Supplementary information to: Atmospheric Inverse Modeling via Sparse Reconstruction

Nils Hase, Scot Miller, Peter Maaß, Justus Notholt, Mathias Palm, Thorsten Warneke

1 Ill-posed inverse problems

In this section we characterize ill-posed inverse problems and illustrate their properties. To keep it simple, we consider a linear inverse problem

$$Ax = y_\delta. \tag{1}$$

of finite dimensions, i.e. the forward model can be represented by a matrix $A \in \mathbb{R}^{m\times n}$. A linear inverse problem is called well-posed in the sense of Hadamard, if

(a) there exists a solution $x \in \mathbb{R}^n$ for all data $y_\delta \in R^m$,

(b) the solution is unique and

(c) the inversion is stable, i.e. the solution depends continuously on the data.

If one of the above properties is not fulfilled, the problem is called ill-posed.

Typically, the forward model A is unable to explain the noise on the data. Consequently, (a) is violated. Also, in most problems multiple measurements of the same parameters will not give the same answer, which leads to an overdetermined linear system without solution.

To guarantee the existence of a solution, one can replace the linear system (1) by the least squares approach

$$x^* = \arg\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax - y_\delta\|_2^2. \tag{2}$$

A minimum can be found, where the derivative of the least squares functional, Eq. (2), is zero. This gives

$$\begin{aligned} & A^t(Ax - y_\delta) = 0 \\ \Leftrightarrow \qquad & A^tAx = A^ty_\delta. \end{aligned} \tag{3}$$

Equation (3) is called the normal equation. If the matrix A does not have full rank, the matrix A^tA is not invertible. Consequently, there exists a nontrivial nullspace and the least squares solution is nonunique. Also, the linear system (1) is underdetermined; there is not enough information in the data to fully determine all parameters of the solution.

To avoid nonuniqueness of the least squares solution, one can select one of these solutions. A common choice is the minimum norm solution

$$x_{MNS} = \arg\min_{x^*} \|x^*\|.$$

The inverse mapping, that maps any data vector y_δ to the uniquely defined minimum norm solution x_{MNS}, is referred to as the pseudoinverse.

We derive an expression for the minimum norm solution based on singular value decomposition, which states that every real matrix A can be decomposed into

$$A = U\Sigma V^t, \tag{4}$$

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are unitary matrices, i.e. $U^t = U^{-1}$, $V^t = V^{-1}$, and $\Sigma \in \mathbb{R}^{m\times n}$ is a matrix of zeros with the singular values on its diagonal. The singular values σ_k are positive and in descending order, i.e. $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$, where $r = rank(A)$. The columns of U and V, u_1, u_2, ..., u_m and v_1, v_2, ..., v_n, are the left and right singular vectors, respectively. Left and right singular vectors each form an orthonormal set.
As in the article we assume for the noisy data

$$y_\delta = Ax^+ + \delta$$

with the true solution x^+. The given data thus consists of the simulated data to the true solution and a noise component. In the following, we derive an expression for the least squares solution and illustrate consequences of properties (b) and (c) on the solution.

With the help of the singular value decomposition, Eq. (4), the minimum norm solution can be expressed by

$$\begin{aligned}
x_{MNS} &= \sum_{k=1}^{r} \frac{1}{\sigma_k}\langle u_k, y_\delta\rangle v_k \\
&= \sum_{k=1}^{r} \frac{1}{\sigma_k}\langle u_k, Ax^+ + \delta\rangle v_k \\
&= \sum_{k=1}^{r} \frac{1}{\sigma_k}\left(\langle u_k, \sum_{k=1}^{r} \sigma_k \langle v_k, x^+\rangle u_k\rangle + \langle u_k, \delta\rangle\right) v_k \\
&= \sum_{k=1}^{r} \langle v_k, x^+\rangle v_k + \sum_{k=1}^{r} \frac{1}{\sigma_k}\langle u_k, \delta\rangle v_k.
\end{aligned}$$

The minimum norm solution consists of a solution term and a noise term. The solution term is the projection of the true solution x^+ onto the subspace spanned by the r right singular vectors. If the matrix A has full rank n, the solution term gives the true solution. The noise term acts in the same subspace as the solution term, but the coefficients are defined by the scalar product of noise and left singular vectors and the factors $\frac{1}{\sigma_k}$. For many ill-posed inverse problems, the singular values decrease rapidly towards zero. Thus, the noise in the data is largely amplified leading to estimates x_{MNS} that are dominated by noise.

