{max}, \\ 0 & \text{otherwise,} \end{cases} \tag{3a}$$

$$d_T(z_i, z_j) := \begin{cases} \exp(-\alpha_T\|z_i^t - z_j^t\|) \text{ if } \|z_i^t - z_j^t\| \le t_{max}, \\ 0 \qquad\qquad \text{otherwise,} \end{cases} \tag{3b}$$

where $\alpha_S \ge 0$, $\alpha_T \ge 0$ and $s_{max} \in [0, \infty]$, $t_{max} \in [0, \infty]$ are given parameters; the last two are known as cutoff distances. Since both quantities in (3) lie in interval $[0, 1]$, we may define the **distance** between z_i and z_j as

$$d(z_i, z_j) := d_S(z_i, z_j) d_T(z_i, z_j). \tag{4}$$

Note that this distance is zero if the measurements are performed at distant places (as specified by s_{max}) or at distant times (as specified by t_{max}). Moreover, the distance decreases with increasing spatial or temporal distance. This rate of decrease is determined by parameters α_S and α_T.

When considering ordinary least squares, instead of minimizing discrepancy $\left|(Mx)_j - y_j\right|$ at given point j, we will minimize the difference between Mx and y on a neighborhood of point j. If we relate this neighborhood to distance d defined in (4), for every measurement $j = 1, \dots, m$ we try to minimize the following quantity

$$\left|\sum_{i=1}^{m} \frac{d(z_j, z_i)}{\sum_{k=1}^{m} d(z_j, z_k)}(Mx)_i - \sum_{i=1}^{m} \frac{d(z_j, z_i)}{\sum_{k=1}^{m} d(z_j, z_k)} y_i\right|, \tag{5}$$

where the denominator is a weighting factor. Thus, adding weighting matrix W moves uncertainties from a point to its neighborhood. When we combine components (5) into one vector, we arrive at minimizing

$$\|W(Mx - y)\|^2, \tag{6}$$

where weighting matrix W consists of elements

$$w_{ij} = \frac{d(z_i, z_j)}{\sum_{k=1}^{m} d(z_i, z_k)}. \tag{7}$$

To conclude this approach, we propose to minimize **weighted least squares** (WLS) under nonnegativity constraints

$$\begin{aligned} &\underset{x}{\text{minimize}}\ \|WMx - Wy\|^2 \\ &\text{subject to } x \ge 0 \end{aligned} \tag{8}$$

instead of problem (1). Similar extensions may be performed for problems with Tikhonov regularization or for any problem based on the ordinary least squares method.

3. Sparse optimization techniques

We will now concentrate on finding sparse solutions to problem (8). Since such solution is uniquely determined provided the system is overdetermined, it is usually assumed that matrix WM has more columns than rows, thus $m < n$. Then there exist multiple solutions and the task of sparse optimization is to select the one with the lowest number of nonzero components. In the opposite case of $m > n$, the solution of (8) is usually uniquely determined but the solution may be dense. In such cases it is possible to trade higher density for a slightly worse error $\|WMx - Wy\|$. An advantage of sparse solutions is that columns corresponding to zero components of a solution are ignored, i.e. we deal with ill–conditioned matrices M in a natural way.

We employ the l_0 "norm", which is defined as

$$\|x\|_0 := \#\{i \mid x_i \neq 0\},$$

where $\#A$ denotes the number of elements in a set A. Thus, sparse optimization tries to minimize $\|x\|_0$, together with criterion (6).

3. Sparse optimization techniques – cont.

In sparse optimization, instead of (8) one usually solves

$$\begin{aligned} &\underset{x}{\text{minimize}}\ \|x\|_0 \\ &\text{subject to } \|WMx - Wy\|^2 \le \varepsilon_{tol}, \\ &\qquad x \ge 0, \end{aligned} \tag{9}$$

where $\varepsilon_{tol} \ge 0$ signifies the maximal possible error between WMx and Wy. Another possibility is to solve

$$\begin{aligned} &\underset{x}{\text{minimize}}\ \|WMx - Wy\|^2 \\ &\text{subject to } \|x\|_0 \le k_{tol}, \\ &\qquad x \ge 0, \end{aligned} \tag{10}$$

where $k_{tol} \in \mathbb{N}$ is a natural number which denotes the maximal number of nonzeros in x. To compare problems (9) and (10), we observe first that parameter ε_{tol} is a real one while k_{tol} is an integer. This implies that it may be simpler to choose the value of k_{tol} which is not sensitive to scaling in variables. Moreover, the optimal value of (10) will provide better error $\|WMx - Wy\|$ under the same sparsity. However, this problem is generally more difficult to solve.

