

APPROACHES TO ITERATIVE ALGORITHMS FOR SOLVING NONLINEAR EQUATIONS WITH AN APPLICATION IN TOMOGRAPHIC ABSORPTION SPECTROSCOPY

FRANCISCO J. ARAGÓN-ARTACHO¹, WEIWEI CAI², YAIR CENSOR³, AVIV GIBALI^{4,*},
CHONGYUAN SHUI² AND DAVID TORREGROSA-BELÉN¹

¹Department of Mathematics, University of Alicante, Alicante, Spain

²Department of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China

³Department of Mathematics, University of Haifa, Haifa, Israel

⁴Applied Mathematics Department, HIT Holon Institute of Technology, Holon, Israel

This work is dedicated to the memory of Professor Hedy Attouch

Abstract. In this paper we propose an approach for solving systems of nonlinear equations without computing function derivatives. Motivated by the application area of tomographic absorption spectroscopy, which is a highly-nonlinear problem with variables coupling, we consider a situation where straightforward translation to a fixed point problem is not possible because the operators that represent the relevant systems of nonlinear equations are not self-mappings, i.e., they operate between spaces of different dimensions.

To overcome this difficulty we suggest an “alternating common fixed points algorithm” that acts alternately on the different vector variables. This approach translates the original problem to a common fixed point problem for which iterative algorithms are abundant and exhibits a viable alternative to translation to an optimization problem, which usually requires derivatives information. However, to apply any of these iterative algorithms requires to ascertain the conditions that appear in their convergence theorems.

To circumvent the need to verify conditions for convergence, we propose and motivate a derivative-free algorithm that better suits the tomographic absorption spectroscopy problem at hand and is even further improved by applying to it the superiorization approach. This is presented along with experimental results that demonstrate our approach.

Keywords. Nonlinear equations; common fixed point; cyclic sequential algorithm; tomographic absorption spectroscopy; alternating fixed points algorithm; descent pairs algorithm; superiorization.

2020 Mathematics Subject Classification. 35Q99, 47J05, 65K05, 90C30.

*Corresponding author.

E-mail address: francisco.aragon@ua.es (F. J. Aragón-Artacho), (W. Cai) cweiwei@sjtu.edu.cn, yair@math.haifa.ac.il (Y. Censor), avivgi@hit.ac.il (A. Gibali), scy1949@sjtu.edu.cn (C. Shui) and david.torregrosa@ua.es (D. Torregrosa-Belén).

Received: January 24, 2024. Revised: May 10, 2024.

©2022 Communications in Optimization Theory

1. INTRODUCTION

Solving systems of nonlinear equations has long been and still is a fundamental problem in mathematics with countless real-world applications that demand efficient methods to accomplish the task, see, e.g., the books by Ortega and Rheinboldt [36] and by Rheinboldt [37]. A Google Scholar search on this topic returns, not surprisingly, numerous entries. A central approach is based on transforming the systems of nonlinear equations to an optimization problem and using methods from that field, mainly methods for unconstrained minimization which are commonly geared toward convex functions, see, e.g., Boyd and Vandenberghe's book [4].

In practical situations it is, however, often the case that no derivatives of the functions that comprise the system of equations are known and even if derivatives exist, they are not calculable. This hinders the applicability of minimization methods and a variety of heuristics have been developed such as, for example, the simulated annealing method, consult, e.g., the information on the auto-generated ScienceDirect Webpage on this subject¹.

An alternative route is to transform the system of nonlinear equations into an operator equation such that every solution to the system is a fixed point of the operator and vice versa. This approach hinges on the premise that the associated operator is a self-mapping from a space into itself. In this paper we investigate the situation when this does not hold and the operator maps one space into a different space. For this scenario we propose an alternating fixed point approach. Specifically, we define a suitable family of fixed point operators that allows constructing an **alternating common fixed points algorithm** that can handle the problem.

However, in order to guarantee the convergence of the algorithm, some quite restrictive conditions on these operators are required. In particular, the approach seems not to be practical for the real-world application in tomographic absorption spectroscopy in which we are interested here.

