A two-stage numerical approach for the sparse initial source identification of a diffusion–advection equation*

OPEN ACCESS
IOP Publishing
Inverse Problems
Inverse Problems **39** (2023) 095003 (30pp)
https://doi.org/10.1088/1361-6420/ace548

A two-stage numerical approach for the sparse initial source identification of a diffusion–advection equation*

Umberto Biccari[2,3], **Yongcun Song**[1,3], **Xiaoming Yuan**[3,**] **and Enrique Zuazua**[1,2,4]

[1] Chair for Dynamics, Control, Machine Learning and Numerics, Alexander von Humboldt-Professorship, Department of Mathematics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen 91058, Germany
[2] Chair of Computational Mathematics, Fundación Deusto, Avenida de las Universidades 24, 48007 Bilbao, Basque Country, Spain
[3] Department of Mathematics, The University of Hong Kong, Pok Fu Lam, Hong Kong, People's Republic of China
[4] Departamento de Matemáticas, Universidad Autónoma de Madrid, 28049 Madrid, Spain

E-mail: xmyuan@hku.hk

Received 6 April 2023; revised 14 June 2023
Accepted for publication 7 July 2023
Published 28 July 2023

Abstract

We consider the problem of identifying a sparse initial source condition to achieve a given state distribution of a diffusion–advection partial differential equation after a given final time. The initial condition is assumed to be a finite combination of Dirac measures. The locations and intensities of this initial

* This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No: 694126-DyCon). The work of U B and E Z is partially supported by the Grant PID2020-112617GB-C22 KILEARN of MINECO (Spain) and the Elkartek Grant KK-2020/00091 CONVADP of the Basque Government. E Z has been funded by the Alexander von Humboldt-Professorship program, the ModConFlex Marie Curie Action, HORIZON-MSCA-2021-DN-01, the COST Action MAT-DYN-NET, the Tran- sregio 154 Project "Mathematical Modelling, Simulation and Optimization Using the Example of Gas Networks" of the DFG, grants PID2020-112617GB-C22 and TED2021-131390B-I00 of MINECO (Spain), and by the Madrid Goverment – UAM Agreement for the Excellence of the University Research Staff in the context of the V PRICIT (Regional Programme of Research and Technological Innovation). The work of: X Y is supported by Seed Fund for Basic Research (Project Number: 202011159106) from The University of Hong Kong.
** Author to whom any correspondence should be addressed.

condition are required to be identified. This problem is known to be exponentially ill-posed because of the strong diffusive and smoothing effects. We propose a two-stage numerical approach to treat this problem. At the first stage, to obtain a sparse initial condition with the desire of achieving the given state subject to a certain tolerance, we propose an optimal control problem involving sparsity-promoting and ill-posedness-avoiding terms in the cost functional, and introduce a generalized primal-dual algorithm for this optimal control problem. At the second stage, the initial condition obtained from the optimal control problem is further enhanced by identifying its locations and intensities in its representation of the combination of Dirac measures. This two-stage numerical approach is shown to be easily implementable and its efficiency in short time horizons is promisingly validated by the results of numerical experiments. Some discussions on long time horizons are also included.

Keywords: initial source identification, inverse problem, optimal control, sparse control, diffusion–advection equations, non-smooth optimization, primal-dual algorithm

(Some figures may appear in colour only in the online journal)

1. Introduction and motivations

Among various inverse problems arising in scientific computing, an important one is the identification of moving pollution sources in either compressible or incompressible fluids that can be described by diffusion–advection systems. See e.g. [15, 33] for accurate estimation of pollution sources in the environmental safeguard of a densely populated city, and [22, 34] for other related problems. As many contributions in the literature have shown [8, 9, 20, 34, 39], this kind of pollution source identification problems can be mathematically modeled by initial source identification problems of diffusion–advection systems. Besides, as pointed out in [8, 9, 15, 32, 34, 39], the initial source is usually assumed to be sparse, i.e. its support is zero in Lebesgue measure. In this paper, we consider the problem of identifying a sparse initial source condition to achieve a given state distribution of a diffusion–advection partial differential equation (PDE) after a given final time. The initial condition is assumed to be a finite combination of Dirac measures, and the locations and intensities of this initial condition are required to be identified.

1.1. Problem statement

Let $\Omega \subset \mathbb{R}^N$ with $N \geqslant 1$ be a bounded domain and $\partial\Omega$ its boundary. We consider the following linear diffusion–advection equation

$$\begin{cases} \partial_t u - d\Delta u + v \cdot \nabla u = 0, & (x,t) \in \Omega \times (0,T), \\ u = 0, & (x,t) \in \partial\Omega \times (0,T), \\ u(x,0) = u_0(x), & x \in \Omega, \end{cases} \tag{1.1}$$

where $0 < T < +\infty$ is a given final time, $d > 0$ is the diffusivity coefficient and the vector $v \in \mathbb{R}^N$ is the velocity field of the advection. Here and in what follows, d and v are both assumed to be constants for simplicity, although our analysis to be presented can be adapted to the case

where both diffusivity and velocity fields vary. We further assume the initial condition $u_0(x)$ to be a finite combination of Dirac measures

$$u_0(x) = \sum_{i=1}^{l} \alpha_i \delta_x(x_i), \quad x_i \in \Omega, \tag{1.2}$$

where $\{\alpha_i\}_{i=1}^l \in \mathbb{R}^l$ and $x_i \in \Omega, 1 \leqslant i \leqslant l$, are the intensities and locations, respectively, with $1 \leqslant l < +\infty$ the number of locations. The Dirac measure $\delta_x(x_i)$ is defined by $\delta_x(x_i) = 1$ if $x = x_i$, and $\delta_x(x_i) = 0$ otherwise. Note that (1.2) implies that the support of $u_0(x)$ is $\{x_i\}_{i=1}^l \subset \Omega$ and its Lebesgue measure is zero. With the assumption (1.2), one can show that there exists a unique solution u of (1.1) and u belongs to the space $L^r(0,T;W_0^{1,p}(\Omega))$ for all $p, r \in [1,2)$, with $\frac{2}{r} + \frac{N}{p} > N+1$, see [7] and the references therein.

Problem 1.1. Consider the diffusion–advection equation (1.1). Let u_T be a given or observed function. We aim at identifying an initial condition $\widehat{u}_0^*$ subject to (1.2), i.e.

$$\widehat{u}_0^*(x) = \sum_{i=1}^{l} \widehat{\alpha}_i^* \delta_x(\widehat{x}_i^*), \quad \text{with } \widehat{\alpha}_i^* \in \mathbb{R}, \widehat{x}_i^* \in \Omega$$

such that the corresponding final state $\widehat{u}^*(\cdot;T)$ of (1.1) is as close as possible to u_T, in the sense that for $\varepsilon > 0$ arbitrary small we have

$$\|\widehat{u}^*(\cdot;T) - u_T\|_{L^2(\Omega)} \leqslant \varepsilon, \quad \text{a.e in } \Omega. \tag{1.3}$$

Problem 1.1 plays an important role in various areas such as pollution sources identification, precision mechanical, industrial mechatronic, hydrologic inversion, and image deblurring. We refer to [41, 42] and references therein for more discussions. As well known (see, e.g. [28]), due to the strong diffusive and smoothing properties of equation (1.1), problem 1.1 is exponentially ill-posed, which means that a small perturbation on the data u_T may cause an arbitrarily large error in $\widehat{u}_0^*$. For instance, if we set $\Omega = [0,\pi], d = 1$ and $v = 0$ in (1.1) and consider a reachable target u_T, then addressing problem 1.1 amounts to solving

$$A_T u_0 := \sum_{n=1}^{\infty} e^{-n^2 T} \langle u_0, v_n \rangle v_n = u_T$$

with v_n defined by $v_n(x) = \sqrt{\frac{2}{\pi}} \sin(nx)$. Since $e^{-n^2 T} \to 0$ as $n \to +\infty$, we see that the operator A_T is compact, which in turn implies that the problem is ill-posed (more discussions on this specific issue can be referred to [2, 16]). Moreover, it is easy to see that if T becomes larger, the problem is increasingly ill-posed. Therefore, it is challenging to design some efficient numerical algorithms for solving problem 1.1.

1.2. State-of-the-art

In the literature, some work has already been done for sparse initial source identification problems, based on the natural idea of taking advantage of the sparse nature of the initial condition. A widely used strategy to address sparse initial source identification problems is to formulate them as optimal control problems modeled by PDEs, in which the initial condition is assumed to play the role of a control term. This is the seminal idea at the basis of some research articles, see e.g. [8, 9, 32, 39].

In [8], sparse optimal control techniques are used to identify sparse initial sources for diffusion-convection equations. The existence and uniqueness of optimal controls are proved, and necessary and sufficient optimality conditions are obtained. Based on these conditions, the sparsity structure of the optimal control is derived. In [9], the adjoint methodology for

sparse initial source identification problems governed by parabolic equations is introduced. It is proved that the sparse initial condition can be recovered by minimizing its measure-norm under the constraint that the corresponding solution and the given target are close at the final time. In [32], the identification of an unknown sparse initial source for a homogeneous parabolic equation is addressed by considering an optimal control problem, where the control variable is considered in the space of regular Borel measures and the corresponding norm is used as a regularization term in the objective functional. Under specific structural assumptions, the authors show that the initial source is a finite combination of Dirac measures as that in (1.2).

It is remarkable that, in the above references, the sparse initial source identification problems are formulated as optimal control problems in measure spaces that can be (equivalently) written as

$$\min_{u_0\in\mathcal{M}(\Omega)} J(u_0) := \frac{1}{2}\|u(\cdot,T)-u_T\|^2_{L^2(\Omega)} + \beta\|u_0\|_{\mathcal{M}(\Omega)}, \tag{1.4}$$

where $u(\cdot,T)$ is the solution at $t=T$ of equation (1.1) corresponding to u_0; $\beta>0$ is a regularization parameter; $\mathcal{M}(\Omega)=C_0(\Omega)^*$ denotes the space of regular Borel measures in Ω, with $C_0(\Omega)$ the space of continuous functions in Ω vanishing on $\partial\Omega$, and the norm in this space is defined by

$$\|u_0\|_{\mathcal{M}(\Omega)} = |u_0|(\Omega) = \sup\left\{\int_\Omega z\mathrm{d}u_0 \ | z\in C_0(\Omega), \|z\|_\infty \leqslant 1\right\},$$

$|u_0|$ being the total variation measure associated to u. Similar models can also be found in [13, 30] and the references therein for sparse peak deconvolution. The presence of measures can guarantee the sparsity of the initial source but entails appropriate discretization for measure-valued quantities and may invalidate the application of some well-known numerical methods. For instance, the first-order optimality condition of (1.4) cannot be reformulated in a non-smooth point-wise form and thus the well-known semi-smooth Newton (SSN) type methods cannot be applied directly, see e.g. [18, 27].

It is shown in [32] that, after some proper discretization, problem (1.4) can be reformulated as a finite-dimensional optimization problem with ℓ^1-regularization, for which various well-developed optimization algorithms can be applied directly. See [30] for related discussions on sparse peak deconvolution. However, in the context of optimal control of PDEs, such a direct application of finite-dimensional optimization algorithms may cause the so-called mesh-dependent issue, which means that the convergence behavior critically depends on the fineness of the discretization, see [32]. Hence, some new numerical algorithms that can be described on the continuous level have to be deliberately designed from scratch. In this regard, a primal-dual active point (PDAP) method is proposed in [32]. At each iteration of the PDAP, one entails the solutions of two parabolic equations to update the adjoint variable, an optimization subproblem to find a new support point, and a non-smooth optimization subproblem to compute a new iterate. This non-smooth optimization problem has no closed-form solution and can only be solved iteratively by some optimization algorithm, such as the SSN method suggested therein. Hence, nested iterations are resulted, which may cause some new challenges in the overall rigorous convergence and additional computational loads in the implementation.

To address problem 1.1, a two-stage numerical approach is proposed in [39]. First, problem 1.1 is formulated as an L^1-regularized optimal control problem, where the initial condition is treated as the control variable and is assumed to be in $L^1(\Omega)$ to promote the sparsity. As a result, measures are avoided. To solve the optimal control problem, a gradient descent (GD) method is suggested. Then, the optimal locations are identified by determining these local maxima/minima of the optimal control, and the corresponding optimal intensities are identified

by solving a least squares problem. Several test cases validate that this two-stage approach can accurately identify the sparse initial sources even in heterogeneous media. Despite this fact, we shall remark that the focus in [39] is on the development and discussion of the numerical algorithm, but from a mathematical viewpoint, the optimal control problem considered in [39] is not well-posed. In particular, since the control variable is considered in the non-reflexive space $L^1(\Omega)$, the existence of a solution in $L^1(\Omega)$ to the optimal control problem cannot be guaranteed. See [9, 45] for some related discussions.

In [34], sparse initial sources are identified from some sparsely sampled solutions of the heat equation, where the initial sources are assumed to satisfy (1.2). After some proper discretization, the initial source identification problem is formulated as a finite-dimensional constrained ℓ^1 minimization problem with respect to the initial condition, under the constraint that the corresponding final states of the discretized heat equation are close to the observations. The classical Bregman iteration method [3] combined with two acceleration strategies (support restriction and domain exclusion) is suggested to solve the constrained ℓ^1 minimization problem. The effectiveness and efficiency of this approach are validated by some numerical experiments, which show that, for two-dimensional spaces, one can recover the sparse initial condition accurately from some point-wise observations at the final time. The Bregman iteration method solves the constrained problem as a sequence of unconstrained subproblems that have no closed-form solutions and can only be solved iteratively. Thus, inner iterations have to be embedded into the implementation of the Bregman iteration method. Hierarchically nested iterations and hence the lack of rigorous analysis for the convergence of the overall scheme are thus caused. Moreover, as mentioned earlier, such a direct application of the Bregman method may lead to the mesh-dependent issue implying that the convergence depends strongly on the fineness of the discretization.

For completeness, we mention that other types of optimal control problems with sparsity properties have also been widely discussed in the existing literature. In [5, 10] for elliptic problems and in [6, 31] for parabolic problems, sparse controls are obtained by considering optimal control problems in the space of measures. Some L^1-regularized elliptic and parabolic optimal control problems are discussed in [44, 45]. The use of L^1-regularization has been shown to be efficient to obtain optimal controls with support in small regions of the domain; and the support can be adjusted by tuning the L^1-regularization parameter in the cost functional.