For finite dimensional problems, the stability of the pseudoinverse can be measured by the condition number, $cond(A) = \frac{\sigma_1}{\sigma_r}$. The condition number ranges from one to technically infinity and describes the maximal noise amplification. Ill-posed inverse problems violating property (c) have an infinite condition number. Problems with finite condition number might technically be well-posed but are called ill-conditioned if the condition number is large. Just as ill-posed problems, ill-conditioned problems require stabil inversion methods.

The existence and the uniqueness property of a linear ill-posed inverse problem can be fixed defining the minimum norm solution of the least squares problem. However, the instability of the inversion requires special techniques such as Tikhonov regularization or Bayesian inversion.

2 Emission estimates of other methods

Our atmospheric inverse modeling szenario is similar to the one used by Miller et al. (2014). This allows a comparison between our inversion methods and the ones used in the previous work. Figure 1 shows the estimates for the methods studied by Miller et al. (2014) in the same colormap as in the main article.

3 Pseudocode for sparse dictionary reconstruction

The standard sparse reconstruction problem is given by

$$x^* = \arg\min_x \frac{1}{2}\|Ax - y_\delta\|_2^2 + \alpha\|x\|_1. \tag{5}$$

The Fast Iterative Shrinkage Thresholding Algorithm (FISTA) (Beck and Teboulle, 2009) is an efficient solver for Eq. (5) for a fixed regularization parameter α. For the suggested method L1 DIC POS, which we refer to as sparse dictionary reconstruction, we need to solve the problem

$$x^* = D\left(\arg\min_c \frac{1}{2}\|L_\delta(ADc - y_\delta)\|_2^2 + \alpha\|c\|_1\right).$$

We apply FISTA with $\tilde{A} = L_\delta AD$ and $\tilde{y} = L_\delta y_\delta$. To determine a suitable regularization parameter, Morozov's discrepancy principle is applied. For this, we need to solve

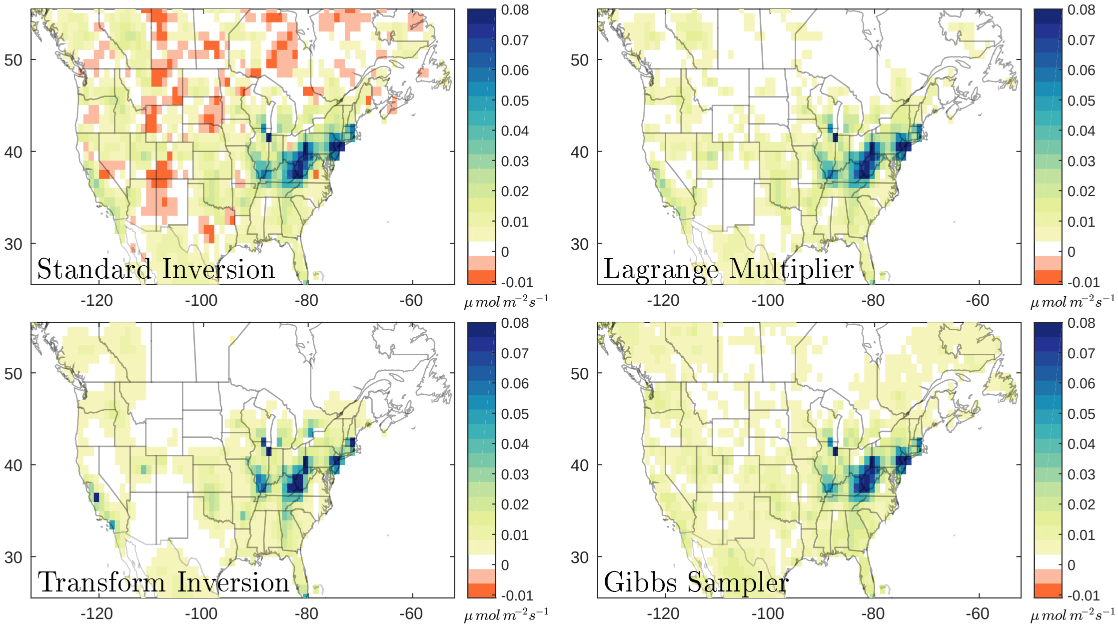

Figure 1: Emission estimates inverted from noisy simulated methane measurements using the Standard Inversion, Lagrange Multiplier method, Transform Inversion and the Gibbs Sampler as examined by Miller et al. (2014). The true flux field from the EDGAR inventory is shown in the article.

a sequence of problems (5). We force nonnegative parameters by setting all negative parameters to zero in each iteration of FISTA. Algorithm 1 gives a detailed pseudocode.