Facing this kind of problems and motivated by the alternating fixed point approach described above, we take a deeper look at the properties of the equations at hand and suggest to use them in a different way. In particular, we address the case in which the equations depend on two variables, and the dependence on one of them is linear. For this problem, we propose a derivative-free algorithm which also acts alternately on each of the variables but makes use of descent directions of the summands of the least squares problem associated to the system of equations. We call it the **descent pairs algorithm** (DPA). The iterative nature of our descent pairs algorithm enables us to introduce a priori conditions into its iterative process. We do this via the **superiorization methodology**. This methodology works by taking an iterative algorithm, investigating its **perturbation resilience**, and then, using proactively such permitted perturbations,

¹<https://www.sciencedirect.com/topics/engineering/simulated-annealing-algorithm>.

it forces the perturbed algorithm to do something useful in addition to what it is originally designed to do. We present a numerical validation of the descent pairs algorithm, with and without superiorization, for a real-world application in tomographic absorption spectroscopy. In this field, problems as the one presented here have been widely studied and a great variety of algorithms have been employed to tackle them. Our experiments show that the approach proposed here yields results that compete well with those obtained by these methods under similar conditions. As a general comment we care to mention that common fixed point iterative methods and related problems are a field of vigorous research with many new directions and developments, see, e.g., [1, 25, 39] and Cegielski's book [9].

The paper is structured as follows. In Section 2, we briefly recall the well-known fixed point approach for tackling systems of nonlinear equations and present our alternating common fixed points algorithm adapted for the case in which the operator is not a self-mapping. In Section 3, the particular instance of the problem with the linear relation is tackled. We present our descent pairs algorithm and mathematically support the idea behind the algorithmic scheme. Section 4, contains a broad view of tomographic absorption spectroscopy theory and in the last subsection, we include experiments that successfully demonstrate the good performance of our descent pairs algorithm, with and without superiorization, in this field.

2. PROBLEM FORMULATION AND AN ALTERNATING COMMON FIXED POINT ALGORITHM

We are interested in solving a system of nonlinear equations as formulated next.

Problem 1. Let \mathbb{R}^M be the Euclidean M -dimensional space. Consider a family of functions $\beta_k : \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}^M$ and vectors $b^k = (b_j^k)_{j=1}^M \in \mathbb{R}^M$, for $k \in \{1, 2, \dots, W\}$.

$$\text{Find } x, y \in \mathbb{R}^M \text{ such that } \beta_k(x, y) = b^k, \quad k = 1, 2, \dots, W. \quad (2.1)$$

The motivation to look at this problem comes from a real-world application in **tomographic absorption spectroscopy** (TAS), see, e.g., Dai et al. [19], that we discuss later in Section 4. In some real-world applications, including TAS, the following condition prevails.

Condition 1. No derivatives of the functions β_k are known and even if they exist they are not calculable.

A common approach for solving systems of nonlinear equations in such a situation consists of translating the system into a **fixed point problem** (FPP), see, e.g., Combettes [18] and Combettes and Pesquet [17].

To illustrate this approach, consider a system of nonlinear equations

$$\gamma_j(z) = c_j, \quad j = 1, 2, \dots, M, \quad (2.2)$$

where $\gamma_j: \mathbb{R}^M \rightarrow \mathbb{R}$, for $j = 1, 2, \dots, M$, are given real-valued functions and $c = (c_j)_{j=1}^M \in \mathbb{R}^M$ is a given vector. Denoting by $\Gamma: \mathbb{R}^M \rightarrow \mathbb{R}^M$ the operator

$$\Gamma := \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_M \end{pmatrix}, \quad (2.3)$$

the system (2.2) becomes an **operator equation**

$$\Gamma(z) = c. \quad (2.4)$$

Since Γ is a self-mapping, it is well-known how to translate the system (2.2) into a problem of finding a point in the set of fixed points $\text{Fix}(T) := \{x \in \mathbb{R}^M \mid T(x) = x\}$ of a suitably defined operator T (see, e.g., Berinde's book [3, Chapter 8, page 179]). This is done by considering the operator $T: \mathbb{R}^M \rightarrow \mathbb{R}^M$ given by

$$T := c + (\text{Id} - \Gamma), \quad (2.5)$$

where $\text{Id}: \mathbb{R}^M \rightarrow \mathbb{R}^M$ is the identity operator, that is, $\text{Id}(x) = x$. Then, it is easy to see that for a point $x^* \in \mathbb{R}^M$

$$T(x^*) = x^* \text{ if and only if } \Gamma(x^*) = c \quad (2.6)$$

and one can solve the system (2.2) by solving the fixed point problem for the operator T .