1.3. Our numerical approach

To address problem 1.1, we propose a new two-stage numerical approach, which consists of a sparsity promotion stage and a structure enhancement stage. Our approach keeps all advantageous features of the framework in [39] while avoids the aforementioned issues encountered therein. First, in the sparsity promotion stage, we treat the initial condition u_0 as a control variable and formulate problem 1.1 as an optimal control problem with L^2+L^1-regularization term. As to be shown in section 2, the presence of the L^1-regularization can promote the sparsity of the initial source. However, the identified initial source from the optimal control problem is not sparse as desired due to the smoothing property of the L^2-regularization term. Hence, a structure enhancement stage should be complemented to ensure that (1.2) holds while identify the locations $\{\widehat{x}_i^*\}_{i=1}^l$ and the intensities $\{\widehat{\alpha}_i^*\}_{i=1}^l$.

Concretely, we formulate problem 1.1 in terms of the following optimal control problem:

$$\min_{u_0\in L^2(\Omega)} J(u_0) := \frac{1}{2}\int_\Omega |u(\cdot,T)-u_T|^2\,\mathrm{d}x + \frac{\tau}{2}\int_\Omega |u_0|^2\,\mathrm{d}x + \beta\int_\Omega |u_0|\,\mathrm{d}x, \tag{1.5}$$

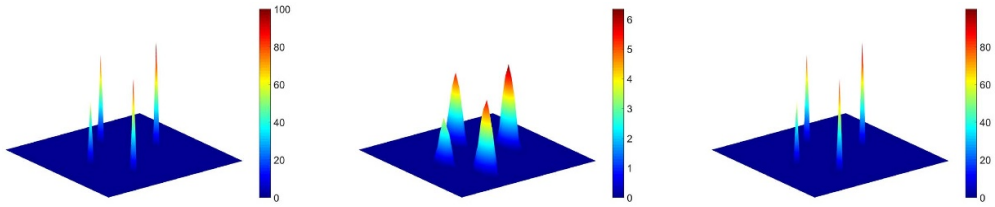


Figure 1. Reference initial datum $\widehat{u}_0$ (left), the recovered initial datum u_0^* (middle) by solving (1.5), and the recovered initial datum $\widehat{u}_0^*$ (right) by the two-stage numerical approach. ($\Omega=(0,2)\times(0,1)$, $T=0.01$, $d=1$, $v=(0,0)^\top$, $\tau=10^{-2}$ and $\beta=3\times10^{-1}$).

where $u(\cdot,T)$ is the solution at $t=T$ of equation (1.1) corresponding to u_0. In (1.5), the constants $\tau>0$ and $\beta>0$ are regularization parameters. Similar as the problem in [39], the first term of $J(u_0)$ seeks for an initial condition u_0 such that the corresponding final state of equation (1.1) is as close as possible to u_T; and the last term promotes the sparsity of the initial source. Meanwhile, inspired by [45], we introduce the L^2-regularization $\frac{\tau}{2}\int_\Omega |u_0|^2\,\mathrm{d}x$ to guarantee the well-posedness of (1.5) while improving the conditioning to allow for a more efficient numerical resolution. For any fixed $\tau>0$, as to be shown in section 2.2, we can always tune β to get an optimal control u_0^* with small support. Note that if $\tau=0$ and $u_0\in L^1(\Omega)$, problem (1.5) is not well-posed. To address this issue, a natural way is to consider $u_0\in\mathcal{M}(\Omega)$ and relax $\beta\int_\Omega|u_0|\,\mathrm{d}x$ to $\beta\|u\|_{\mathcal{M}(\Omega)}$ so that problem (1.4) is obtained. From this perspective, problem (1.5) can be viewed as a regularized version of (1.4); related discussions can be referred to [12].

Notice that the control variable u_0 in (1.5) is considered as a general function in $L^2(\Omega)$ and it is not assumed to satisfy (1.2). To identify the locations and intensities directly, one may further assume that $u_0(x)=\sum_{i=1}^l \alpha_i\delta_x(x_i)$, with $\alpha_i\in\mathbb{R}$ and $x_i\in\Omega$, in the formulation of (1.5). As a result, the intensities $\{\alpha_i\}_{i=1}^l$ and the locations $\{x_i\}_{i=1}^l$ become the control variables. However, this leads to a non-convex optimization problem which is challenging to be solved both in terms of theory and algorithms. Meanwhile, it causes practical difficulties related to the computation of the derivatives with respect to $\{x_i\}_{i=1}^l$. By contrast, problem (1.5) is convex and the computation of the derivatives with respect to u_0 is relatively easier.

Clearly, problem (1.5) operates in function spaces and avoids the employment of measures. As a consequence, it can be easily addressed numerically and various well-developed optimization algorithms can be applied directly. Furthermore, due to the introduction of the L^2-regularization term, problem (1.5) allows identifying the sparse initial sources much more efficiently than the one in [39], as to be validated in section 6. Notwithstanding that, due to the presence of the L^2-regularization term and its smoothing property, the recovered initial condition u_0 by solving (1.5) is not sparse as desired in (1.2).

To validate this fact, we set $\Omega=(0,2)\times(0,1)$, $T=0.01$, $d=1$, $v=(0,0)^\top$, $\tau=10^{-2}$ and $\beta=3\times10^{-1}$, then solve (1.5) by the primal-dual algorithm described in section 3. Additional details are presented in section 6. The numerical results are visualized in figure 1, where the left plot corresponds to the reference initial datum $\widehat{u}_0$ assigned *a priori* in the form of (1.2), while the middle plot shows the recovered initial datum u_0^* by solving (1.5). We can clearly see that $\widehat{u}_0$ and u_0^* do not coincide. In particular, the recovered initial datum u_0^* has a small support but it is not sparse as the reference $\widehat{u}_0$. The intensities of u_0^* are below the ones of $\widehat{u}_0$.

For the above reasons, once a numerical solution of (1.5) is computed, a structure enhancement stage exploiting (1.2) is necessary to identify the optimal locations $\{\widehat{x}_i^*\}_{i=1}^l$ and the

intensities $\{\widehat{\alpha}_i^*\}_{i=1}^l$. To this end, we propose to solve two simple and low-dimensional optimization problems. More precisely, to identify the optimal locations $\{\widehat{x}_i^*\}_{i=1}^l$, we consider an optimization problem in terms of the spatial variable $x \in \Omega$. Then, motivated by the facts that the initial source $\widehat{u}_0^*$ to be recovered is a finite combination of Dirac measures and the associated final state $\widehat{u}^*(\cdot, T)$ should be as close as possible to u_T, we solve a least squares problem to identify the optimal intensities $\{\widehat{\alpha}_i^*\}_{i=1}^l$. A two-stage numerical approach is thus proposed for solving problem 1.1. The right plot in figure 1 depicts the recovered initial datum $\widehat{u}_0^*$ by the two-stage numerical approach, which clearly is a highly accurate approximation to the reference initial datum $\widehat{u}_0$. Therefore, the proposed two-stage numerical approach allows identifying the sparse initial sources very accurately, even for some heterogeneous materials or coupled models as validated by some numerical experiments in section 6.

1.4. Primal-dual algorithms for the solution of (1.5)

Note that the identification of the optimal locations and intensities is based on the solution of (1.5). Thus it is crucial to solve (1.5) efficiently. Recall that (1.5) is modeled in function spaces. Hence, various well-developed optimization algorithms can be applied directly. For instance, SSN-type methods [46] and the alternating direction method of multipliers (ADMM) [17] can be conceptually applied and they indeed have been successful in solving some other types of optimal control problems in the literature (see [18, 19, 27] and the references therein). Nevertheless, we note that at each iteration of SSN and ADMM, a complicated large-scale and ill-conditioned saddle point system and an optimal control subproblem should be iteratively solved, respectively. Both of them are numerically challenging and expensive for such a time-dependent model. Consequently, some numerical algorithms tailored for these subproblems have to be deliberately designed. The same concerns apply to the Bregman iteration method in [34], which can also be considered for solving (1.5).

To avoid the above issues, we advocate the primal-dual algorithm proposed in [11], which has been widely used in various areas such as image processing, inverse problems, and statistical learning. As to be shown in section 3, when the primal-dual algorithm in [11] is applied to problem (1.5), the main computation at each iteration is solving only two PDEs which can be efficiently addressed by various well-developed PDE solvers. Hence, the implementation of the primal-dual algorithm in [11] is easy and computationally cheap for (1.5). To further speed up the convergence, we propose a generalized version of the primal-dual algorithm mainly by following the ideas in [21, 23, 25]. Moreover, we show that the generalized primal-dual algorithm performs significantly better than the GD described in [39] for the initial source identification procedure.

1.5. Organization

The rest of this paper is organized as follows. Some preliminaries including the existence and uniqueness of a solution, the first-order optimality condition, and the structural property of the solution are given in section 2. A generalized primal-dual algorithm and its implementation details for solving (1.5) are discussed in section 3, and its strong global convergence and worst-case convergence rate are analyzed in section 4. A structure enhancement stage is introduced in section 5 to identify the optimal locations and intensities. A two-stage numerical approach is thus proposed, and its efficiency is illustrated in section 6 through some numerical experiments. Finally, section 7 gathers some final remarks and future perspectives.

2. Preliminaries

In this section, we analyze some properties of the optimal control problem (1.5). First, the existence and uniqueness of an optimal control u_0^* are discussed. Then, we derive the optimality conditions and deduce some structural properties of u_0^*.

2.1. Analysis of the optimal control problem (1.5)

Let us start by discussing the existence and uniqueness of an optimal control u_0^* to (1.5). This comes from a very standard argument and can be easily obtained by adapting the proof of [9, lemma 2.3].

Theorem 2.1. *There exists a unique solution $u_0^* \in L^2(\Omega)$ of the optimal control problem (1.5).*

Then, using some similar arguments as those in [4, 45], we have the following result.

Theorem 2.2. *Suppose that $u_0^* \in L^2(\Omega)$ is the unique solution of the optimal control problem (1.5). Then, the following first-order optimality condition holds:*

$$\psi^*(\cdot,0) + \tau u_0^* + \lambda_{u_0}^* = 0, \tag{2.1}$$

where $\lambda_{u_0}^* \in \partial\varphi(u_0^*)$ with $\varphi(u_0^*) = \beta \int_\Omega |u_0^*| \mathrm{d}x$, and ψ^* is the corresponding adjoint variable that is the successive solution of the state equation (1.1) and the adjoint equation

$$\begin{cases} \partial_t \psi + d\Delta\psi + v\cdot\nabla\psi = 0, & (x,t) \in \Omega \times (0,T), \\ \psi = 0, & (x,t) \in \partial\Omega \times (0,T), \\ \psi(\cdot,T) = u(\cdot,T) - u_T := \psi_T, & x \in \Omega, \end{cases} \tag{2.2}$$

provided the initial datum u_0^*.

2.2. Structural properties of u_0^*

Recall that $\lambda_{u_0}^* \in \partial\varphi(u_0^*) = \beta\partial \int_\Omega |u_0^*| \mathrm{d}x$. Moreover, it follows from the results of [29] that

$$\lambda_{u_0}^* \in \beta \mathrm{sign}(u_0^*),$$

where the set-valued function $\mathrm{sign}(\cdot)$ is given by

$$\mathrm{sign}(v) = \begin{cases} \dfrac{v}{|v|}, & \text{if } v \neq 0, \\ \{\eta : |\eta| \leqslant 1\}, & \text{otherwise.} \end{cases}$$

Then, one can consider the optimality condition (2.1) for all $x \in \Omega$ and get a point-wise relation of u_0^* and $\psi^*(\cdot,0)$ as displayed in figure 2. To be concrete, for any $x \in \Omega$, we have

$$\begin{cases} u_0^*(x) = \dfrac{1}{\tau}(-\psi^*(x,0) - \beta), & \text{if } u_0^*(x) > 0, \\ u_0^*(x) = \dfrac{1}{\tau}(-\psi^*(x,0) + \beta), & \text{if } u_0^*(x) < 0, \\ |\psi^*(x,0)| \leqslant \beta, & \text{if } u_0^*(x) = 0, \end{cases}$$

which implies that

$$u_0^*(x) = -\mathrm{sign}(\psi^*(x,0)) \max\left\{ \frac{1}{\tau}\Big(|\psi^*(x,0)| - \beta\Big), 0 \right\}.$$

We thus have the following structural property of u_0^*.

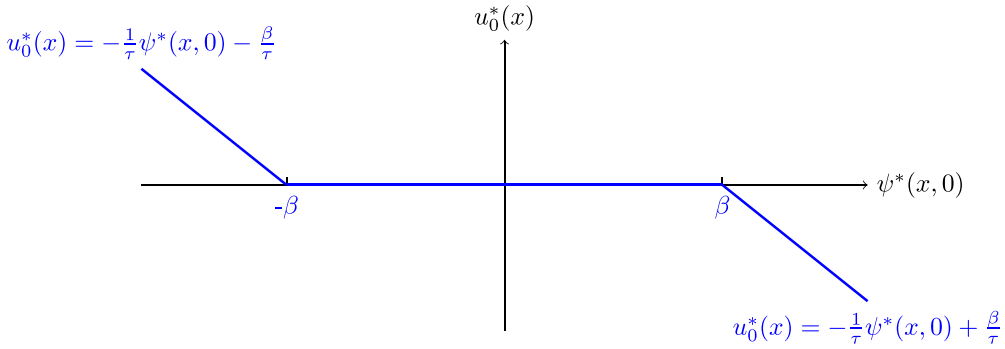


Figure 2. Relationship between $\psi^*(x,0)$ and $u_0^*(x)$.

Theorem 2.3. *Let $u_0^* \in L^2(\Omega)$ be the unique solution of problem (1.5), and ψ^* be the corresponding adjoint variable. Then, for a.e. $x \in \Omega$, we have that $|\psi^*(x,0)| \leqslant \beta$ implies $u_0^*(x) = 0$.*

When β is sufficient large, using some similar arguments as those in [45], we can prove that $u_0^* = 0$ on the whole domain Ω.

