Parallel speedup analysis of an adjoint ensemble-based source identification algorithm

Alexey Penenko$^{1,2}$ and Alexander Gochakov$^3$

$^1$Institute of Computational Mathematics and Mathematical Geophysics SB RAS, ICMMG SB RAS, prospekt Akademika Lavrentjeva 6, Novosibirsk, 630090, Russian Federation
$^2$Novosibirsk State University, NSU, Pirogova st 1, Novosibirsk, 630090, Russian Federation
$^3$Siberian Regional Scientific-Research Hydrometeorological Institute, Sovetskaya s 30, Novosibirsk, 630099, Russian Federation

E-mail: aleks@ommgp.sscc.ru

Abstract. In the approach to the inverse problems, based on the sensitivity operators, the inverse problem is transformed into the family of quasi-linear operator equations. The sensitivity operator is constructed of the ensemble of the adjoint problem solutions corresponding to different adjoint problems source functions. Hence, from the theoretical point of view, sensitivity operator-based algorithms can be efficiently parallelized. The paper’s objective is to carry out the experimental study of the algorithm computation time for the different number of computation cores.

1. Introduction

The sensitivity operator-based approach [1, 2] seems to be a universal tool for solving various inverse modeling problems. In the test problems, inspired by the environmental [1, 2] and biological inverse modeling problems [3], the developed algorithms show relatively good performance. The approach takes the solution of ensembles of independent adjoint problems. Therefore it can be naturally parallelized. The paper’s objective is to present the results of the experimental parallelization efficiency study of the current implementation of the adjoint ensemble-based source identification algorithm.

2. Methods

2.1. Algorithm

A convection-diffusion-reaction model for $l = 1, \ldots, N_c$ is considered in the domain $\Omega_T = \Omega \times (0, T)$, where $\Omega$ is a sufficiently smooth approximation of the bounded rectangular domain $[0, X] \times [0, Y]$ in $\mathbb{R}^2$, $T > 0$. The domain $\Omega_T$ is bounded by $\partial\Omega_T = \partial\Omega \times [0, T]$.

$$\frac{\partial \varphi_l}{\partial t} - \nabla \cdot (\text{diag} (\mu_l) \nabla \varphi_l - u_l \varphi_l) + P_l(t, \varphi) \varphi_l = \Pi_l(t, \varphi) + f_l + r_l, \quad (x, t) \in \Omega_T, \quad (1)$$

$$u \cdot (\text{diag} (\mu_l) \nabla \varphi_l) + \beta_l \varphi_l = \alpha_l, \quad (x, t) \in \Gamma_{\text{out}} \subset \partial\Omega \times [0, T], \quad (2)$$

$$\varphi_l = \alpha_l, \quad (x, t) \in \Gamma_{\text{in}} \subset \partial\Omega \times [0, T], \quad \quad \quad (3)$$

$$\varphi_l = \varphi^0_l, \quad x \in \Omega, \quad t = 0, \quad \quad \quad (4)$$
where \( N_c \) is the number of considered substances, \( \varphi_l = \varphi_l(x, t) \) denotes the concentration of the \( l^{th} \) substance at a point \((x, t) \in \Omega_T\), \( \varphi \) is the vector of \( \varphi_l(x, t) \) for \( l = 1, \ldots, N_c \) will be called the state function, \( L = \{1, \ldots, N_c\} \). The functions \( \mu_l(x, t) \in \mathbb{R}^2 \) correspond to the diffusion coefficients, \( \text{diag}(a) \) is the diagonal matrix with the vector \( a \) on the diagonal, \( u(x, t) \in \mathbb{R}^2 \) is the underlying flow speed. \( \Gamma_{in} \) and \( \Gamma_{out} \) are parts of domain boundary \( \partial \Omega_T \) in which the vector \( u(x, t) \) points inwards the domain \( \Omega_T \) and is zero or points outwards the domain \( \Omega_T \) correspondingly, \( n \) is the outer normal. The functions \( \alpha_l(x, t), \varphi_0^l(x) \) are boundary and initial conditions, correspondingly, \( f_l(x, t) \) is the \textit{a priori} known source function, \( r_l(x, t) \) is a source function to be determined with the inverse problem solution (the uncertainty function). Let \( r \in Q \), where \( Q \) is the set of admissible sources such that the direct problem has a solution. Loss and production operator elements \( P_l, \Pi_l : [0, T] \times \mathbb{R}^{N_c} \rightarrow \mathbb{R}^+ \) are defined by the transformation model (in the considered case of the chemical transformations, they are polynomials with positive coefficients depending on time). We suppose all the functions and model parameters are smooth enough for the solutions to exist and the further transformations to make sense. Direct problem: given \( f_l, r_1, \mu_l, u_l, \alpha_l, \varphi_0^l \), find \( \varphi \) from (1) – (4).