The difficulty in applying this fixed point approach to the problem posed in (2.1) lies in the fact that $\beta_k: \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}^M$ are not self-mappings. To the best of our knowledge, the challenge of adapting the fixed point theory to this setting has not been attended before. Therefore, we create here an alternating common fixed point algorithm that applies the approach of fixed point theory alternately to each of the two vector variables x and y of the functions β_k of (2.1).

The adaptation of the fixed point theory to our problem works as follows. For any pair $x, y \in \mathbb{R}^M$ and for any $k = 1, 2, \dots, W$, define the operators

$$B_k \begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} \beta_k(x, y) \\ \beta_k(x, y) \end{pmatrix}. \quad (2.7)$$

Then $B_k : \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}^M \times \mathbb{R}^M$ are self-mappings and, for all $k = 1, 2, \dots, W$,

$$\beta_k(x, y) = b^k \iff B_k \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b^k \\ b^k \end{pmatrix}. \quad (2.8)$$

Now use the technique of Equation (2.5) and define, for all $k = 1, 2, \dots, W$, the operators

$$T_k := \begin{pmatrix} b^k \\ b^k \end{pmatrix} + (\text{Id} - B_k), \quad (2.9)$$

i.e.,

$$T_k \begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} b^k \\ b^k \end{pmatrix} + \left(\begin{pmatrix} x \\ y \end{pmatrix} - B_k \begin{pmatrix} x \\ y \end{pmatrix} \right). \quad (2.10)$$

Then, for all $k = 1, 2, \dots, W$,

$$T_k \begin{pmatrix} x^* \\ y^* \end{pmatrix} = \begin{pmatrix} x^* \\ y^* \end{pmatrix} \iff \beta_k(x^*, y^*) = b^k, \quad (2.11)$$

and finding a solution to Problem 1 is equivalent to solving the common fixed point problem (CFPP)

$$\text{Find } \begin{pmatrix} x^* \\ y^* \end{pmatrix} \in \bigcap_{k=1}^W \text{Fix}(T_k). \quad (2.12)$$

There is an extensive literature of algorithms devoted to solving the CFPP in (2.12), see, e.g., Zaslavski's book [40]. As an example, the algorithm presented here in Algorithm 1 is the cyclic version of the **almost cyclic sequential algorithm** (ACSA) for the common fixed point problem, in Censor and Segal [14, Algorithm 5], which is, in turn, a special case of the algorithm in Combettes [16, Algorithm 6.1].

For the operators defined in (2.10), Algorithm 1 leads to the proposed **alternating common fixed points algorithm**. It is the iterative process, that starts with an arbitrary

$\begin{pmatrix} x^0 \\ y^0 \end{pmatrix} \in \mathbb{R}^M \times \mathbb{R}^M$, and then, for all $\ell \geq 0$, updates according to

$$\begin{pmatrix} x^{\ell+1} \\ y^{\ell+1} \end{pmatrix} = \begin{pmatrix} x^\ell \\ y^\ell \end{pmatrix} + \lambda_\ell \left(\begin{pmatrix} b^{i(\ell)} \\ b^{i(\ell)} \end{pmatrix} - \begin{pmatrix} \beta_{i(\ell)}(x^\ell, y^\ell) \\ \beta_{i(\ell)}(x^\ell, y^\ell) \end{pmatrix} \right). \quad (2.13)$$

However, the convergence theorem for ACSA, see, e.g., [14, Theorem 6], requires additional assumptions on the involved operators. They should be “directed operators”² such that, for all k , $T_k - \text{Id}$ should be closed at 0. An (uninteresting) example verifying these properties are the projection operators onto the diagonal subspace.