Theorem 2.4. *Let $\mathcal{L} : L^2(\Omega) \to L^2(\Omega)$ be the solution operator associated with the diffusion–advection equation (1.1), i.e. $\mathcal{L}u_0 = u(\cdot,T)$, and let $\mathcal{L}^*$ denote its adjoint. Let $\beta_0 := \|\mathcal{L}^* u_T\|_{L^\infty(\Omega)}$, where $\mathcal{L}^* u_T = \psi(\cdot,0)$ with ψ the solution of (2.2) corresponding to $\psi(\cdot,T) = u_T$. Then, if $\beta \geqslant \beta_0$, the unique solution of problem (1.5) is $u_0^* = 0$.*

Proof. We first note that, with $\mathcal{L}u_0 = u(\cdot,T)$, the objective functional $J(u_0)$ in (1.5) can be rewritten as

$$J(u_0) = \frac{1}{2}\int_\Omega |\mathcal{L}u_0 - u_T|^2 \,\mathrm{d}x + \frac{\tau}{2}\int_\Omega |u_0|^2 \,\mathrm{d}x + \beta \int_\Omega |u_0| \,\mathrm{d}x.$$

Then, it is easy to obtain that

$$\begin{aligned} J(u_0) - J(0) &= \frac{1}{2}\int_\Omega |\mathcal{L}u_0|^2 \,\mathrm{d}x - \int_\Omega \mathcal{L}u_0 u_T \,\mathrm{d}x + \frac{\tau}{2}\int_\Omega |u_0|^2 \,\mathrm{d}x + \beta \int_\Omega |u_0| \,\mathrm{d}x \\ &= \frac{1}{2}\|\mathcal{L}u_0\|^2_{L^2(\Omega)} - \int_\Omega u_0 \mathcal{L}^* u_T \,\mathrm{d}x + \frac{\tau}{2}\|u_0\|^2_{L^2(\Omega)} + \beta \|u_0\|_{L^1(\Omega)} \\ &\geqslant \frac{1}{2}\|\mathcal{L}u_0\|^2_{L^2(\Omega)} - \|u_0\|_{L^1(\Omega)}\|\mathcal{L}^* u_T\|_{L^\infty(\Omega)} + \frac{\tau}{2}\|u_0\|^2_{L^2(\Omega)} + \beta \|u_0\|_{L^1(\Omega)} \\ &= \frac{1}{2}\|\mathcal{L}u_0\|^2_{L^2(\Omega)} + (\beta - \|\mathcal{L}^* u_T\|_{L^\infty(\Omega)})\|u_0\|_{L^1(\Omega)} + \frac{\tau}{2}\|u_0\|^2_{L^2(\Omega)}. \end{aligned}$$

If $\beta \geqslant \beta_0$, we have that $J(0) \leqslant J(u_0)$ for any $u_0 \in L^2(\Omega)$, which implies that the unique solution of problem (1.5) is $u_0^* = 0$.

□

Moreover, for $\beta = 0$, it follows from (2.1) that u_0^* is not zero whenever $\psi^*(\cdot,0)$ is not zero. Typically in this case, u_0^* is nonzero almost everywhere in Ω. Therefore, we can tune β in the interval $(0,\beta_0)$ to get an optimal control u_0^* with small support.

3. A generalized primal-dual algorithm for the optimal control problem (1.5)

In this section, we propose a generalized primal-dual algorithm for the optimal control problem (1.5) and delineate its implementation details. We are inspired by a number of existing works including [11, 21, 23, 25].

3.1. A generalized primal-dual algorithmic framework

Let us define

$$f(\mathcal{L}u_0) = \frac{1}{2}\int_\Omega |\mathcal{L}u_0 - u_T|^2\,\mathrm{d}x \quad \text{and} \quad g(u_0) = \frac{\tau}{2}\int_\Omega |u_0|^2\,\mathrm{d}x + \beta\int_\Omega |u_0|\,\mathrm{d}x.$$

Then, the optimal control problem (1.5) can be reformulated as

$$\min_{u_0\in L^2(\Omega)} \Big(f(\mathcal{L}u_0) + g(u_0)\Big). \tag{3.1}$$

With an auxiliary variable $p \in L^2(\Omega)$, it follows from the Frenchel duality [1, 43] that (3.1) can be reformulated as the saddle point problem

$$\min_{u_0\in L^2(\Omega)} \max_{p\in L^2(\Omega)} \Big(g(u_0) + \int_\Omega p\mathcal{L}u_0\,\mathrm{d}x - f^*(p)\Big), \tag{3.2}$$

where $f^*(p) := \sup_{q\in L^2(\Omega)} \Big(\int_\Omega pq\,\mathrm{d}x - f(q)\Big)$ is the convex conjugate of $f(q)$ and can be specified as

$$f^*(p) = \frac{1}{2}\int_\Omega |p|^2\,\mathrm{d}x + \int_\Omega pu_T\,\mathrm{d}x.$$

Inspired by [11, 23], we propose a generalized primal-dual algorithmic framework for solving problem (3.2).

Algorithm 1. A generalized primal-dual algorithm for (3.2).

---

input: initial values $u_0^0 \in L^2(\Omega)$ and $p^0 \in L^2(\Omega)$. Choose constants $\theta \in (0,1]$, $r>0$ and $s>0$ satisfying

$$rs < \tfrac{1}{\|\mathcal{L}\mathcal{L}^*\|}, \tag{3.3}$$

and ρ and σ satisfying

$$\begin{cases} \rho = \sigma \in (0,2), \quad \text{if } \theta = 1, & (3.4a)\\ \rho \in \Big(0, 1+\theta-\sqrt{1-\theta}\Big] \text{ and } \sigma = \frac{\theta}{\rho}, \quad \text{if } \theta \in (0,1). & (3.4b) \end{cases}$$

while not converged **do**

$$\begin{cases} \widetilde{u}_0^k = \arg\min\limits_{u_0\in L^2(\Omega)} \Big(g(u_0) + \int_\Omega p^k\mathcal{L}u_0\,dx + \frac{1}{2r}\|u_0 - u_0^k\|^2_{L^2(\Omega)}\Big), & (3.5a)\\ \bar{u}_0^k = \widetilde{u}_0^k + \theta(\widetilde{u}_0^k - u_0^k), & (3.5b)\\ \widetilde{p}^k = \arg\max\limits_{p\in L^2(\Omega)} \Big(\int_\Omega p\mathcal{L}\bar{u}_0^k\,dx - f^*(p) - \frac{1}{2s}\|p - p^k\|^2_{L^2(\Omega)}\Big), & (3.5c)\\ u_0^{k+1} = u_0^k - \rho(u_0^k - \widetilde{u}_0^k), & (3.5d)\\ p^{k+1} = p^k - \sigma(p^k - \widetilde{p}^k.) & (3.5e) \end{cases}$$

end while

Algorithm 1 includes some existing works as special cases. For example, when $\rho=\sigma=0$ and $\theta=1$, it reduces to the application of the primal-dual algorithm in [11] to (3.2). That is,

$$\begin{cases} u_0^{k+1} = \underset{u_0\in L^2(\Omega)}{\arg\min}\left(g(u_0)+\int_\Omega p^k\mathcal{L}u_0\,\mathrm{d}x+\frac{1}{2r}\|u_0-u_0^k\|^2_{L^2(\Omega)}\right), & (3.6a)\\ \bar{u}_0^k = 2u_0^{k+1}-u_0^k, & (3.6b)\\ p^{k+1} = \underset{p\in L^2(\Omega)}{\arg\max}\left(\int_\Omega p\mathcal{L}\bar{u}_0^k\,\mathrm{d}x-f^*(p)-\frac{1}{2s}\|p-p^k\|^2_{L^2(\Omega)}\right). & (3.6c)\end{cases}$$

Thus, algorithm 1 generalizes the primal-dual algorithm (3.6) with more flexible choices for ρ, σ, and θ, which may result in numerical accelerations accordingly.

3.2. *Implementation of algorithm 1*

In this subsection, we discuss the implementation details of algorithm 1. To this end, it is sufficient to focus on the solutions of subproblems (3.5a) and (3.5c).

First of all, we observe that the u-subproblem (3.5a) can be reformulated as

$$\widetilde{u}_0^k = \underset{u_0\in L^2(\Omega)}{\arg\min}\left(\frac{\tau}{2}\int_\Omega |u_0|^2\,\mathrm{d}x+\beta\int_\Omega|u_0|\,\mathrm{d}x+\frac{1}{2r}\|u_0-u_0^k+r\mathcal{L}^*p^k\|^2_{L^2(\Omega)}\right), \tag{3.7}$$

where $\mathcal{L}^*p^k:=\zeta^k(\cdot,0)$ is the solution at time $t=0$ of the following backward equation:

$$\begin{cases} \partial_t\zeta^k+d\Delta\zeta^k+v\cdot\nabla\zeta^k=0, & (x,t)\in\Omega\times(0,T),\\ \zeta^k=0, & (x,t)\in\partial\Omega\times(0,T),\\ \zeta^k(\cdot,T)=p^k, & x\in\Omega. \end{cases} \tag{3.8}$$

In addition, it can be readily checked (see e.g. [29]) that problem (3.7) has the following closed-form solution

$$\widetilde{u}_0^k = \mathcal{S}_{\frac{\beta r}{\tau r+1}}\left(\frac{u_0^k-r\zeta^k(\cdot,0)}{\tau r+1}\right),$$

where, for any constant $\gamma>0$, we denoted by $\mathcal{S}_\gamma$ the Shrinkage operator defined as

$$\mathcal{S}_\gamma(a)=\begin{cases} a-\gamma, & a>\gamma,\\ 0, & |a|\leqslant\gamma,\\ a+\gamma, & a<-\gamma. \end{cases}$$

Concerning the solution of the p-subproblem (3.5c), it can be computed explicitly by taking into account that $\widetilde{p}^k$ has to satisfy

$$\nabla_p\left(\int_\Omega p\mathcal{L}\bar{u}_0^k\,\mathrm{d}x-f^*(p)-\frac{1}{2s}\|p-p^k\|^2_{L^2(\Omega)}\right)\Big|_{p=\widetilde{p}^k}=0.$$

In particular, we have

$$\widetilde{p}^k=\frac{1}{s+1}p^k+\frac{s}{s+1}\left(\mathcal{L}\bar{u}_0^k-u_T\right),$$

where $\mathcal{L}\bar{u}_0^k:=\bar{u}^k(\cdot,T)$ is the solution at time $t=T$ of the equation (1.1).

At each iteration, the main computation of algorithm 1 only requires the solutions of one forward equation (1.1) and one backward equation (3.8), and both of them can be efficiently solved by various well-developed PDE solvers. Hence, algorithm 1 is easy and computationally cheap to implement.

4. Convergence analysis of algorithm 1

In this section, we prove the strong global convergence and derive the worst-case $O(1/K)$ convergence rate measured by the iteration complexity in both the ergodic and non-ergodic senses for algorithm 1 in the context of optimal control problems. All the results can be directly extended to the primal-dual algorithm (3.6) and its relaxed version since they are special cases of algorithm 1 with specific choices of parameters. For ease of presentation, we denote by $(\cdot,\cdot)$ the canonical inner product in L^2 spaces in the following discussions.

4.1. Preliminaries

Denote $(u_0^*,p^*)^\top \in L^2(\Omega)\times L^2(\Omega)$ the saddle point of (3.2), which in particular means that u_0^* is the unique solution of (1.5). Then, the following variational inequalities (VIs) hold (see e.g. [25, 48]):

$$\varphi(u_0)-\varphi(u_0^*)+\left(u_0-u_0^*,\tau u_0^*+\mathcal{L}^*p^*\right)\geqslant 0,\quad \forall u_0\in L^2(\Omega), \tag{4.1a}$$

$$\left(p-p^*,p^*+u_T-\mathcal{L}u_0^*\right)\geqslant 0,\qquad \forall p\in L^2(\Omega), \tag{4.1b}$$

where $\varphi(u_0^*)=\beta\int_\Omega |u_0^*|dx$. We observe that the VIs (4.1*a*) and (4.1*b*) can be written in a compact form:

$$\varphi(u_0)-\varphi(u_0^*)+\left(w-w^*,F(w^*)\right)\geqslant 0,\quad \forall w\in W, \tag{4.2}$$

where

$$W=L^2(\Omega)\times L^2(\Omega),\quad w=\begin{pmatrix}u_0\\ p\end{pmatrix},\quad F(w)=\begin{pmatrix}\tau u_0+\mathcal{L}^*p\\ p-\mathcal{L}u_0+u_T\end{pmatrix}. \tag{4.3}$$

Moreover, a direct calculation shows that, for all $w_1,w_2\in W$,

$$\left(w_1-w_2,F(w_1)-F(w_2)\right)=\|p_1-p_2\|^2_{L^2(\Omega)}+\tau\|u_{0,1}-u_{0,2}\|^2_{L^2(\Omega)}, \tag{4.4}$$

which implies that F is strongly monotone.