Let there be an “exact” source function \( r^{(*)} \) to be found and \( L_{\text{meas}} \) denote the set of indices of the measured substances. We consider the time series of concentrations \( \{\varphi_l^{(*)}(x, t) | t \in [0, T], x \in \chi, l \in L_{\text{meas}}\} \) in a given set of measurement sites \( \chi \). Here \( \varphi^{(*)} \) is the solution of the direct problem with the source function \( r^{(*)} \).

For the solution of the source identification problem, we use the algorithm based on the ensembles of the adjoint problem solutions, described in [1, 2]. In the abstract form the relation between the model state function variation and uncertainty function variation is given by the sensitivity relation

\[
\left\langle S[r^{(2)}, r^{(1)}]; H \right| r^{(2)} - r^{(1)} \right\rangle_Q = \langle H, \delta \varphi \rangle_H,
\]

where \( S[r^{(2)}, r^{(1)}]; H \) denotes the sensitivity function which is calculated by the solution of the adjoint problem (the details can be found in [2]). The solution of the adjoint problem is determined by its source function \( H \), which is defined by the measurement operator.

If we consider a set of functions \( U = \{H^{(\xi)}\}_{\xi=1}^{\Xi} \), then combining the corresponding relations we obtain a sensitivity operator relation

\[
M_U \left[ r^{(2)}, r^{(1)} \right] \left( r^{(2)} - r^{(1)} \right) = H_U \varphi \left[ r^{(2)} \right] - H_U \varphi \left[ r^{(1)} \right],
\]

where

\[
M_U \left[ r^{(2)}, r^{(1)} \right] = \sum_{\xi=1}^{\Xi} e^{(\xi)} \left\langle S[r^{(2)}, r^{(1)}]; H^{(\xi)} \right| z \right\rangle_Q,
\]

\[
H_U \varphi = \sum_{\xi=1}^{\Xi} e^{(\xi)} \left\langle H^{(\xi)}, \varphi \right| H \right\rangle.
\]

The adjoint functions needed to compose the sensitivity operator can be evaluated in parallel, as an ensemble. If \( r^{(*)} \) is the exact solution of the source identification problem, \( I \) is the measurement data, aggregated in the state-function form, and \( \delta I \) is its perturbation, then for any \( U \) and \( r \) the relation holds:

\[
M_U \left[ r^{(*)}, r \right] \left( r^{(*)} - r \right) = H_U I + H_U \delta I - H_U \varphi \left[ r \right].
\]

This quasi-linear operator equations are solved based on the regularized TSVD Newton-Kantorovich method. The ensemble of the adjoint problem solutions is optimized according
to the maximal projection of the initial discrepancy to the appropriate trigonometric cosine-basis. Analogous algorithms have been applied to the linear passive transport problem [4] and the nonlinear transport-transformation model with point-wise sources and in situ measurements [5].

In the current statement, we can consider several relevant inverse modeling tasks. The obvious one is the inverse source identification problem stated above. The second one is the “continuation” problem, which consists of reconstructing $\varphi^{(\ast)}$ using the inverse source problem as an auxiliary one. It can be called a “continuation” problem because we continue the measurements to the whole domain using the advection-diffusion-reaction model. Hence, the obtained concentration field obeys the advection-diffusion-reaction model.

### 2.2. Inverse modeling scenario

For example, we consider the RADM2 transformation mechanism [6] with 61 reacting species with time-dependent reaction rates [7, 8]. To prepare the coefficients of the chemical transport model (1)-(4), the meteorological parameters $\mathbf{u}$ and $\mu$ are calculated with the WRF model [9] in the area limited by geographic coordinates 54.75$^\circ$ - 55.16$^\circ$, 82.66$^\circ$ - 83.37$^\circ$, corresponding to Novosibirsk city. The calculations were made for the model period 12:00-14:00 July 09, 2019. To obtain 2D spatial wind speed fields, the 3D WRF fields were vertically averaged. The domain parameters for the numerical inverse problem solution are $X = 35613m$, $Y = 34697m$, $T = 2 \times 3600s$ and the grid parameters are: $N_x = 50$, $N_y = 49$, $N_t = 232$.