Algorithm 1: The Cyclic Sequential Algorithm (CSA) for the common fixed point problem

- 1 **Initialization:** Let $\{T_k : \mathbb{R}^M \rightarrow \mathbb{R}^M\}_{k=1}^W$ be a finite family of operators. Set $\{\lambda_\ell\}_{\ell=0}^\infty$ a sequence in $[0, 2]$. Choose $x^0 \in \mathbb{R}^M$;
 - 2 set $\ell = 0$;
 - 3 **repeat**
 - 4 set $i(\ell) = \ell \bmod W + 1$;
 - 5 set $x^{\ell+1} = x^\ell + \lambda_\ell (T_{i(\ell)}(x^\ell) - x^\ell)$;
-

The difficulty of having operators whose block coordinates are identically defined, as is the case for the operators in (2.10), and verifying the assumptions of the convergence theorem for ACSA, lead us to search for an alternative approach to handle the problem. In the next section we develop a different algorithmic scheme which is inspired by the iterative process (2.13).

3. THE DESCENT PAIRS ALGORITHM

In this section we propose and motivate a derivative-free algorithm for tackling Problem 1 in the case when the operators fulfill the following two assumptions. We denote the set of vectors whose components are all different from zero by

$$\mathbb{R}_{\neq 0}^M := \{x \in \mathbb{R}^M \mid x_j \neq 0 \text{ for all } j = 1, 2, \dots, M\}. \quad (3.1)$$

Assumption 1. For every $k \in \{1, 2, \dots, W\}$, the operators β_k depend linearly on the second variable and nonlinearly on the first variable, such that they can be expressed, in matrix form, as

$$\beta_k(x, y) = \text{diag} \left(\tilde{\beta}_k(x) \right) y, \quad k = 1, 2, \dots, W, \quad (3.2)$$

²Directed operators are nowadays called “cutters”, see [9, Subsection 2.1.3].

where $\tilde{\beta}_k : \mathbb{R}^M \rightarrow \mathbb{R}_{\neq 0}^M$ and $\text{diag}(u)$ denotes the diagonal matrix with the vector u along its diagonal.

For our second assumption, we need to introduce the following definitions.

Definition 3.1. (i) **Descent direction** (see, e.g., [2, Definition 5.1]). Given a function $g : \mathbb{R}^M \rightarrow \mathbb{R}$ which is differentiable at some vector $x \in \mathbb{R}^M$, then a vector $v \in \mathbb{R}^M$ is called a **descent direction for g at the point x** if

$$g'(x; v) = \lim_{t \rightarrow 0} \frac{g(x + tv) - g(x)}{t} = \nabla g(x)^T v < 0. \quad (3.3)$$

(ii) **Descent pair.** Given a mapping $f : \mathbb{R}^M \rightarrow \mathbb{R}^N$ and a vector $u \in \mathbb{R}^N$ we say that (f, u) is a **descent pair** if, for every $x \in \mathbb{R}^M$, the vector

$$v = v(x) := u - f(x) \quad (3.4)$$

is a descent direction at the point x for the function $g : \mathbb{R}^M \rightarrow \mathbb{R}$ defined as

$$g(\cdot) := \frac{1}{2} \|u - f(\cdot)\|^2. \quad (3.5)$$

Assumption 2. There exists an index $t \in \{1, 2, \dots, W\}$ such that $b^t \in \mathbb{R}_{\neq 0}^M$ and such that for all $k = 1, 2, \dots, W$, the pairs (f_k, u^k) , with

$$f_k(\cdot) := \text{diag}(\tilde{\beta}_t(\cdot)) \tilde{\beta}_k(\cdot) \text{ and } u^k := \text{diag}(b^t)^{-1} b^k, \quad (3.6)$$

and where the “ -1 ” power represents matrix inversion, are descent pairs. Moreover, for each $x \in \mathbb{R}^M$, the pairs (h_k, b^k) , where

$$h_k(\cdot) := \beta_k(x, \cdot), \quad (3.7)$$

are descent pairs for all $k = 1, 2, \dots, W$.