Then, we rewrite also the iterative scheme (3.5a)–(3.5c) in a VI form. For this purpose, we first note that the optimality conditions of (3.5a) and (3.5c) are

$$\begin{aligned}&\varphi(u_0)-\varphi(\widetilde{u}_0^k)+\left(u_0-\widetilde{u}_0^k,\tau\widetilde{u}_0^k+\mathcal{L}^*p^k+\frac{1}{r}(\widetilde{u}_0^k-u_0^k)\right)\geqslant 0,\quad \forall u_0\in L^2(\Omega),\\ &\left(p-\widetilde{p}^k,\widetilde{p}^k+u_T-\mathcal{L}\bar{u}_0^k+\frac{1}{s}(\widetilde{p}^k-p^k)\right)\geqslant 0,\qquad \forall p\in L^2(\Omega),\end{aligned} \tag{4.5}$$

respectively. Taking (3.5b) into account, we obtain the following VIs:

$$\varphi(u_0)-\varphi(\widetilde{u}_0^k)+\left(u_0-\widetilde{u}_0^k,\tau\widetilde{u}_0^k+\mathcal{L}^*\widetilde{p}^k-\mathcal{L}^*(\widetilde{p}^k-p^k)+\frac{1}{r}(\widetilde{u}_0^k-u_0^k)\right)\geqslant 0,\quad \forall u_0\in L^2(\Omega), \tag{4.6a}$$

$$\left(p-\widetilde{p}^k,\widetilde{p}^k+u_T-\mathcal{L}\widetilde{u}_0^k-\theta\mathcal{L}(\widetilde{u}_0^k-u_0^k)+\frac{1}{s}(\widetilde{p}^k-p^k)\right)\geqslant 0,\quad \forall p\in L^2(\Omega). \tag{4.6b}$$

To simplify the notation, we define the following matrix-form operators

$$\mathcal{D}:=\begin{pmatrix}\rho I & 0\\ 0 & \sigma I\end{pmatrix},\quad G:=\begin{pmatrix}\frac{1}{r}I & -\mathcal{L}^*\\ -\theta\mathcal{L} & \frac{1}{s}I\end{pmatrix},\quad \mathcal{K}:=G\mathcal{D}^{-1},\quad \mathcal{N}:=G+G^*-\mathcal{D}^*\mathcal{K}\mathcal{D}. \tag{4.7}$$

With the notations in (4.3) and (4.7), the VIs (4.6*a*) and (4.6*b*), as well as the correction steps (3.5c) and (3.5d), can be respectively written in the following compact forms

$$\varphi(u_0)-\varphi(\widetilde{u}_0^k)+\left(w-\widetilde{w}^k,F(\widetilde{w}^k)+G(\widetilde{w}^k-w^k)\right)\geqslant 0,\quad \forall w\in W, \tag{4.8}$$

and

$$w^{k+1}=w^k-\mathcal{D}(w^k-\widetilde{w}^k). \tag{4.9}$$

Using some similar arguments as those in [23], we have the following result.

Lemma 4.1. *Let $\theta\in(0,1]$, r and s satisfy (3.3), ρ and σ satisfy (3.4). Then, the matrix-form operators $\mathcal{K}$ and $\mathcal{N}$ defined in (4.7) are self-adjoint and positive definite, namely,*

$$\begin{aligned}&\mathcal{K}=\mathcal{K}^*\quad \text{and}\quad (\mathcal{K}w,w)\geqslant c_1\|w\|^2_{L^2(\Omega)},\\ &\mathcal{N}=\mathcal{N}^*\quad \text{and}\quad (\mathcal{N}w,w)\geqslant c_2\|w\|^2_{L^2(\Omega)},\quad \forall w\in W, w\neq 0,\end{aligned} \tag{4.10}$$

where c_1 and c_2 are two positive constants.

In the following discussions, we denote by $\|w\|_{\mathcal{A}}:=(\mathcal{A}w,w),\forall w\in W$, the norm induced by a self-adjoint and positive definite matrix-form operator $\mathcal{A}$. Clearly, it follows from (4.10) that the norms $\|w\|_{\mathcal{K}}$ and $\|w\|_{\mathcal{N}}$, $\forall w\in W$, are well-defined.

4.2. Global convergence of algorithm 1

In this subsection, we prove the convergence of algorithm 1 under the conditions (3.3) and (3.4). First, we show that the sequence $\{w^k=(u_0^k,p^k)^\top\}_{k\geqslant 1}$ generated by algorithm 1 is strictly contractive.

Theorem 4.2. *Let $\{w^k=(u_0^k,p^k)^\top\}_{k\geqslant 1}$ be the sequence generated by algorithm 1 and $w^*=(u_0^*,p^*)^\top$ be the solution of problem (3.2). Suppose that the conditions (3.3) and (3.4) hold. Then, we have*

$$\|w^{k+1}-w^*\|^2_{\mathcal{K}}\leqslant\|w^k-w^*\|^2_{\mathcal{K}}-\|w^k-\widetilde{w}^k\|^2_{\mathcal{N}}-2\|\widetilde{p}^k-p^*\|^2_{L^2(\Omega)}-2\tau\|\widetilde{u}_0^k-u_0^*\|^2_{L^2(\Omega)}. \tag{4.11}$$

Proof. First of all, it follows from (4.7) and (4.9) that the VI (4.8) can be written as

$$\varphi(u_0)-\varphi(\widetilde{u}_0^k)+\left(w-\widetilde{w}^k,F(\widetilde{w}^k)\right)\geqslant\left(w-\widetilde{w}^k,\mathcal{K}(w^k-w^{k+1})\right),\quad \forall w\in W. \tag{4.12}$$

Then, we apply the identity

$$(a-b,\mathcal{K}(c-d))=\frac{1}{2}\left(\|a-d\|^2_{\mathcal{K}}-\|a-c\|^2_{\mathcal{K}}\right)+\frac{1}{2}\left(\|c-b\|^2_{\mathcal{K}}-\|d-b\|^2_{\mathcal{K}}\right)$$

to the right-hand side of (4.12) with

$$a=w,\quad b=\widetilde{w}^k,\quad c=w^k,\quad \text{and}\quad d=w^{k+1}.$$

We thus obtain

$$\begin{aligned}\left(w-\widetilde{w}^k,\mathcal{K}(w^k-w^{k+1})\right)=&\frac{1}{2}\left(\|w-w^{k+1}\|^2_{\mathcal{K}}-\|w-w^k\|^2_{\mathcal{K}}\right)\\ &+\frac{1}{2}\left(\|w^k-\widetilde{w}^k\|^2_{\mathcal{K}}-\|w^{k+1}-\widetilde{w}^k\|^2_{\mathcal{K}}\right).\end{aligned} \tag{4.13}$$

Considering the last two terms in (4.13) and using (4.7) and (4.9), we have

$$\begin{aligned}\|w^k-\widetilde{w}^k\|_{\mathcal{K}}^2-\|w^{k+1}-\widetilde{w}^k\|_{\mathcal{K}}^2&=\|w^k-\widetilde{w}^k\|_{\mathcal{K}}^2-\|(w^k-\widetilde{w}^k)-(w^k-w^{k+1})\|_{\mathcal{K}}^2\\&=\|w^k-\widetilde{w}^k\|_{\mathcal{K}}^2-\|(w^k-\widetilde{w}^k)-\mathcal{D}(w^k-\widetilde{w}^k)\|_{\mathcal{K}}^2\\&=2\Big(w^k-\widetilde{w}^k,\mathcal{K}\mathcal{D}(w^k-\widetilde{w}^k)\Big)-\Big(\mathcal{D}(w^k-\widetilde{w}^k),\mathcal{K}\mathcal{D}(w^k-\widetilde{w}^k)\Big)\\&=2\Big(w^k-\widetilde{w}^k,G(w^k-\widetilde{w}^k)\Big)-\Big(w^k-\widetilde{w}^k,\mathcal{D}^*\mathcal{K}\mathcal{D}(w^k-\widetilde{w}^k)\Big)\\&=\Big(w^k-\widetilde{w}^k,(G+G^*-\mathcal{D}^*\mathcal{K}\mathcal{D})(w^k-\widetilde{w}^k)\Big)\\&=\|w^k-\widetilde{w}^k\|_{\mathcal{N}}^2.\end{aligned}\tag{4.14}$$

Combining (4.12)–(4.14), we obtain that

$$\begin{aligned}\varphi(u_0)-\varphi(\widetilde{u}_0^k)+\Big(w-\widetilde{w}^k,F(\widetilde{w}^k)\Big)\geqslant\frac{1}{2}\Big(\|w-w^{k+1}\|_{\mathcal{K}}^2-\|w-w^k\|_{\mathcal{K}}^2\Big)\\+\frac{1}{2}\|w^k-\widetilde{w}^k\|_{\mathcal{N}}^2,\quad\forall w\in W.\end{aligned}\tag{4.15}$$

It follows from (4.15) that, for all $w\in W$,

$$\begin{aligned}\varphi(\widetilde{u}_0^k)-\varphi(u_0)+\Big(\widetilde{w}^k-w,F(w)\Big)+\Big(\widetilde{w}^k-w,F(\widetilde{w}^k)-F(w)\Big)\\\leqslant\frac{1}{2}\Big(\|w^k-w\|_{\mathcal{K}}^2-\|w^{k+1}-w\|_{\mathcal{K}}^2\Big)-\frac{1}{2}\|w^k-\widetilde{w}^k\|_{\mathcal{N}}^2.\end{aligned}\tag{4.16}$$

Moreover, we recall that (see (4.4))

$$\Big(\widetilde{w}^k-w,F(\widetilde{w}^k)-F(w)\Big)=\|\widetilde{p}^k-p\|_{L^2(\Omega)}^2+\tau\|\widetilde{u}_0^k-u_0\|_{L^2(\Omega)}^2.$$

Hence, setting $w=w^*$ in (4.16), and using (4.2), we finally obtain

$$\|w^{k+1}-w^*\|_{\mathcal{K}}^2\leqslant\|w^k-w^*\|_{\mathcal{K}}^2-\|w^k-\widetilde{w}^k\|_{\mathcal{N}}^2-2\|\widetilde{p}^k-p^*\|_{L^2(\Omega)}^2-2\tau\|\widetilde{u}_0^k-u_0^*\|_{L^2(\Omega)}^2.$$

□

Theorem 4.2 shows that the square of distance to a solution point can be reduced by the quantity $\|w^k-\widetilde{w}^k\|_{\mathcal{N}}^2+2\|\widetilde{p}^k-p^*\|_{L^2(\Omega)}^2+2\tau\|\widetilde{u}_0^k-u_0^*\|_{L^2(\Omega)}^2$ at the $(k+1)$th iteration. Hence, the sequence $\{w^k=(u_0^k,p^k)^\top\}_{k\geqslant1}$ generated by algorithm 1 is strictly contractive with respect to the solution w^*. This, in turn, implies the convergence of w^k to the solution point w^* of problem (3.2), as we shall see in the following theorem.

Theorem 4.3. *Let $\{w^k=(u_0^k,p^k)^\top\}_{k\geqslant1}$ be the sequence generated by algorithm 1 and $w^*=(u_0^*,p^*)^\top$ be the solution of problem (3.2). Suppose that the conditions (3.3) and (3.4) hold. Then, $\{u_0^k\}$ converges to u_0^* strongly in $L^2(\Omega)$ and p^k converges to p^* strongly in $L^2(\Omega)$.*

Proof. First of all, it follows from (4.11) that

$$\sum_{k=0}^{\infty}\Big(\|\widetilde{w}^k-w^k\|_{\mathcal{N}}^2+2\|\widetilde{p}^k-p^*\|_{L^2(\Omega)}^2+2\tau\|\widetilde{u}_0^k-u_0^*\|_{L^2(\Omega)}^2\Big)\leqslant\|w^0-w^*\|_{\mathcal{K}}^2.$$

This means that the series

$$\sum_{k=0}^{\infty}\Big(\|\widetilde{w}^k-w^k\|_{\mathcal{N}}^2+2\|\widetilde{p}^k-p^*\|_{L^2(\Omega)}^2+2\tau\|\widetilde{u}_0^k-u_0^*\|_{L^2(\Omega)}^2\Big)$$

is convergent which, in particular, implies

$$\|\widetilde{w}^k - w^k\|^2_{\mathcal{N}} \to 0, \quad \|\widetilde{u}_0^k - u_0^*\|^2_{L^2(\Omega)} \to 0, \quad \text{and} \quad \|\widetilde{p}^k - p^*\|^2_{L^2(\Omega)} \to 0, \quad \text{as } k \to \infty. \tag{4.17}$$

Thus

$$\widetilde{p}^k \to p^*, \ \widetilde{u}_0^k \to u_0^*, \quad \text{strongly in } L^2(\Omega). \tag{4.18}$$

It follows from (4.10) and (4.17) that

$$\|\widetilde{w}^k - w^k\|^2_{L^2(\Omega)} = \|\widetilde{u}_0^k - u_0^k\|^2_{L^2(\Omega)} + \|\widetilde{p}^k - p^k\|^2_{L^2(\Omega)} \to 0,$$

which, in particular, yields

$$\|\widetilde{p}^k - p^k\|^2_{L^2(\Omega)} \to 0, \quad \text{and} \quad \|\widetilde{u}_0^k - u_0^k\|^2_{L^2(\Omega)} \to 0, \quad \text{as} \quad k \to \infty.$$

This, together with (4.18), implies that

$$p^k \to p^*, \ u_0^k \to u_0^* \quad \text{strongly in } L^2(\Omega).$$

Our proof is then concluded. □

4.3. Convergence rate of algorithm 1

In this subsection, we analyze the convergence rate of algorithm 1. In particular, we establish an $O(1/K)$ worst-case convergence rate in both ergodic and non-ergodic senses.

Recall that an $O(1/K)$ worst-case convergence rate means that an iterate whose accuracy to the solution under certain criterion is of the order $O(1/K)$ can be found after K iterations of an iterative scheme. This can also be understood as the need of at most $O(1/\varepsilon)$ iterations to find an approximate solution with an accuracy of ε. Besides, we emphasize that such a convergence rate is in the worst-case nature, meaning that it provides a worst-case but universal estimate on the speed of convergence. Hence, it does not contradict with some much faster speeds which might be observed empirically for a specific application (as to be shown in section 6).

4.3.1. Convergence rate in the ergodic sense. We first establish the $O(1/K)$ worst-case convergence rate in the ergodic sense for algorithm 1 by following the work [24].