The sources were located in the places of the roads of the city, marked by contours in figure 1 (a). The sources emitted NO with constant rates, which were proportional to the total number of cars in the specified location for the whole interval. 2GIS Company provided information about the road traffic intensity. The emitted substance name and the constant emission regime are known in the inverse modeling scenario. Emission rates are a priori considered as non-negative. Another information about the sources of emission was considered as unavailable.

The locations of the measurement sites (marked by red circles in figure 1 (a)) were taken from the state report [10]. Measurement sites provide time-series of the concentrations for the model time interval. We consider 720 and 70 ensemble members, correspondingly. The number of ensemble members describes the amount of data elements that are used by the algorithm. In the numerical experiments, the emitted substance NO concentration were measured.

### 3. Results

![Figure 1](image-url)  
**Figure 1.** “Exact” solution of the source identification problem a). Source identification results with the different numbers of the adjoint ensemble members: $\Xi = 720$ b) and $\Xi = 70$ c).
First, we shall compare the reconstruction results for the different numbers of the ensemble members: $\Xi = 70$ and $\Xi = 720$. This number regulates the dimensionality of the finite-dimensional equivalent of the operator equations that are solved in the algorithm. The results of the emission source identification for the direct measurements are presented in figure 1, and the results for the “continuation” problem solution are presented in figure 2. Since the concentration field is the dynamical picture, in figure 2, we show the distribution of mean concentrations with respect to time. The relative error dynamics is presented in figure 3. In figure 3 (a) and (b), the relative error is shown with respect to computation time measured in direct problem solution times ($t_{\text{inverse}}/t_{\text{direct}}$). “Background” in (c) denotes the solution of the direct problem with zero emission sources.

In the next experiment, we constrained the computation time by one hour, in the sense that the algorithm stops on the first iteration that exceeds one hour computation time. The experiment was carried out on the server with two Intel Xeon Gold-5220 processors with 18 cores on each processor (72 computation threads in total) and 256 GB of RAM. The results are presented in Tables 1 and 2. The first four columns contain the primary computation process
Table 1. Computation time characteristics for different number of computation cores, Ξ = 720. Common last iteration is 3.

| threads | $t_{\text{inverse}}$ | # iter. | mean iteration time | mean iteration speedup | $t_{\text{inverse}}$ at common last iter. | speedup at common last iter. |
|---------|----------------------|---------|---------------------|------------------------|----------------------------------------|-----------------------------|
| 72      | 3746.80              | 21      | 178.42              | 11.14                  | 400.63                                 | 14.88                       |
| 36      | 3652.67              | 20      | 182.63              | 10.88                  | 407.38                                 | 14.63                       |
| 18      | 3663.26              | 15      | 244.22              | 8.14                   | 520.74                                 | 11.45                       |
| 9       | 3614.15              | 10      | 361.42              | 5.5                    | 799.78                                 | 7.45                        |
| 3       | 3723.52              | 4       | 930.88              | 2.13                   | 2481.42                                | 2.40                        |
| 1       | 5960.60              | 3       | 1986.87             | 1.00                   | 5960.60                                | 1.00                        |

Table 2. Computation time characteristics for different number of computation cores, Ξ = 70. Common last iteration is 11.

| threads | $t_{\text{inverse}}$ | # iter. | mean iteration time | mean iteration speedup | $t_{\text{inverse}}$ at common last iter. | speedup at common last iter. |
|---------|----------------------|---------|---------------------|------------------------|----------------------------------------|-----------------------------|
| 72      | 3621.74              | 94      | 38.53               | 8.9                    | 467.75                                 | 8.06                        |
| 36      | 3616.21              | 110     | 32.87               | 10.43                  | 391.88                                 | 9.63                        |
| 18      | 3636.66              | 89      | 40.86               | 8.39                   | 477.02                                 | 7.91                        |
| 9       | 3647.88              | 66      | 55.27               | 6.20                   | 650.52                                 | 5.80                        |
| 3       | 3737.74              | 26      | 143.76              | 2.36                   | 1616.89                                | 2.33                        |
| 1       | 3771.83              | 11      | 342.89              | 1.00                   | 3771.83                                | 1.00                        |