In the tomographic absorption spectroscopy problem, modeled as Problem 1 and studied in subsection 4.3 below, Assumptions 1 and 2 are fulfilled.

The next lemma shows that, although Problem 1 is a system which depends on both variables x and y , the Assumption 1 on the operators β_k implies that the variable vector y does not interfere with the suitability of the variable vector x for solving the system.

Lemma 3.2. Consider a family of operators $\beta_k : \mathbb{R}^M \times \mathbb{R}_{\neq 0}^M \rightarrow \mathbb{R}_{\neq 0}^M$, for $k \in \{1, 2, \dots, W\}$, for which Assumption 1 holds. Then, a point $x^* \in \mathbb{R}^M$ belongs to a solution pair (x^*, y^*)

of the system of equations given by (2.1) if and only if x^* is a solution of the system

$$\text{diag}\left(\tilde{\beta}_t(x)\right)^{-1}\tilde{\beta}_k(x)=\text{diag}(b^t)^{-1}b^k, \text{ for all } k=1,2,\dots,W, k\neq t, \quad (3.8)$$

with $t\in\{1,2,\dots,W\}$ such that $b^t\in\mathbb{R}_{\neq 0}^M$.

Proof. Let (x^*,y^*) be a solution of the system given by (2.1). Then, by Assumption 1, for all $k,t\in\{1,2,\dots,W\}$, and using elementary rules for matrix inversion and diagonal matrices, we have

$$\begin{aligned} \beta_k(x^*,y^*)=b^k &\Leftrightarrow \text{diag}\left(\tilde{\beta}_k(x^*)\right)y^*=b^k \\ &\Leftrightarrow \text{diag}\left(\text{diag}\left(\tilde{\beta}_t(x^*)\right)y^*\right)^{-1}\text{diag}\left(\tilde{\beta}_k(x^*)\right)y^*=\text{diag}(b^t)^{-1}b^k \\ &\Leftrightarrow \text{diag}\left(\tilde{\beta}_t(x^*)\right)^{-1}\text{diag}(y^*)^{-1}\text{diag}\left(\tilde{\beta}_k(x^*)\right)y^*=\text{diag}(b^t)^{-1}b^k \\ &\Leftrightarrow \text{diag}\left(\tilde{\beta}_t(x^*)\right)^{-1}\tilde{\beta}_k(x^*)=\text{diag}(b^t)^{-1}b^k. \end{aligned} \quad (3.9)$$

Thus, x^* is a solution for the single vector variable system (3.8). \square

In Algorithm 2 below, which we call the **descent pairs algorithm**, we present a method for tackling a system of the form (2.1) that obeys Assumptions 1 and 2. The motivation of the algorithm is as follows. In order to obtain a solution pair (x^*,y^*) to the system (2.1) we generate two separate iterative sequences. The first sequence is denoted by $\{x^\ell\}_{\ell=0}^\infty$ and employs Lemma 3.2 to find x^* as a solution to the system (3.8). To achieve this, we consider the least squares problem associated with (3.8), which is

$$\underset{x\in\mathbb{R}^M}{\text{argmin}} \sum_{k=1,k\neq t}^W g_k(x)=\sum_{k=1,k\neq t}^W \frac{1}{2}\left\|\text{diag}\left(\tilde{\beta}_t(x)\right)^{-1}\tilde{\beta}_k(x)-\text{diag}(b^t)^{-1}b^k\right\|^2. \quad (3.10)$$

Lines 4 to 8 of the algorithm generate $x^{\ell+1}$ from x^ℓ by sequentially updating x^ℓ with constant step-size line searches for each of the summands in (3.10), in the descent direction given by the descent pair provided by Assumption 2.

The purpose of the second sequence $\{y^\ell\}_{\ell=0}^\infty$ is to find the point y^* . Lines 9 to 13 show how to obtain $y^{\ell+1}$ from y^ℓ and $x^{\ell+1}$. Indeed, y^ℓ is sequentially updated by performing a constant step-size line search for each of the functions

$$h_k(y):=\frac{1}{2}\|b^k-\beta_k(x^{\ell+1},y)\|^2, \quad k=1,2,\dots,W, \quad (3.11)$$

in the descent direction determined by the descent pair given by Assumption 2.