Theorem 4.4. *Let* $\{w^k = (u_0^k, p^k)^\top\}_{k\geqslant 1}$ *and* $\{\widetilde{w}^k = (\widetilde{u}_0^k, \widetilde{p}^k)^\top\}_{k\geqslant 1}$ *be the sequences generated by algorithm 1 and* $w^* = (u_0^*, p^*)^\top$ *be the solution of problem (3.2). For any* $K \in \mathbb{N}$*, define*

$$w_K = \frac{1}{K+1}\sum_{k=0}^{K} \widetilde{w}^k \quad \text{and} \quad u_{0,K} = \frac{1}{K+1}\sum_{k=0}^{K} \widetilde{u}_0^k. \tag{4.19}$$

Then, we have

$$\varphi(u_{0,K}) - \varphi(u_0^*) + \left(w_K - w^*, F(w^*)\right) \leqslant \frac{1}{2(K+1)}\|w^0 - w^*\|^2_{\mathcal{K}}. \tag{4.20}$$

Proof. Setting $w = w^*$ in (4.16), it follows from the monotonicity of F that

$$\varphi(\widetilde{u}_0^k) - \varphi(u_0^*) + \left(\widetilde{w}^k - w^*, F(w^*)\right) \leqslant \frac{1}{2}\left(\|w^k - w^*\|^2_{\mathcal{K}} - \|w^{k+1} - w^*\|^2_{\mathcal{K}}\right). \tag{4.21}$$

Summing the inequality (4.21) over $k = 0, \ldots K$, we then have

$$\frac{1}{K+1}\sum_{k=0}^{K}\left(\varphi(\widetilde{u}_0^k) - \varphi(u_0^*)\right) + \left(\frac{1}{K+1}\sum_{k=0}^{K}\widetilde{w}^k - w^*, F(w^*)\right) \leqslant \frac{1}{2(K+1)}\|w^0 - w^*\|^2_{\mathcal{K}}.$$

Then, from the convexity of φ and (4.19), we immediately obtain

$$\varphi(u_{0,K}) - \varphi(u_0^*) + \Big(w_K - w^*, F(w^*)\Big) \leqslant \frac{1}{2(K+1)}\|w^0 - w^*\|_{\mathcal{K}}^2,$$

and complete the proof. □

The above theorem shows that, after K iterations of algorithm 1, we can find an approximate solution with an $O(1/K)$ accuracy. This approximate solution is given by w_K, and it is the average of all the points $\widetilde{w}^k$ which can be computed by all the known iterates generated algorithm 1. Hence, this is an $O(1/K)$ worst-case convergence rate in the ergodic sense for algorithm 1.

As a corollary of theorem 4.4, we have the following convergence rate estimate for algorithm 1 with $\theta = 1$.

Corollary 4.5. *Let $\{w^k = (u_0^k, p^k)^\top\}_{k\geqslant 1}$ and $\{\widetilde{w}^k = (\widetilde{u}_0^k, \widetilde{p}^k)^\top\}_{k\geqslant 1}$ be the sequences generated by algorithm 1 with $\theta = 1$, and $w^* = (u_0^*, p^*)^\top$ be the solution of problem (3.2). For any $K \in \mathbb{N}$, w_K and $u_{0,K}$ are defined in (4.19). Then, for a constant $c \in (0,2)$, we have*

$$\varphi(u_{0,K}) - \varphi(u_0^*) + \Big(w_K - w^*, F(w^*)\Big) \leqslant \frac{1}{2(K+1)c}\|w^0 - w^*\|_{\widetilde{G}}^2, \tag{4.22}$$

where $\widetilde{G}$ is obtained from G in (4.7) by setting $\theta = 1$.

Proof. When $\theta = 1$, the condition (3.4) implies that ρ and σ should be chosen such that $\rho = \sigma \in (0,2)$. Moreover, if we let $c = \rho = \sigma$, the matrix-form operator $\mathcal{K}$ in (4.7) turns out to be $c^{-1}\widetilde{G}$. Then, the desired result (4.22) follows from (4.20) directly. □

The above result implies that, to implement algorithm 1 with $\theta = 1$, it is beneficial to choose c (i.e. ρ and σ) as close to 2 as possible, in order to reduce the constant on the right hand side of (4.22) and thus improve the convergence rate. Moreover, recall that the original primal-dual algorithm (3.6) is obtained by setting $\rho = \sigma = 1$ (i.e. $c = 1$) and $\theta = 1$ in algorithm 1. Hence, algorithm 1 converges faster than the original primal-dual algorithm (3.6), and this will be validated by some numerical experiments in section 6.

4.3.2. Convergence rate in the non-ergodic sense. Next, we establish the $O(1/K)$ worst-case convergence rate in a non-ergodic sense for algorithm 1 by following the work [26]. For this purpose, we first need to define a criterion to precisely measure the accuracy of an iterate.

It follows from (4.8) and $G = \mathcal{K}\mathcal{D}$ that the sequence $\{w^k\}_{k\geqslant 1}$ generated by algorithm 1 is a solution point of (4.2) if $\|\mathcal{D}(w^k - \widetilde{w}^k)\|_{\mathcal{K}} = 0$. Hence, it is reasonable to use $\|\mathcal{D}(w^k - \widetilde{w}^k)\|_{\mathcal{K}}$ or $\|\mathcal{D}(w^k - \widetilde{w}^k)\|_{\mathcal{K}}^2$ to measure the accuracy of an iterate w^k to a solution point. We have the following result.

Theorem 4.6. *Let $\{w^k = (u_0^k, p^k)^\top\}_{k\geqslant 1}$ and $\{\widetilde{w}^k = (\widetilde{u}_0^k, \widetilde{p}^k)^\top\}_{k\geqslant 1}$ be the sequences generated by algorithm 1 and $w^* = (u_0^*, p^*)^\top$ be the solution of problem (3.2). Then, for any $K \in \mathbb{N}$, we have*

$$\|\mathcal{D}(w^K - \widetilde{w}^K)\|_{\mathcal{K}}^2 \leqslant \frac{1}{c_0(K+1)}\|w^0 - w^*\|_{\mathcal{K}}^2. \tag{4.23}$$

Proof. We set $w = \widetilde{w}^{k+1}$ in (4.8) and obtain

$$\varphi(\widetilde{u}_0^{k+1}) - \varphi(\widetilde{u}_0^k) + \Big(\widetilde{w}^{k+1} - \widetilde{w}^k, F(\widetilde{w}^k) + G(\widetilde{w}^k - w^k)\Big) \geqslant 0. \tag{4.24}$$

Moreover, we notice that (4.8) also holds for $k := k+1$, which yields

$$\varphi(u_0) - \varphi(\widetilde{u}_0^{k+1}) + \Big(w - \widetilde{w}^{k+1}, F(\widetilde{w}^{k+1}) + G(\widetilde{w}^{k+1} - w^{k+1})\Big) \geqslant 0, \quad \forall w \in W.$$

Let $w = \widetilde{w}^k$ in the above inequality. Hence, we have that

$$\varphi(\widetilde{u}_0^k) - \varphi(\widetilde{u}_0^{k+1}) + \Big(\widetilde{w}^k - \widetilde{w}^{k+1}, F(\widetilde{w}^{k+1}) + G(\widetilde{w}^{k+1} - w^{k+1})\Big) \geqslant 0. \tag{4.25}$$

Adding up (4.24) and (4.25), and taking into account (4.4), we obtain that

$$\Big(\widetilde{w}^k - \widetilde{w}^{k+1}, G(\widetilde{w}^{k+1} - w^{k+1}) - G(\widetilde{w}^k - w^k)\Big) \geqslant 0.$$

Furthermore, observing that $\widetilde{w}^k - \widetilde{w}^{k+1} = \widetilde{w}^k - \widetilde{w}^{k+1} + w^k - w^k + w^{k+1} - w^{k+1}$, the above inequality yields

$$\Big(w^k - w^{k+1}, G(\widetilde{w}^{k+1} - w^{k+1}) - G(\widetilde{w}^k - w^k)\Big) \geqslant \frac{1}{2}\|(\widetilde{w}^k - w^k) - (\widetilde{w}^{k+1} - w^{k+1})\|^2_{G^*+G}, \tag{4.26}$$

where we used the fact that

$$\Big(w, Gw\Big) = \frac{1}{2}\Big(w, (G^* + G)w\Big), \ \forall w \in W.$$

It follows from (4.7) and (4.9) that (4.26) is equivalent to

$$\Big(w^k - \widetilde{w}^k, \mathcal{D}^*\mathcal{K}\mathcal{D}\big((\widetilde{w}^{k+1} - w^{k+1}) - (\widetilde{w}^k - w^k)\big)\Big) \geqslant \frac{1}{2}\|(\widetilde{w}^k - w^k) - (\widetilde{w}^{k+1} - w^{k+1})\|^2_{G^*+G}. \tag{4.27}$$

Applying the identity

$$\big(a, \mathcal{K}(a-b)\big) = \frac{1}{2}\left(\|a\|^2_{\mathcal{K}} - \|b\|^2_{\mathcal{K}} + \|a-b\|^2_{\mathcal{K}}\right)$$

to the left-hand side of (4.27) with $a = \mathcal{D}(w^k - \widetilde{w}^k)$ and $b = \mathcal{D}(w^{k+1} - \widetilde{w}^{k+1})$, we obtain

$$\begin{aligned}\Big(w^k - \widetilde{w}^k, \mathcal{D}^*\mathcal{K}\mathcal{D}\big((\widetilde{w}^{k+1} - w^{k+1}) - (\widetilde{w}^k - w^k)\big)\Big) &= \frac{1}{2}\|\mathcal{D}(w^k - \widetilde{w}^k)\|^2_{\mathcal{K}} - \frac{1}{2}\|\mathcal{D}(w^{k+1} - \widetilde{w}^{k+1})\|^2_{\mathcal{K}} \\ &\quad + \frac{1}{2}\|\mathcal{D}(w^k - \widetilde{w}^k) - \mathcal{D}(w^{k+1} - \widetilde{w}^{k+1})\|^2_{\mathcal{K}}.\end{aligned} \tag{4.28}$$

Combining (4.27) and (4.28), we thus obtain

$$\begin{aligned}&\|\mathcal{D}(w^k - \widetilde{w}^k)\|^2_{\mathcal{K}} - \|\mathcal{D}(w^{k+1} - \widetilde{w}^{k+1})\|^2_{\mathcal{K}} \\ &\quad \geqslant \|(\widetilde{w}^k - w^k) - (\widetilde{w}^{k+1} - w^{k+1})\|^2_{G^*+G} - \|\mathcal{D}(w^k - \widetilde{w}^k) - \mathcal{D}(w^{k+1} - \widetilde{w}^{k+1})\|^2_{\mathcal{K}} \\ &\quad = \|(\widetilde{w}^k - w^k) - (\widetilde{w}^{k+1} - w^{k+1})\|^2_{G^*+G-\mathcal{D}^*\mathcal{K}\mathcal{D}} \geqslant 0.\end{aligned}$$

This implies that the sequence $\|\mathcal{D}(w^k - \widetilde{w}^k)\|^2_{\mathcal{K}}$ is non-increasing, i.e.

$$\|\mathcal{D}(w^{k+1} - \widetilde{w}^{k+1})\|^2_{\mathcal{K}} \leqslant \|\mathcal{D}(w^k - \widetilde{w}^k)\|^2_{\mathcal{K}}, \quad \forall k \geqslant 0. \tag{4.29}$$

Furthermore, it follows from (4.10) and (4.11) that there exists a positive constant $c_0 > 0$ such that

$$\|w^{k+1} - w^*\|^2_{\mathcal{K}} \leqslant \|w^k - w^*\|^2_{\mathcal{K}} - c_0\|\mathcal{D}(w^k - \widetilde{w}^k)\|^2_{\mathcal{K}},$$

which implies that

$$c_0 \sum_{k=0}^{\infty} \|\mathcal{D}(w^k - \widetilde{w}^k)\|^2_{\mathcal{K}} \leqslant \|w^0 - w^*\|^2_{\mathcal{K}}. \tag{4.30}$$

Therefore, it follows from (4.29) and (4.30) that for any integer $K>0$, we have

$$(K+1)\|\mathcal{D}(w^K-\widetilde{w}^K)\|_{\mathcal{K}}^2 \leqslant \sum_{k=0}^{K}\|\mathcal{D}(w^k-\widetilde{w}^k)\|_{\mathcal{K}}^2 \leqslant \frac{1}{c_0}\|w^0-w^*\|_{\mathcal{K}}^2.$$

Our proof is then complete. □

We note that the number in the right-hand side of (4.23) is of order $O(1/K)$. Therefore, theorem 4.6 provides an $O(1/K)$ worst-case convergence rate in a non-ergodic sense for algorithm 1.

5. A structure enhancement stage for identifying the optimal locations and intensities

As discussed in the introduction, the numerical solution of the optimal control problem (1.5) is not sparse as desired. This suggests the need of introducing a second procedure to project the obtained non-sparse initial source into the set of admissible sparse solutions in the form of (1.2) and identify the locations $\widehat{x}^* := \{\widehat{x}_i^*\}_{i=1}^l$ and the intensities $\widehat{\boldsymbol{\alpha}}^* := \{\widehat{\alpha}_i^*\}_{i=1}^l$. We thus obtain a two-stage numerical approach for solving problem 1.1.

5.1. Optimal locations identification

To identify the optimal locations, we recall (see (1.2)) that the initial condition to be identified is assumed to be a finite combination of Dirac measures.

It was numerically observed in [39] that all local maxima of $|u_0^*(x)|$ fall into the optimal locations. Consequently, one can consider identifying the optimal locations $\widehat{x}^*$ by solving

$$\widehat{x}^* = \arg\max_{x\in\mathrm{supp}(u_0^*)}|u_0^*(x)|, \tag{5.1}$$

where $\mathrm{supp}(u_0^*)$ denotes the support of u_0^* and the notation 'max' refers to local maximum. Recall that by tuning the regularization parameter β, one can always obtain an optimal control u_0^* with small support. Hence, problem (5.1) is usually low-dimensional and computationally cheap to solve. Let us stress that this approach is a heuristic that has been verified by numerical observations, and it is very interesting to address its related theoretical arguments.