We suggest to set $q = 0.9$, $\tau = 1.05$ and $max\_iter = 250$. $c_0 \in \mathbb{R}^N$ is typically just a vector of zeros. The regularization parameter α needs to be chosen large enough, such that the discrepancy principle in line 25 is not fulfilled in the first execution of the while-loop. However, choosing it too large will require more executions of the while-loop. In order to simplify the notation, we introduced $K = L_\delta AD$. For large scale problems, explicit computation of K in line 3 and its norm in line 5 might be impossible. However, the matrix is never needed explicitly. The matrix-vector product Kc (and $K^t y$) can instead be performed in several steps $Kx = L_\delta(A(Dc))$. Often, the matrices L^δ, A and D are sparse (having many zero elements). Sparse matrix-vector products can speed-up the execution. It is sufficient to estimate the spectral norm of K from above using the Frobenius-norm.
As mentioned in the article the projection step in lines 17*ff.* is critical. It involves the problem of finding a representation c in the dictionary to represent x^+. As we are interested in a sparse representation the most direct way would be to solve

$$c^+ = \arg\min_c \frac{1}{2}\|Dc - x^+\|_2^2 + \alpha\|c\|_1. \tag{6}$$

We avoid solving another inverse problem to speed up the calculation. As our dictionary includes the pixel basis, finding a dictionary representation c for any parameters x is straightforward by working only on a subset I of all dictionary elements using only those elements that represent the pixel basis. An identical mapping yields the dictionary representation, $c_i = x_i,\ i \in I,\ c_i = 0,\ i \notin I$. For this approach there is no guarantee that the representation is sparse. However, the representation for an update of the sinks c^- should not have too many nonzero entries as no sinks are expected in the final estimate.

4 Source code

The most relevant parts of the source code to calculate emission estimates and generate some basic graphics are available as a supplement. The code is written in Matlab 2014b and commented. It consists of five data files, three scripts and a number of functions that are called by the scripts for calculation and plotting. We recommend starting the scripts to begin with. These are:

- *inverse_methods.m*:
  This script produces the figures with the surface flux estimates for all inversion methods. The estimates for the Tikhonov based methods can either be calculated, which requires a couple of minutes ($< 15\ min$), or they can just be loaded from the data files (set flag $load\_saved = 1$).

- *compare_results.m*:
  This script loads the estimates for all methods and calculates local, regional and total measures of error.

- *barnett_scenario.m*:
  This script produces the figures for the Barnett scenario for all inversion methods. Similarly to *inverse_methods.m*, the estimates for the Tikhonov based methods can either be calculated or they can just be loaded from the data files (set flag $load\_saved = 1$).

For further questions please contact N. Hase (nilshase@math.uni-bremen.de).

Algorithm 1 Sparse dictionary reconstruction

function $[x^*, \alpha^*] = \mathrm{L1\ DIC\ POS}(A, D, L_\delta, y^\delta, x_a)$
Set c_0, $\alpha > 0$, $q < 1$, $\tau > 1$, max_iter, $exit\_flag = 0$
$K = L_\delta A D$
$z = L_\delta(y^\delta - Ax_a)$
$\beta = \frac{1}{\|K\|^2}$
while $exit\_flag == 0$ **do**
$c_{old} = c_0$
$c = c_0$
for $k = 1, 2, ..., max\_iter$ **do**
\# Update step
$\hat{c} = c + \frac{k-2}{k+1}(c - c_{old}),\ c_{old} = c$
$c = \hat{c} - \beta K^t(K\hat{c} - z)$
\# Shrinkage step
for $i = 1, 2, ..., dim\_c$ **do**
$c_i = sign(c_i) max(|c_i| - \alpha\beta, 0)$
end for
\# Projection step
$x = Dc$
$x^+ = P_+(x),\ x^- = x - x^+$
$x^- \rightarrow c^-$
$c = c - c^-$
end for
$x^* = Dc + x_a,\ \alpha^* = \alpha$
\# Morozov's discrepancy principle
if $\|L_\delta(Ax^* - y^\delta)\|_2 < \tau\sqrt{dim\_y}$ **then**
$exit\_flag = 1$
else
$\alpha = q\alpha$
end if
end while
return x^*, α^*
end function