Algorithm 2: The Descent Pairs Algorithm (DPA).

```

1 Initialization. Choose  $x^0, y^0 \in \mathbb{R}^M$ . Set an index  $t \in \{1, 2, \dots, W\}$  in compliance
   with Assumption 2 and pick real fixed relaxation parameters  $\lambda_x > 0$  and  $\lambda_y > 0$ ;
2 Set  $\ell = 0$ ;
3 repeat
4   Set  $x^{\ell,0} = x^\ell$ ;
5   for  $q = 1, 2, \dots, W$  do
6      $x^{\ell,q} = x^{\ell,q-1} + \lambda_x \left( \text{diag}(b^t)^{-1} b^q - \text{diag} \left( \tilde{\beta}_t(x^{\ell,q-1}) \right)^{-1} \tilde{\beta}_q(x^{\ell,q-1}) \right)$ ;
7   end
8   Set  $x^{\ell+1} = x^{\ell,W}$ ;
9   Set  $y^{\ell,0} = y^\ell$ ;
10  for  $q = 1, 2, \dots, W$  do
11     $y^{\ell,q} = y^{\ell,q-1} + \lambda_y (b^q - \beta_q(x^{\ell+1}, y^{\ell,q-1}))$ ;
12  end
13  Set  $y^{\ell+1} = y^{\ell,W}$ ;

```

In Section 4, we illustrate the good performance of the proposed Algorithm 2 in a demonstrative numerical experiment arising from a real-world problem in tomographic absorption spectroscopy, in which our algorithm shows a competitive potential vis-à-vis with state-of-the-art methods in the field.

4. AN EXPERIMENTAL DEMONSTRATION IN TOMOGRAPHIC ABSORPTION SPECTROSCOPY

4.1. Introduction to tomographic absorption spectroscopy. Absorption spectroscopy is a popular technique for gas sensing which can simultaneously retrieve thermophysical properties such as temperature, species concentration and pressure [7]. When a laser beam penetrates a **region of interest** (ROI) filled with gaseous medium, its intensity is attenuated due to the absorption of the gas molecules along the **line-of-sight** (LOS).

The aim of absorption spectroscopy is to obtain information of the gaseous medium by measuring spectrum-specified absorbance. The Beer-Lambert law, which relates the attenuation of light to the properties of the material through which the light is traveling, provides the relationship between the gas properties (i.e., pressure, temperature and concentration) and the absorbance, see [33],

$$b(\mathbf{v}) := \ln(I_0(\mathbf{v})/I_t(\mathbf{v})) = \int PS(\mathbf{v}, x) y dL. \quad (4.1)$$

The absorbance $b(\nu)$ is the logarithmic ratio of the incident $I_0(\nu)$ and transmitted $I_t(\nu)$ intensities for the absorption line at wavelength ν . In (4.1), P is the pressure, which is supposed to be a known constant, and S is the line strength function, whose value depends on the wavelength ν and the temperature x . For each wavelength of interest, the line function S has an approximately negative exponent relation with the reciprocal of the temperature, i.e. $1/x$. The concentration of the absorbing species is y and L is the length of the LOS.

In practical applications, the gas properties are usually non-uniform along the LOS, therefore, tomography is needed to enable the spatial resolution of absorption spectroscopy. In order to represent a mathematically tractable model, full discretization is done, meaning that the light spectrum is discretized into a finite number of wavelengths, the ROI is discretized into a finite number of pixels/voxels and the external light sources are discretized into a finite number of individual beams. This modeling of tomographic absorption spectroscopy (TAS) is illustrated in Figure 1.