5.2. Optimal intensities identification

In this subsection, we explain how to find the intensities $\{\widehat{\alpha}_i^*\}_{i=1}^l$ of the initial source once we have identified their locations $\{\widehat{x}_i^*\}_{i=1}^l$ by solving (5.1). To this end, we first note that the state equation (1.1) is linear. As a consequence, for any $u_0(x)=\sum_{i=1}^l \alpha_i\delta_x(x_i)$ with $\alpha_i\in\mathbb{R}$ and $x_i\in\Omega$, the solution operator $\mathcal{L}$ verifies

$$\mathcal{L}u_0 = \sum_{i=1}^{l}\alpha_i\mathcal{L}\delta_x(x_i), \quad x_i\in\Omega.$$

Recall that we aim at identifying a sparse initial condition u_0 such that $\mathcal{L}u_0$ is as close as possible to the given target u_T. Hence, to find the optimal intensities of the initial source, it is sufficient to consider the following least squares problem:

$$\{\widehat{\alpha}_i^*\}_{i=1}^l = \arg\min_{\{\alpha_i\}_{i=1}^l\in\mathbb{R}^l}\frac{1}{2}\left\|\sum_{i=1}^{l}\alpha_i\mathcal{L}\delta_x(\widehat{x}_i^*)-u_T\right\|_{L^2(\Omega)}^2. \tag{5.2}$$

After a suitable space-time discretization, the discretized formulation of (5.2) reads

$$\widehat{\boldsymbol{\alpha}}^* = \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^l} \frac{1}{2}\|\mathbf{L}\boldsymbol{\alpha} - \boldsymbol{u_T}\|^2, \tag{5.3}$$

where $\boldsymbol{\alpha} = \{\alpha_i\}_{i=1}^l$, the vector $\boldsymbol{u_T} \in \mathbb{R}^{N_x}$ is a discretized version of u_T with N_x the number of grid points on Ω, and each column of the matrix $\mathbf{L} \in \mathbb{R}^{N_x\times l}$ contains the solution of (1.1) with $u(x,0) = \delta_x(\widehat{x}_i^*), 1 \leqslant i \leqslant l$. Note that the support of the desired sparse initial source usually consists of a few points, i.e. l is generally small. Hence, the dimension of problem (5.3) is low and it can be solved efficiently through various existing techniques. Here, we suggest to solve the corresponding normal equation

$$\mathbf{L}^\top\mathbf{L}\widehat{\boldsymbol{\alpha}}^* = \mathbf{L}^\top\boldsymbol{u_T}, \tag{5.4}$$

to find the vector of intensities $\widehat{\boldsymbol{\alpha}}^*$. Clearly, problem (5.4) is a $l\times l$ symmetric positive definite linear system and can be easily solved.

Finally, with the computed locations $\{\widehat{x}_i^*\}_{i=1}^l$ and intensities $\{\widehat{\alpha}_i^*\}_{i=1}^l$, the recovered initial source is thus given by

$$\widehat{u}_0^* = \sum_{i=1}^{l} \widehat{\alpha}_i^*\,\delta_x(\widehat{x}_i^*).$$

5.3. A two-stage numerical approach for problem 1.1

In view of the above considerations, the procedure for our initial source identification problem 1.1 needs to be complemented with the structure enhancement stage we just described. The complete methodology is given by algorithm 2.

Algorithm 2. A two-stage numerical approach for solving problem 1.1.

procedure Sparse Identification(u_T)
compute u_0^* from the optimal control problem (1.5) by algorithm 1;
compute $\psi^*(\cdot,0)$ by solving the state equation (1.1) and the adjoint equation (2.2);
find the locations by solving (5.1).
for $i = 1,2,\ldots,l$ **do**
  compute $\mathbf{L}(:,i)$ by solving (1.1) with $u(x,0) = \delta_x(\widehat{x}_i^*)$
end for
$\widehat{\boldsymbol{\alpha}}^* = (\mathbf{L}^\top\mathbf{L})\backslash\mathbf{L}^\top\boldsymbol{u_T}$
compute $\widehat{u}_0^* = \sum_{i=1}^l \widehat{\alpha}_i^*\,\delta_x(\widehat{x}_i^*)$

6. Numerical experiments

In this section, we show several test cases to validate that algorithm 2 allows identifying the sparse initial sources accurately from reachable targets or noisy observations, even for some heterogeneous materials or coupled models. For numerical discretization, we employ the backward Euler finite difference method (with step size Δt) for the time discretization and the finite element method (with mesh size Δx) described in [39, 47] for the space discretization. All our numerical results have been produced by implementing algorithm 2 in MATLAB R2016b on a Surface Pro 5 laptop with 64-bit Windows 10.0 operation system, Intel(R) Core(TM) i7-7660U CPU (2.50 GHz), and 16 GB RAM.

6.1. Generalities

We consider problem 1.1 on the domain $\Omega \times (0,T)$ with $\Omega = (0,2) \times (0,1)$ and $T = 0.1$; and we test algorithm 2 for two scenarios:

- **Scenario 1:** the given function u_T is reachable.
- **Scenario 2:** the given function u_T is observed with noise.

For each scenario, we further consider the following three cases:

(a) **Case I:** diffusivity coefficient $d = 0.05$; advection vector $v = (2,-2)^\top$ on Ω. In this case, several initial sources are to be identified in a homogeneous medium, namely, the domain Ω is constituted by materials with same diffusivity constants.

(b) **Case II:** diffusivity coefficient $d = 0.08$ on $\Omega_1 = (0,1) \times (0,1)$ and $d = 0.05$ on $\Omega_2 = (1,2) \times (0,1)$; advection vector $v = (1,2)^\top$ on Ω. Here, we consider the advection–diffusion equation modeled in a heterogeneous medium. To be concrete, the left half subdomain $\Omega_1 = (0,1) \times (0,1)$ and the right half one $\Omega_2 = (1,2) \times (0,1)$ are constituted by materials with different diffusivity constants. Consequently, the dynamics of the problem behaves differently in each of them.

(c) **Case III:** diffusivity coefficient $d = 0.05$ on Ω; advection vector $v = (0,0)^\top$ on $\Omega_1 = (0,1) \times (0,1)$ and $v = (0,-3)^\top$ on $\Omega_2 = (1,2) \times (0,1)$. This means that we identify several initial sources for coupled-models, namely, different equations are modeled on the left half ($\Omega_1 = (0,1) \times (0,1)$) and the right half ($\Omega_2 = (1,2) \times (0,1)$) of the domain Ω. More precisely, the heat equation is used on Ω_1 and the diffusion–advection equation is used on Ω_2.

The reference initial datum $\widehat{u}_0$ to be recovered for all cases is set as

$$\widehat{u}_0 = 100\delta(1.5,0.5) + 85\delta(1,0.75) + 60\delta(0.5,0.5) + 90\delta(0.75,0.25). \tag{6.1}$$

We implement the original primal-dual algorithm (3.6) and algorithm 1 to solve the optimal control problem (1.5). Both of them are repeated until the following stopping criterion is fulfilled:

$$e_k := \max\left\{\|u_0^{k+1} - u_0^k\|_{L^2(\Omega)}/\|u_0^{k+1}\|_{L^2(\Omega)}, \|p^{k+1} - p^k\|_{L^2(\Omega)}/\|p^{k+1}\|_{L^2(\Omega)}\right\} \leqslant \text{tol}$$

with $\text{tol} = 10^{-5}$ or until we reach a maximum number of iterations $k_{\max} = 1000$. Moreover, if there are no other specifications, we always use the following parameters:

- Mesh sizes: $\Delta x = 0.02$ and $\Delta t = 0.05$.
- Regularization parameters: $\beta = (\Delta x)^4, \tau = 10^{-2}$.
- The original primal-dual algorithm (3.6): $r = 6, s = 0.193 (\approx \frac{0.999}{r\|\mathcal{L}^*\mathcal{L}\|})$.
- Algorithm 1: $\theta = 1$, $r = 6, s = 0.193, \rho = \sigma = 1.9$.
- Initial values: $u_0^0 = 0, p^0 = 0$.

Moreover, we compare the numerical efficiency of our approach with the one described in [39], and show that our methodology yields significant improvements in the performance of the initial source identification procedure. For completeness, we review the approach in [39] briefly.

In [39], problem 1.1 was formulated as an optimal control problem but in the absence of an L^2-regularization in the cost functional (that is, taking $\tau = 0$ in (1.5)). To address the resulting

optimal control problem numerically, a GD approach was employed, which consists of looking for the minimizer u_0^* as the limit $k \to +\infty$ of the following iterative process:

$$u_0^{k+1} = u_0^k - \eta_k \nabla J(u_0^k).$$

Applying the above iterative scheme to the optimal control problem (1.5) yields

$$u_0^{k+1} = u_0^k - \eta_k(\psi_0^k + \tau u_0^k + \lambda_{u_0^k}), \tag{6.2}$$

where $\psi_0^k = \psi^k(\cdot, 0)$ with ψ^k the solution of (2.2). It is clear that the computational load of each GD iteration (6.2) is the same as that of algorithm 1. It is worth noting that the L^1-regularization is nonsmooth in the optimal control problem (1.5). Thus, a subgradient of the objective functional is used as the proxy of its gradient for implementation. For the convenience of comparison, we follow the notation in [39] and still call it a GD method.

In (6.2), the parameter $\eta_k > 0$ is called the step-size and plays a fundamental role in the convergence of the scheme. It is by now well-known that, if one takes η_k constant small enough and the objective functional is sufficiently regular (convex, differentiable, and with Lipschitz gradient), then (6.2) will eventually converge to the minimum (see, e.g. [40, section 2.1.5]).

Nevertheless, the choice of a constant step-size is most often not optimal: if η_k is too small, the convergence velocity of GD may drastically decrease while, if η_k is too large, one can generate overshooting phenomena and not be able to reach the minimum of J. Hence, in numerical implementations, an adaptive choice of the step-size is usually introduced (e.g. Armijo line search). In this regard, it is worth recalling that these adaptive strategies require the evaluation of the objective function value repeatedly, which in our case is numerically expensive because each one of these evaluations requires solving (1.1). For the above reasons, in our implementation of GD we always considered a constant step-size although, as we shall see, this choice contributes to making the GD methodology less efficient.

6.2. *Reachable target u_T*

We first test algorithm 2 for problem 1.1 where the target function u_T is reachable. In particular, we set the target function u_T as the solution of (1.1) at $T = 0.1$ corresponding to the initial condition $u(x;0) = \widehat{u}_0$ in (6.1).

We apply the original primal-dual algorithm (3.6), algorithm 1, and the GD method in [39] to the optimal control problem (1.5). The efficiency (in terms of the number of iterations to converge) is collected in table 1. First of all, we observe that the iteration numbers of the algorithm (3.6) and algorithm 1 are almost unchanged for different cases. We thus conclude that their convergence are robust with respect to the diffusion coefficient d and the convection coefficient v, at least for the cases we considered. We also observe from table 1 that algorithm 1 improves the numerical efficiency of the original primal-dual algorithm (3.6) by a factor about 40%, and both of them are more efficient than the GD method.

For comparison purposes, we also implement the original primal-dual algorithm (3.6), algorithm 1, and the GD method for the model introduced in [39]. The efficiency of each methodology is once again collected in table 1. It is not surprising that a significantly higher number of iterations is required because the model considered in [39] excludes the term $\frac{\tau}{2}\int_\Omega |u_0|^2 \,\mathrm{d}x$ and is much more ill-conditioned than (1.5).

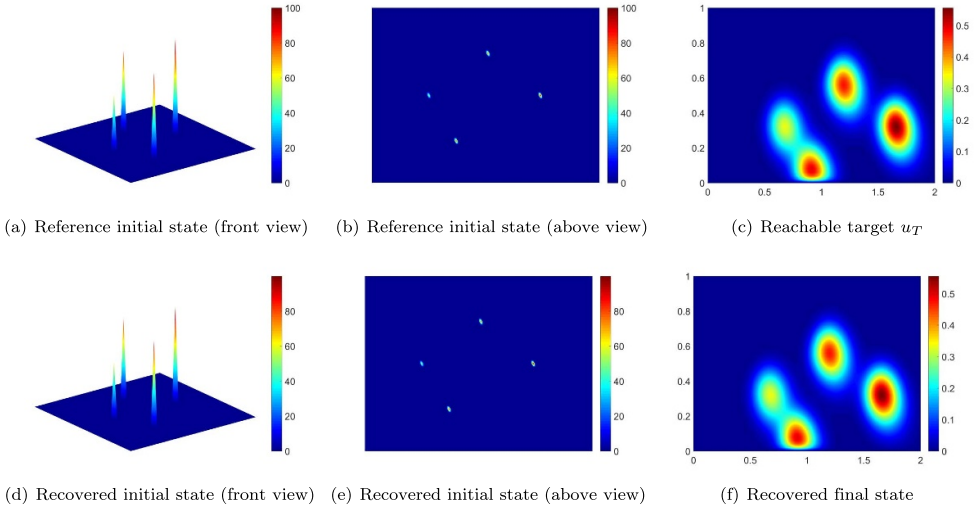
Furthermore, we recall that algorithm 1 is described on the continuous level and its convergence property is analyzed in function spaces. Hence, mesh independent property of algorithm 1 can be expected in practice, which means that the convergence behavior is independent of the fineness of the discretization. This is confirmed by our numerical results presented in table 2. The same conclusion also applies to the original primal-dual algorithm (3.6).

Table 1. Numerical comparisons of different algorithms for Cases I–III. ('Iter': the number of iterations to converge; 'Err': the relative error $\|u_0^{k+1}-u_0^k\|_{L^2(\Omega)}/\|u_0^{k+1}\|_{L^2(\Omega)}$; 'CPU': the CPU time listed in seconds) .

	Model (1.5)			Model in [39]		
	Algorithm (3.6)	Algorithm 1	GD	Algorithm (3.6)	Algorithm 1	GD
	Iter/Err/CPU	Iter/Err/CPU	Iter/Err/CPU	Iter/Err/CPU	Iter/Err/CPU	Iter/Err/CPU
Case I	$53/3\times10^{-6}/22$	$32/4\times10^{-6}/13$	$86/9\times10^{-6}/39$	$629/8\times10^{-6}/260$	$589/8\times10^{-6}/242$	$673/9\times10^{-6}/270$
Case II	$54/3\times10^{-6}/22$	$32/4\times10^{-6}/13$	$87/9\times10^{-6}/40$	$632/8\times10^{-6}/261$	$612/8\times10^{-6}/256$	$650/9\times10^{-6}/265$
Case III	$52/3\times10^{-6}/21$	$32/4\times10^{-6}/13$	$87/9\times10^{-6}/40$	$648/8\times10^{-6}/266$	$601/8\times10^{-6}/251$	$667/9\times10^{-6}/269$

Table 2. Iteration numbers with respect to different mesh sizes for Case I.