We may approximate nonconvex problem (9) by a **convex** one using l_1 norm instead of $\|x\|_0$ leading to

$$\begin{aligned} &\underset{x}{\text{minimize}}\ \sum_{i=1}^{n} x_i \\ &\text{subject to } \|WMx - Wy\|^2 \le \varepsilon_{tol}, \\ &\qquad x \ge 0. \end{aligned} \tag{11}$$

Using artificial binary variables $z_i \in \{0, 1\}$ such that

$$\begin{aligned} z_i = 0 &\iff x_i = 0, \\ z_i = 1 &\iff x_i > 0, \end{aligned} \tag{12}$$

we obtain

$$\|x\|_0 = \sum_{i=1}^{n} z_i.$$

Then, instead of (10) we consider **mixed-integer problem**

$$\begin{aligned} &\underset{x,z}{\text{minimize}}\ \|WMx - Wy\|^2 \\ &\text{subject to } \sum_{i=1}^{n} z_i \le k_{tol}, \\ &\qquad z_i \cdot \text{lb}_i \le x_i \le z_i \cdot \text{ub}_i,\ i = 1, \dots, n, \\ &\qquad z_i \in \{0, 1\}. \end{aligned} \tag{13}$$

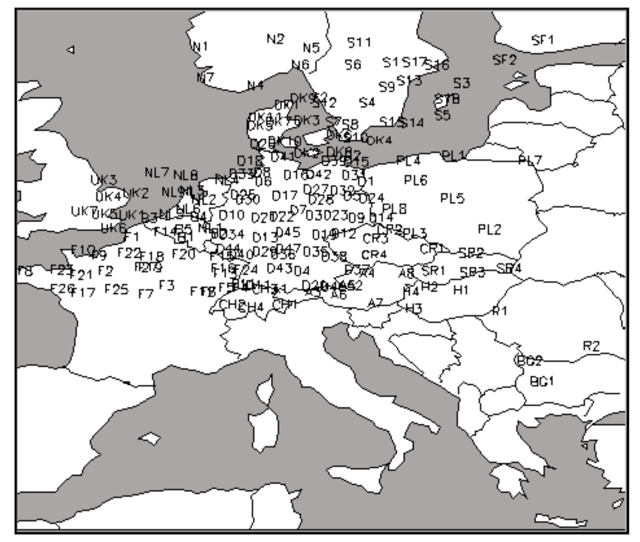
For a solution x of this problem, we always have that at most k_{tol} components x_i are positive and if this is the case, then they are greater than lb_i. We can also control the highest value by setting ub_i.

4. Application to ETEX

Now we are ready to compare the methods on real data. ETEX (European Tracer Experiment) is a controlled tracer experiment performed in 1994 near Rennes in France with detailed information about the release. This experiment was performed twice, for the first time on 23 October 1994 and for the second time on 14 November 1994. A total of 340kg of PMCH was released into the atmosphere during the course of 12 hours. The sampling network consisted of 168 stations. These stations are depicted in Figure 1. Each station was supposed to sample over the period of 72 hours with the time difference between two subsequent measurements being 3 hours. Thus, every station was to provide 24 measurements. The stations closest to the release point started to sample 3 hours before the release was performed while the stations far away from the release ended their sampling activity 90 hours after the release had started.

4. Application to ETEX – cont.

Figure 1: Locations of measurement stations for the ETEX experiment. The first letter denotes a country in which the station is located..