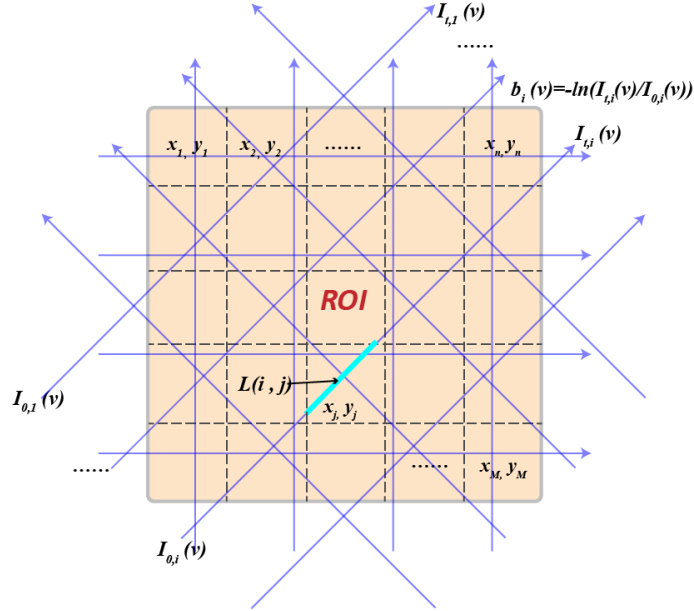


FIGURE 1. Illustration of the fully-discretized model for tomographic absorption spectroscopy.

The discretized model for TAS is obtained as follows. First, the absorbance is measured along a finite number of beams, indexed by $i = 1, 2, \dots, N$, covering the ROI. The light spectrum is discretized into a finite number of wavelengths, indexed by $k = 1, 2, \dots, W$. When N beams are employed and W absorption features are probed in the measurements, the relationship between the temperature x , the concentration of the absorbing species y and the measured absorbance b can be expressed as

$$b_i^k = \int_{L_i} \alpha_{v_k}(x, y) dL, \quad k = 1, 2, \dots, W, \quad i = 1, 2, \dots, N, \quad (4.2)$$

where b_i^k is the absorbance along the i -th beam with length L_i at the k -th wavelength v_k , and α_{v_k} is the absorption coefficient, related to the wavelength v_k and the local values of x and y .

Finally, we assume that the ROI is discretized and consists of a two-dimensional square, meshed into $M = n \times n$ square pixels, indexed by $j = 1, 2, \dots, M$. Thus, denoting by $L_{i,j}$ the length of intersection of the i -th light beam within the j -th pixel, equation (4.2) is reinterpreted in its fully-discretized form as

$$b_i^k = \sum_{j=1}^M \alpha_{v_k}(x_j, y_j) \cdot L_{i,j} = \sum_{j=1}^M \alpha_k(x_j, y_j) \cdot L_{i,j}, \quad (4.3)$$

for $k = 1, 2, \dots, W$, $i = 1, 2, \dots, N$, $j = 1, 2, \dots, M$. The $\alpha_k : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ are nonlinear operators which measure the absorption coefficient at the k -th wavelength. Note that after the discretization, x and y are redefined as vectors in \mathbb{R}^M .

Repeating equation (4.3) for all the beams at the k -th wavelength, a set of equations is formulated in matrix form as

$$b^k = \begin{pmatrix} b_1^k \\ b_2^k \\ \vdots \\ b_i^k \\ \vdots \\ b_N^k \end{pmatrix} = \begin{pmatrix} L_{1,1} & \cdots & L_{1,j} & \cdots & L_{1,M} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ L_{i,1} & \cdots & L_{i,j} & \cdots & L_{i,M} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ L_{N,1} & \cdots & L_{N,j} & \cdots & L_{N,M} \end{pmatrix} \begin{pmatrix} \alpha_k(x_1, y_1) \\ \alpha_k(x_2, y_2) \\ \vdots \\ \alpha_k(x_j, y_j) \\ \vdots \\ \alpha_k(x_M, y_M) \end{pmatrix} = L\alpha_k(x, y), \quad (4.4)$$

for all $k = 1, 2, \dots, W$, where L denotes the matrix $L = (L_{i,j})_{i=1, j=1}^{N, M}$. Thus, the fully-discretized modeling of the tomographic problem in TAS results in solving the nonlinear system of equations

$$b^k = L\alpha_k(x, y), \quad k = 1, 2, \dots, W, \quad (4.5)$$

where the vectors b^k are known from measurements for all wavelengths and the matrix L is calculated according to the beam arrangement. In the past decades, numerous studies have made progress in this problem. Next, we give a brief glance of it.