Mesh size	$\Delta t=0.1$, $\Delta x=0.05$	$\Delta t=0.05$, $\Delta x=0.02$	$\Delta t=0.025$, $\Delta x=0.0125$	$\Delta t=0.0156$, $\Delta x=0.00781$
Algorithm (3.6)	61	53	49	46
Algorithm 1	37	32	29	27



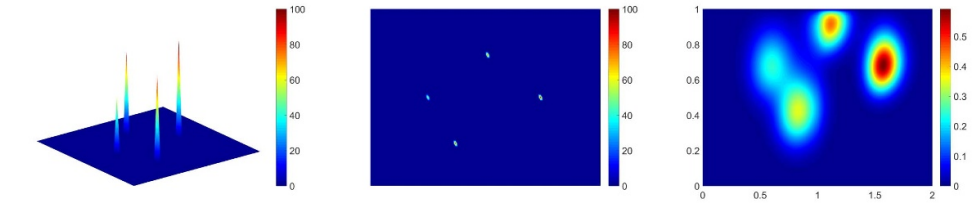
(a) Reference initial state (front view) (b) Reference initial state (above view) (c) Reachable target u_T

(d) Recovered initial state (front view) (e) Recovered initial state (above view) (f) Recovered final state

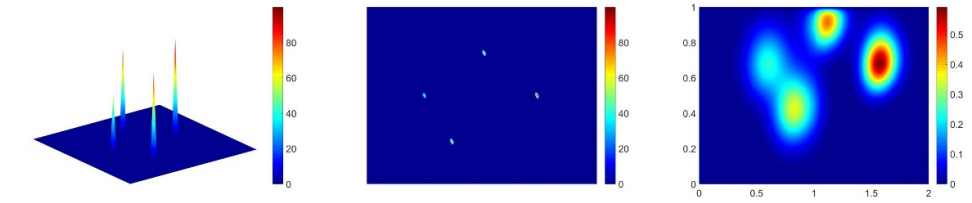
Figure 3. Sparse initial sources identification by algorithm 2 for Case I ($d=0.05, v=(2,-2)^\top$ on Ω) with a reachable target u_T at $T=0.1$.

For Case I, the recovered initial datum $\widehat{u}_0^*$ by algorithm 2 and the corresponding final state $\widehat{u}^*(\cdot,T)$ are displayed in figure 3. One can observe that both the locations and the intensities of the initial condition are recovered very accurately, which validates the effectiveness and efficiency of algorithm 2.

Similarly, the results in table 1 show that also in Case II and Case III, algorithm 1 is the most efficient one. Moreover, problem (1.5) allows for a much less expensive numerical resolution than the one in [39]. The recovered initial datum $\widehat{u}_0^*$ by algorithm 2 and the corresponding final state $\widehat{u}^*(\cdot,T)$ are displayed in figure 4(Case II) and figure 5(Case III). We observe that the locations and the intensities of the sparse initial sources are also recovered very accurately for heterogeneous materials and coupled models.

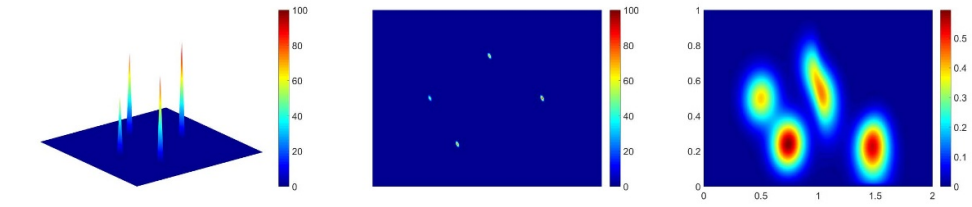


(a) Reference initial state (front view) (b) Reference initial state (above view) (c) Reachable target u_T

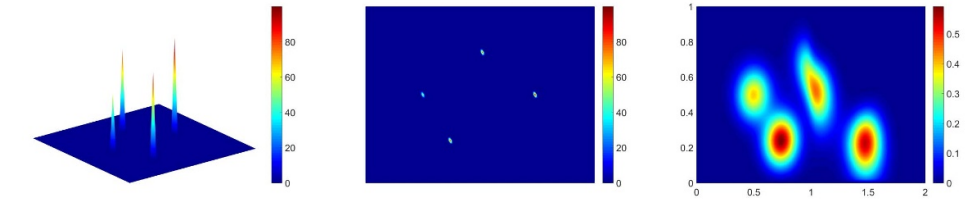


(d) Recovered initial state (front view) (e) Recovered initial state (above view) (f) Recovered final state

Figure 4. Sparse initial sources identification by algorithm 2 for Case II ($d = 0.08$ on $\Omega_1 = (0,1) \times (0,1)$ and $d = 0.05$ on $\Omega_2 = (1,2) \times (0,1)$; $v = (1,2)^\top$ on Ω) with a reachable target u_T at $T = 0.1$.

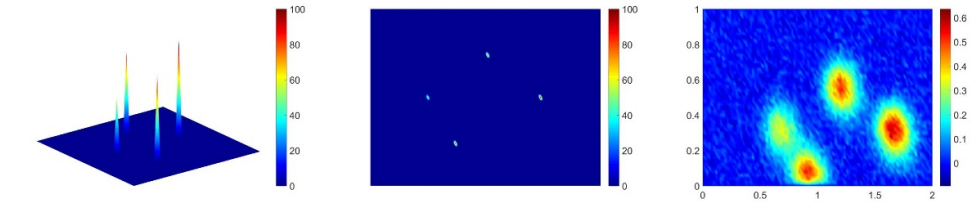


(a) Reference initial state (front view) (b) Reference initial state (above view) (c) Reachable target u_T

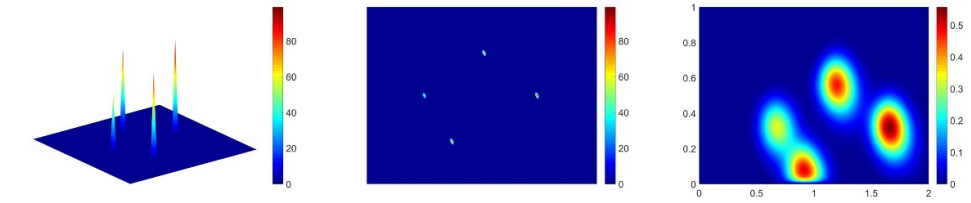


(d) Recovered initial state (front view) (e) Recovered initial state (above view) (f) Recovered final state

Figure 5. Sparse initial sources identification by algorithm 2 for Case III ($d = 0.05$ on Ω; $v = (0,0)^\top$ on $\Omega_1 = (0,1) \times (0,1)$ and $v = (0,-3)^\top$ on $\Omega_2 = (1,2) \times (0,1)$) with a reachable target u_T at $T = 0.1$.



(a) Reference initial state (front view) (b) Reference initial state (above view) (c) Noisy observation u_T



(d) Recovered initial state (front view) (e) Recovered initial state (above view) (f) Recovered final state

Figure 6. Sparse initial sources identification by algorithm 2 for Case I ($d = 0.05, v = (2,-2)^\top$ on Ω) with a noisy observation u_T at $T = 0.1$.

6.3. Noisy observation u_T

In this subsection, we aim to validate the effectiveness and efficiency of algorithm 2 for identifying sparse initial sources from some noisy observations. For convenience, we still consider the reference initial datum $\widehat{u}_0$ in (6.1), and the noisy observations at $T = 0.1$ are given by $u_T = \mathcal{L}u_0 + \delta$, where $\delta \in L^2(\Omega)$ is a noise term satisfying $\frac{\|\mathcal{L}u_0 - u_T\|_{L^2(\Omega)}}{\|\mathcal{L}u_0\|_{L^2(\Omega)}} \approx 10\%$.

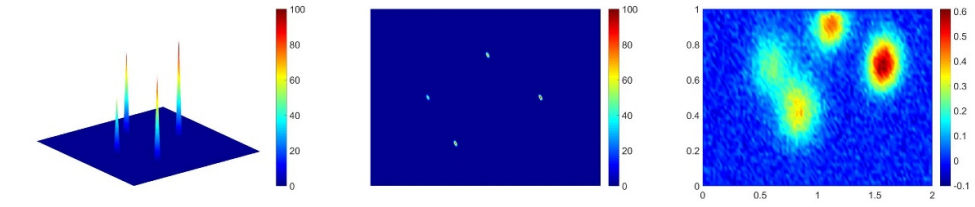
As in the previous subsections, we employ algorithm 1 to solve the optimal control problem (1.5). We observe that the iteration numbers of algorithm 1 for all test cases are almost the same as the reachable target case. Furthermore, mesh-independent property can also be observed. Hence, we can conclude that the numerical efficiency of algorithm 1 is robust with respect to noisy observations.

The initial datum $\widehat{u}_0^*$ recovered from the noisy observations u_T by algorithm 2 and the associated final state $\widehat{u}^*(\cdot, T)$ for Cases I–III are respectively presented in figures 6–8. It is easy to observe that both the locations and the intensities of the sparse initial source are recovered accurately from the noisy observations.

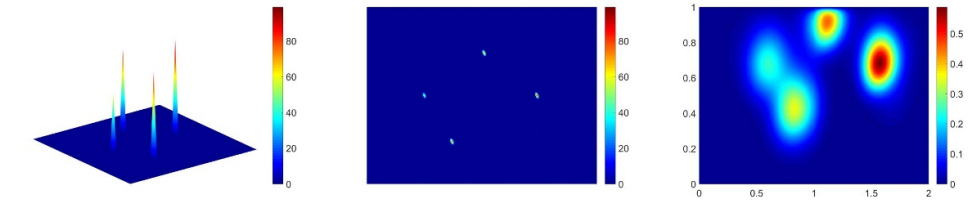
6.4. Long time horizon cases

Our simulations have shown that algorithm 2 is capable of accurately recovering the sparse initial source from a reachable target or noisy observation u_T at $T = 0.1$. On the other hand, if the final time T increases, problem 1.1 becomes strongly ill-posed and algorithm 2 cannot identify a sparse initial condition correctly, as it can be appreciated in figure 9. We observe that the recovered final state $u^*(T)$ is close to the target u_T, but the recovered initial source $\widehat{u}_0^*$ and the reference $\widehat{u}_0$ do not coincide. This validate the extreme ill-posedness of the sparse initial source identification problem in long time horizons, as it shows that a small perturbation on the final state may cause an arbitrarily large error on the initial datum.

The above issue caused by long time horizons has also been observed in some research works on backward heat conduction problems (BHCPs), see e.g. [36, 38]. Typically, a BHCP

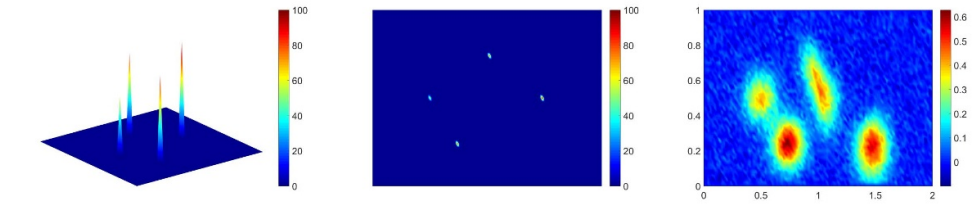


(a) Reference initial state (front view) (b) Reference initial state (above view) (c) Noisy observation u_T

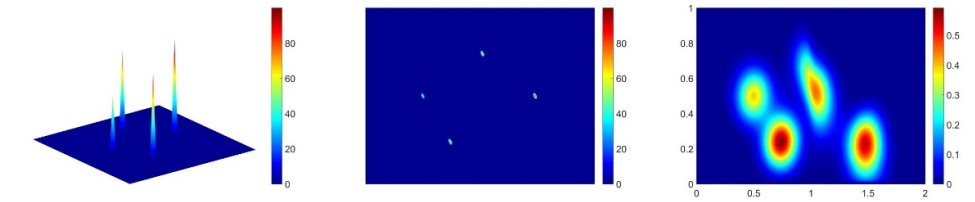


(d) Recovered initial state (front view) (e) Recovered initial state (above view) (f) Recovered final state

Figure 7. Sparse initial sources identification by algorithm 2 for Case II ($d = 0.08$ on $\Omega_1 = (0,1) \times (0,1)$ and $d = 0.05$ on $\Omega_2 = (1,2) \times (0,1)$; $v = (1,2)^\top$ on Ω) with a noisy observation u_T at $T = 0.1$.



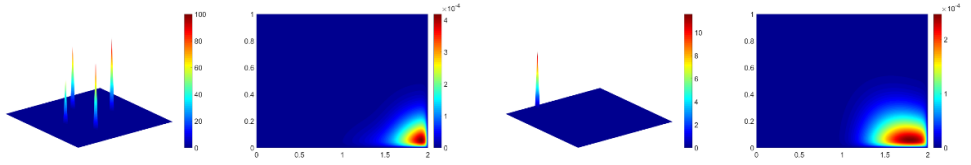
(a) Reference initial state (front view) (b) Reference initial state (above view) (c) Noisy observation u_T



(d) Recovered initial state (front view) (e) Recovered initial state (above view) (f) Recovered final state

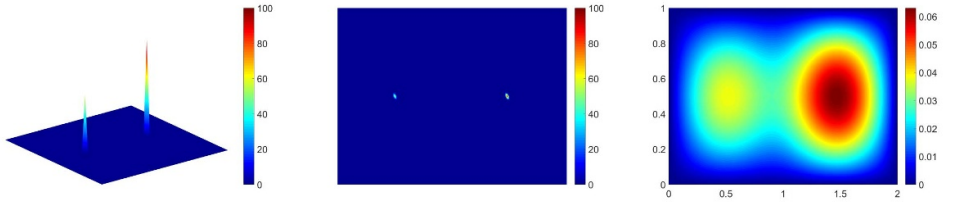
Figure 8. Sparse initial sources identification by algorithm 2 for Case III ($d = 0.05$ on Ω; $v = (0,0)^\top$ on $\Omega_1 = (0,1) \times (0,1)$ and $v = (0,-3)^\top$ on $\Omega_2 = (1,2) \times (0,1)$) with a noisy observation u_T at $T = 0.1$.

aims at estimating an initial condition of the heat equation for a given final state distribution, which is closely related to problem 1.1 but without the sparsity assumption (1.2). Based on the group preserving scheme [35], a Lie-group shooting method was proposed in [14]. When the initial condition to be estimated is smooth or its support is sufficiently large, this Lie-group shooting method can address BHCPs in long time horizons successfully. However, the Lie-group shooting method cannot be extended directly to problem 1.1 because the initial

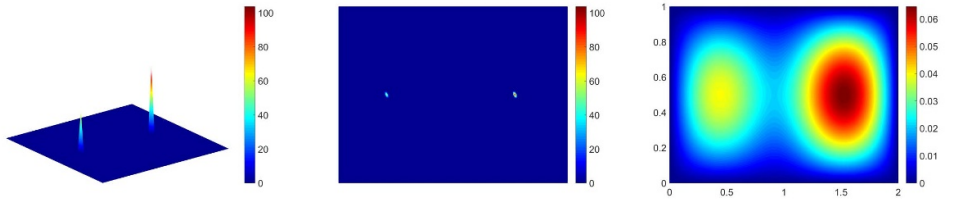


(a) Reference initial datum $\widehat{u}_0$ (b) Referene final state u_T (c) Recovered initial datum $\widehat{u}_0^*$ (d) Recovered final state $u(T)$

Figure 9. Sparse initial sources identification by algorithm 2 for Case I ($d = 0.05, v = (2,-2)^\top$ on Ω) with a reachable target u_T at $T = 1$.