characteristics. The first one contains the number of the CPU threads used, $t_{\text{inverse}}$ is the computation time for solving inverse problems, # iter is the number of complete iterations of the inverse problem solution algorithm. We measured the parallel speedup in two ways. First, we compared the mean iteration time by dividing the computation time by the number of iterations. Since the iterations are not equivalent in time, we used the second criteria. We found the smallest number of iterations (# iter.) that fit a one-hour computation interval (common last iteration) for the experiments with the different number of computation threads and compare the times ($t_{\text{inverse}}$ at common last iter.) when this iteration number was achieved. The speedup for N threads is calculated as the computation time characteristic for a single thread divided by the corresponding characteristic for N threads.

4. Discussion
In figure 1,2 and 3, we can see that taking more ensemble members allows improving the solution quality, but in turn, it takes more time to obtain the needed result quality. The speedup for 36 and 72 threads is almost the same. Analyzing the computation time, we can conclude that the speedup for Ξ = 720 is higher than for Ξ = 70. During the iteration, the sensitivity operator computations consume most of the time. From the theoretical point of view, this speedup is still seemed to be far from the perfect one, since the adjoint equations can be theoretically evaluated in parallel. So the current implementation of the algorithm still has room for optimization.
Currently, the parallelization is carried out on a single time step of the adjoint ensemble calculation. The computations on every time step include parallel evaluation of common adjoint equation scheme coefficients and parallel evaluation of the ensemble of adjoint equations solutions using the calculated coefficients. The first part is much shorter than the second one. Therefore the switching between the fragments takes a relatively long time. Hence, parallelization efficiency can be improved by enlarging the coefficient calculation section and considering several time steps.

5. Conclusion
In the paper, we measured the parallelization efficiency of the current version of the source identification algorithm based on the adjoint ensemble members. The speedup and the benefit of using parallel computation technologies are clearly observable, but there is still significant room for further algorithm optimizations.

Acknowledgments
The construction of the scenario for modeling, the adaptation of the algorithm to the Novosibirsk city conditions, and numerical experiments were carried out with the financial support of the Russian Foundation for Basic Research and the Novosibirsk Region Government within the framework of the scientific project No. 19-47-540011. 2GIS Company granted the information about the road traffic intensity. The continuation problem analysis was supported by Russian Foundation for Basic Research project No. 20-01-00560. The development of the algorithm modification with weaker misfit monotonicity and its analysis were supported by State Budget Program No 0315-2019-0004.

References
[1] Penenko V V, Penenko A V, Tsvetova E A and Gochakov A V 2019 Journal of Applied Mechanics and Technical Physics 60 392–399
[2] Penenko A 2020 Inverse Problems & Imaging 14 757–782
[3] Penenko A, Nikolaev S, Golushko S, Romashenko A and Kirilova I 2016 Mathematical Biology and Bioinformatics 11 426–444 (In Russian)
[4] Issartel J P 2005 Atmospheric Chemistry and Physics 5 249–273
[5] Mamonov A V and Tsai Y H R 2013 Inverse Problems 29 035009
[6] Stockwell W R, Middleton P, Chang J S and Tang X 1990 Journal of Geophysical Research 95 16343
[7] Goris N 2011 Singular Vector Based Targeted Observations of Atmospheric Chemical Compounds Ph.D. thesis Universität zu Köln
[8] Saunders S M, Jenkin M E, Derwent R G and Pilling M J 2003 Atmospheric Chemistry and Physics 3 161–180
[9] Skamarock W C, Klemp J B, Dudhia J, Gill D O, Barker D M, Duda M G, Huang X Y, Wang W and Powers J G 2008 A Description of the Advanced Research WRF Version 3 Mesoscale and Microscale Meteorology Division National Center for Atmospheric Research Boulder, Colorado, USA near technical note ed
[10] Department of natural resources and environmental protection of the Novosibirsk region 2016 On the state and protection of the environmental of the novosibirsk region in 2015 State report of department of natural resources and environmental protection of the Novosibirsk region Novosibirsk (In Russian) URL http://www.nso.ru/sites/test.new.nso.ru/wodby_files/files/wiki/2014/01/korrektura_gosdoklad-2015.compressed.pdf