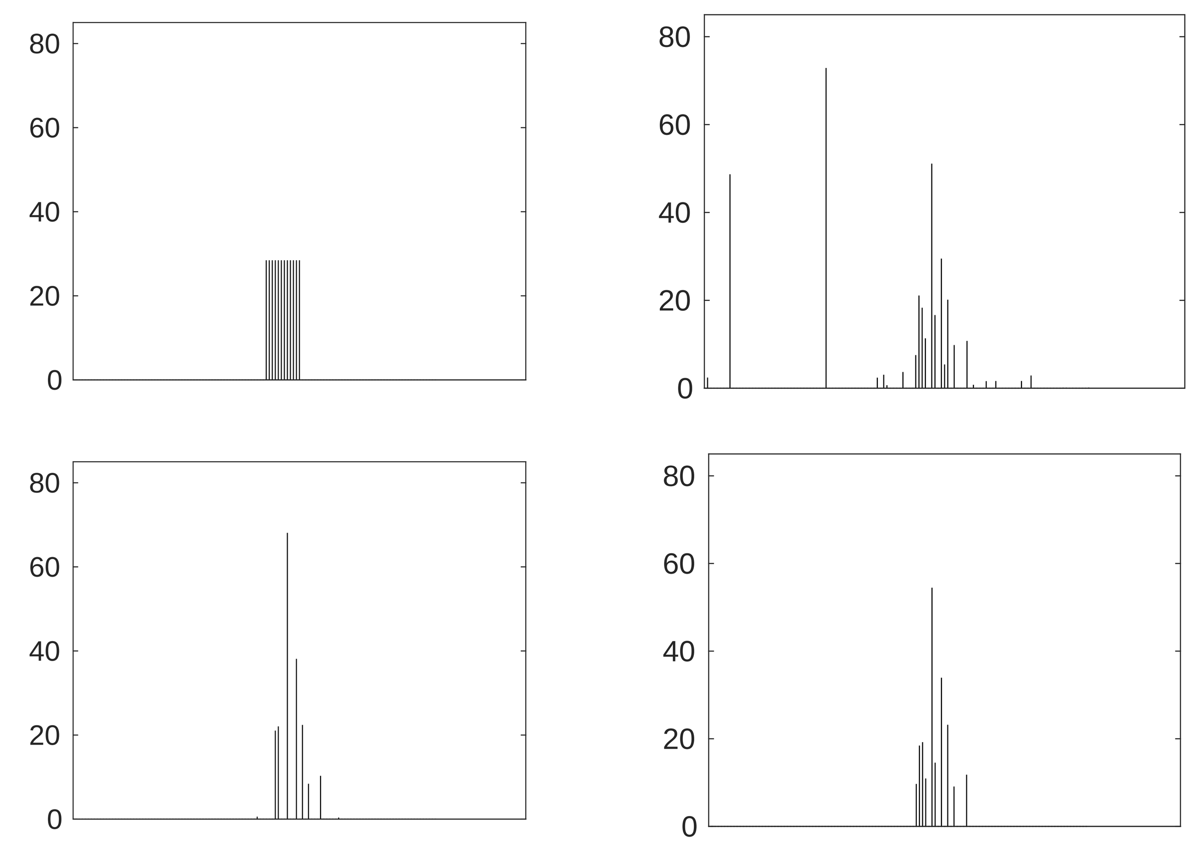


We set parameters to $\alpha_S = 2$, $\alpha_T = 1$, $t_{max} = 1$ and slightly modified s_{max} by considering

$$d_S(z_i, z_j) := \begin{cases} \exp(-\alpha_S\|(z_i^x, z_i^y) - (z_j^x, z_j^y)\|) & \text{if } \exp(-\alpha_S\|(z_i^x, z_i^y) - (z_j^x, z_j^y)\|) > 10^{-5}, \\ 0 & \text{otherwise.} \end{cases}$$

We can compare the original release with the solutions of the methods introduced above.

Figure 2: Original release (Upper Left plot), solution of (8) (Upper Right), (11) (Lower Left), (13) (Lower Right).



It is clear from Figure 2UR that the WLS solution is not sparse. It can be seen that the solutions of (11) , (13) estimate the true release plotted in Figure 2UL in a good way. In particular, the time profile of the release is very similar to the true one. This holds true especially for Figure 2LR which is based on the sparse optimization technique with maximal allowed sparsity $k_{tol} = 10$. This complies Table 1 where one can see that for $k_{tol} = 10$, error $\|WMx - Wy\|$ is the lowest possible one.

Table 1: Methods comparison.

Method	Error $\|WMx - Wy\|$	Sparsity $\|x\|_0$
WLS (8)	9.77e-12	30
l_1 approximation (11)	9.41e-12	9
Mixed-integer problem (13)	9.38e-12	10

A reference

- L. Adam and M. Branda: Sparse optimization for inverse problems in atmospheric modelling. Optimization Online (submitted).

Data and codes

We emphasize that all data and Matlab codes are available online at

- *http://staff.utia.cas.cz/adam/research.html*

ACKNOWLEDGEMENTS: This research is supported by EEA/Norwegian Financial Mechanism under project 7F14287 STRADI.

WWW (with manuscript and codes): *http://staff.utia.cas.cz/adam/research.html*