(a) Reference initial state (front view) (b) Reference initial state (above view) (c) Reachable target u_T



(d) Recovered initial state (front view) (e) Recovered initial state (above view) (f) Recovered final state

Figure 10. Sparse initial sources identification by algorithm 2 for Case III ($d = 0.05$ and $v = (0,0)^\top$ on Ω) with a reachable target u_T at $T = 1$.

condition to be recovered therein is nonsmooth and has a support of Lebesgue measure zero. We also combined the group preserving scheme into algorithm 2 and obtained a new numerical approach for addressing problem 1.1. By some numerical simulations, we found that this new approach cannot improve the performance of algorithm 2 when T is large, while it is less efficient than algorithm 2 when T is small.

Additionally, we note that the admissible final time at which the sparse initial source can be identified numerically varies from case to case. It is highly related to the diffusivity parameter, the velocity field of the advection, the geometry of the domain, and the locations and intensities of the initial source to be identified, etc. To elaborate, we remove two Dirac deltas from (6.1) and consider the following reference initial datum:

$$\widehat{u}_0 = 100\delta(1.5, 0.5) + 60\delta(0.5, 0.5).$$

We set $d = 0.05$ and $v = (0,0)^\top$ on Ω, and $T = 1$. We implement algorithm 2 to this test case and the numerical results are reported in figure 10. We observe that the initial datum can be accurately recovered from the final target u_T at $T = 1$. Compared with the results in figure 9, it is easy to see that the admissible final time varies from case to case.

7. Conclusions and perspectives

In this paper, we discussed the sparse initial source identification of diffusion–advection equations. The initial source is assumed to be a finite combination of Dirac measures indicating the locations, with their weights representing the intensities; and the locations and intensities are required to be identified. We designed an algorithm capable of identifying a sparse initial condition and leading the solution of our model to match with a prescribed final target in a given time horizon T. The algorithm we proposed to solve the initial source identification problem is comprised of two stages. Firstly, we formulated an optimal control problem with a cost functional consisting of three terms:

(1) a least squares term seeking for an initial condition u_0 such that the corresponding solution, at time $t = T$, is as close as possible to the desired target;
(2) an L^1-regularization term of the initial condition u_0 to promote sparsity;
(3) an L^2-regularization term, introduced to guarantee the well-posedness of the problem while improving the conditioning of the optimal control problem;

and we introduced a generalized primal-dual algorithm to solve the optimal control problem. Secondly, an optimization problem in terms of the locations and a least squares fitting corresponding to the intensities are considered to find the optimal locations and intensities of the initial source, respectively. In our numerical simulations, by comparing with the approach in [39], the effectiveness and efficiency of the proposed two-stage numerical approach were validated by several test cases. We observed that, when the final time is not large, the initial sources from reachable targets or noisy observations were accurately identified, even for some heterogeneous materials or coupled models. When the final time becomes larger, the problem becomes increasingly ill-posed, and the sparse initial source may not be identified correctly. By some preliminary numerical tests, we found that the admissible final time, at which the sparse initial source can be identified accurately, varies from case to case. To the best of our knowledge, there is still no numerical approach in the literature that can address sparse initial source identification problems in arbitrarily long time horizons.

Nevertheless, our work left unaddressed several key aspects of initial source identification problems, which are beyond the scope of the paper and will be subject of future investigation.

(1) A natural extension of this work is to design novel and efficient algorithms allowing to address the sparse initial source identification of advection–diffusion systems in some relatively longer time horizons. We observe from figure 9(c) that the recovered location is close to the boundary of the domain, and this is mainly caused by the advection, which is the transport of a substance by bulk motion. Meanwhile, the recovered intensity is affected by the diffusion of the system. Hence, the sparse initial source identification of diffusion–advection systems can be viewed as a two-scale process: one is the inverse transport to determine the locations of the initial source, and the other is to determine the intensities of the initial source from the diffusion process. It is thus natural to consider some multiscale methods, for which some further investigation is needed.
(2) It would be interesting to address a complete analysis of the maximum admissible final time at which the sparse initial source can still be identified. This is highly related to the diffusivity parameter, the velocity field of the advection, the geometry of the domain, and the locations and intensities of the initial source to be identified. For instance, it is easy to see that a smaller diffusivity parameter or velocity field admits a larger maximum final time.

(3) In section 6, the L^2- and L^1-regularization parameters were chosen empirically. Although we observe that the proposed two-stage approach works well for different roughly selected regularization parameter, it is important to discuss the optimal combination of these two regularizations. In particular, some regularization parameter choice rules have to be deliberately designed in order to find an optimal balance between the L^2-regularization that aims to avoid ill-conditioning and the L^1-regularization that promotes sparsity.
(4) To further simplify the implementation and to improve the numerical efficiency, it would be attractive to address the sparse initial source identification problem in one shot. In this regard, one may consider modifying the optimal control problem (1.5) by taking into account the sparsity assumption (1.2) and designing some more sophisticated numerical approaches.
(5) In section 5.1, a heuristic approach was studied for identifying the locations. Its numerical efficiency inspires us to investigate its related theoretical arguments in the future.
(6) Finally, it is worth designing algorithms for the sparse initial source identification of equations that are nonlinear or modeled on more complicated geometries. For instance, recall (5.2) that the identification of the optimal intensities relies on the linearity of the diffusion–advection equation (1.1). Hence, the proposed two-stage numerical approach cannot be directly extended to the sparse initial source identification of nonlinear systems [37] and some more sophisticated techniques have to be involved in developing efficient numerical algorithms in this specific setting.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

The authors wish to acknowledge Leon Bungert (Hausdorff Center for Mathematics, University of Bonn, Bonn, Germany) for fruitful discussions on the topics of the paper. The authors are grateful to three anonymous referees for their very valuable comments which have helped them improve the paper substantially.

ORCID iDs

Umberto Biccari https://orcid.org/0000-0003-0096-5630
Xiaoming Yuan https://orcid.org/0000-0002-6900-6983

References

[1] Bauschke H H and Combettes P L 2011 *Convex Analysis and Monotone Operator Theory in Hilbert Spaces* vol 408 (Springer)
[2] Beck J V, Blaekwell B and Clair C R 1985 *Inverse Heat Conduction: Illposed Problems* (Wiley)
[3] Bregman L M 1967 The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming *USSR Comput. Math. Math. Phys.* **7** 200–17
[4] Casas E 2017 A review on sparse solutions in optimal control of partial differential equations *SeMA J.* **74** 319–44

[5] Casas E, Clason C and Kunisch K 2012 Approximation of elliptic control problems in measure spaces with sparse solutions *SIAM J. Control Optim.* **50** 1735–52
[6] Casas E, Clason C and Kunisch K 2013 Parabolic control problems in measure spaces with sparse solutions *SIAM J. Control Optim.* **51** 28–63
[7] Casas E and Kunisch K 2016 Parabolic control problems in space-time measure spaces *ESAIM: Contr. Optim. Calc.* **22** 355–70
[8] Casas E and Kunisch K 2019 Using sparse control methods to identify sources in linear diffusion-convection equations *Inverse Problems* **35** 114002
[9] Casas E, Vexler B Z and Zuazua E 2015 Sparse initial data identification for parabolic PDE and its finite element approximations *Math. Control. Relat. Fields* **5** 377–99
[10] Casas E and Zuazua E 2013 Spike controls for elliptic and parabolic PDEs *Syst. Control Lett.* **62** 311–8
[11] Chambolle A and Pock T 2011 A first-order primal-dual algorithm for convex problems with applications to imaging *J. Math. Imaging Vis.* **40** 120–45
[12] Clason C and Kunisch K 2011 A duality-based approach to elliptic control problems in non-reflexive Banach spaces *ESAIM Control Optim. Calc. Var.* **17** 243–66
[13] Duval V and Peyré G 2015 Exact support recovery for sparse spikes deconvolution *Found. Comput. Math.* **15** 1315–55
[14] Chen Y-W 2018 A modified Lie-group shooting method for multi-dimensional backward heat conduction problems under long time span *Int. J. Heat Mass Transfer* **127** 948–60
[15] El Badia A, Ha-Duong T and Hamdi A 2005 Identification of a point source in a linear advection-dispersion-reaction equation: application to a pollution source problem *Inverse Problems* **21** 1121
[16] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* vol 375 (Springer Science & Business Media)
[17] Glowinski R and Marroco A 1975 Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires *ESAIM Math. Model. Numer. Anal.* **9** 41–76
[18] Glowinski R, Song Y and Yuan X 2020 An ADMM numerical approach to linear parabolic state constrained optimal control problems *Numer. Math.* **144** 1–36
[19] Glowinski R, Song Y, Yuan X and Yue H 2022 Application of the alternating direction method of multipliers to control constrained parabolic optimal control problems and beyond *Ann. Appl. Math.* **38** 115–58
[20] Gorelick S M, Evans B and Remson I 1983 Identifying sources of groundwater pollution: an optimization approach *Water Resour. Res.* **19** 779–90
[21] Gol'shtein E G and Tret'yakov N V 1979 Modified Lagrangians in convex programming and their generalizations *Point-to-Set Maps and Mathematical Programming* (*Mathematical Programming Studies* vol 10) (Springer) pp 86–97
[22] Gurarslan G and Karahan H 2015 Solving inverse problems of groundwater-pollution-source identification using a differential evolution algorithm *Hydrogeol. J.* **23** 1109–19
[23] He B, Ma F and Yuan X 2017 An algorithmic framework of generalized primal-dual hybrid gradient methods for saddle point problems *J. Math. Imaging Vis.* **58** 279–93
[24] He B and Yuan X 2012 On the $O(1/n)$ convergence rate of Douglas-Rachford alternating direction method *SIAM J. Numer. Anal.* **50** 700–9
[25] He B and Yuan X 2012 Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective *SIAM J. Imaging Sci.* **5** 119–49
[26] He B and Yuan X 2015 On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers *Numer. Math.* **130** 567–77
[27] Hinze M, Pinnau R, Ulbrich M and Ulbrich S 2008 *Optimization with PDE Constraints* vol 23 (Springer Science & Business Media)
[28] Isakov V 2017 *Inverse Problems for Partial Differential Equations of Applied Mathematical Sciences* vol 127, 3rd edn (Springer)
[29] Justen L and Ramlau R 2009 A general framework for soft-shrinkage with applications to blind deconvolution and wavelet denoising *Appl. Comput. Harmon. Anal.* **26** 43–63
[30] Koulouri A, Heins P and Burger M 2020 Adaptive superresolution in deconvolution of sparse peaks *IEEE Trans. Signal Process.* **69** 165–78
[31] Kunisch K, Pieper K and Vexler B 2014 Measure valued directional sparsity for parabolic optimal control problems *SIAM J. Control Optim.* **52** 3078–108

[32] Leykekhman D, Vexler B and Walter D 2020 Numerical analysis of sparse initial data identification for parabolic problems *ESAIM Math. Model. Numer. Anal.* **54** 1139–80
[33] Li G, Tan Y, Cheng J and Wang X 2006 Determining magnitude of groundwater pollution sources by data compatibility analysis *Inverse Problems Sci. Eng.* **14** 287–300
[34] Tsai R, Osher S and Li Y 2014 Heat source identification based on l^1 constrained minimization *Inverse Problems Imaging* **8** 199–221
[35] Liu C S 2001 Cone of non-linear dynamical system and group preserving schemes *Int. J. Heat Mass Transfer* **36** 1047–68
[36] Liu C-S 2004 Group preserving scheme for backward heat conduction problems *Int. J. Heat Mass Transfer* **47** 2567–76
[37] Mamonov A V and Tsai Y-H R 2013 Point source identification in nonlinear advection-diffusion-reaction systems *Inverse Problems* **29** 035009
[38] Mera N S 2005 The method of fundamental solutions for the backward heat conduction problem *Inverse Problems Sci. Eng.* **13** 65–78
[39] Monge A and Zuazua E 2020 Sparse source identification of linear diffusion-advection equations by adjoint methods *Syst. Control Lett.* **145** 104801
[40] Nesterov Y 2004 *Introductory Lectures on Convex Optimization: A Basic Course* (Springer Science & Bussines Media)
[41] Ohnaka K and Uosaki K 1989 Boundary element approach for identification of point forces of distributed parameter systems *Int. J. Control* **49** 119–27
[42] Ozişik M N and Orlande H R B 2000 *Inverse Heat Transfer: Fundamentals and Applications* (Hemisphere Pub)
[43] Rockafellar T 1970 *Convex Analysis* (Princeton University Press)
[44] Schindele A and Borzì A 2017 Proximal schemes for parabolic optimal control problems with sparsity promoting cost functionals *Int. J. Control* **90** 2349–67
[45] Stadler G 2009 Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices *Comput. Optim. Appl.* **44** 159–81
[46] Ulbrich M 2011 *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces* (SIAM)
[47] Wachsmuth G and Wachsmuth D 2011 Convergence and regularization results for optimal control problems with sparsity functional *ESAIM Control Optim. Calc. Var.* **17** 858–86
[48] Zhu M and Chan T 2008 An efficient primal-dual hybrid gradient algorithm for total variation image restoration *UCLA CAM Report* 34 University of California, Los Angeles