4.2. Techniques for solving TAS problems. Both nonlinear or linear approaches have been applied in the literature to solve system (4.5). Depending on the approach for solving the tomographic problem, TAS can be divided into **nonlinear TAS** and **linear TAS**. In nonlinear TAS, all the equation systems are considered together and the inversion is recast into a one-step optimization problem as [30]:

$$\operatorname{argmin}_{x,y \in \mathbb{R}^M} \frac{1}{2} \sum_{k=1}^W \|b^k - L\alpha_k(x,y)\|^2. \quad (4.6)$$

The x and y distributions are directly solved from (4.6). This optimization problem is usually solved by a global heuristic optimization algorithm, such as **simulated annealing** (SA) [6, 30].

The SA algorithm was first proposed by Kirkpatrick et al. in 1983 [31] and has been widely employed in large-scale optimization problems [32, 8]. It is a heuristic for finding the global minimum inspired by annealing in metallurgy. A prominent advantage of SA is its insensitivity to the initial guesses [34], while the major drawback is its high computational cost. Besides, a priori information can be taken into consideration [30], which is another advantage of SA.

On the other hand, in linear TAS the problem is divided into two stages. In the first stage, for every $k \in \{1, 2, \dots, W\}$, equation (4.4) is solved for each individual wavelength as a linear equation system whose variables are the local absorption coefficients α_k . Classical algorithms, including **algebraic reconstruction technique** (ART) [26], **maximum likelihood expectation maximization** (MLEM) [22] and Tikhonov reconstruction [20], have been extensively adopted for this stage.

For the second stage, absorption coefficients $a^k = (a_j^k)_{j=1}^M \in \mathbb{R}^M$, at each pixel j and for each wavelength k , are supposed to have been recovered in the first stage, i.e.,

$$a^k = \begin{pmatrix} a_1^k \\ a_2^k \\ \vdots \\ a_j^k \\ \vdots \\ a_M^k \end{pmatrix} = \begin{pmatrix} \alpha_k(x_1, y_1) \\ \alpha_k(x_2, y_2) \\ \vdots \\ \alpha_k(x_j, y_j) \\ \vdots \\ \alpha_k(x_M, y_M) \end{pmatrix}, \quad k = 1, 2, \dots, W. \quad (4.7)$$

Now the properties x and y need to be solved according to the nonlinear relationship between them and the absorption coefficients. While, for every $k \in \{1, 2, \dots, W\}$, the absorption coefficients a^k have been recovered in the first stage, the $\alpha_k : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ become here operators. Commonly, in the literature, gas properties within each pixel were calculated from the a^k s using a **nonlinear fitting** (NF) method [27]. This nonlinear fitting process employs a trust-region-reflective algorithm [15] to find the least-squares

of the discrepancy between a^k and $\alpha_k(x, y)$. The iterative process of this algorithm is briefly described in Algorithm 3, see [35]. In the algorithm, g and H refer to the first and second order derivatives of the function to be solved, respectively, and $s = (s_1, s_2)$ is the step of the iteration. It should be noted that an analytical expression of the nonlinear tomographic absorption spectroscopy problem is not derivable. Here, the g and H are results of an approximation of the nonlinear tomography formulation.

Algorithm 3: The Nonlinear Fitting (NF) method applied in multi-spectral TAS.

```

1 Initialization: Set  $\delta \in \mathbb{R}$ . Choose  $x_j^0 \geq 0$  and  $y_j^0 \geq 0$ ;
2 set  $\ell = 0$ ;
3 repeat
4   set  $H = \nabla^2(\|a_j^k - \alpha_k(\cdot, \cdot)\|^2)(x_j^\ell, y_j^\ell)$ ,  $g = \nabla(\|a_j^k - \alpha_k(\cdot, \cdot)\|^2)(x_j^\ell, y_j^\ell)$ ;
5   set  $(s_1, s_2) = \operatorname{argmin}_{\|s\| \leq \delta} (\frac{1}{2}s^T H s + g^T s)$ ;
6   if  $\|a_j^k - \alpha_k(x_j^\ell + s_1, y_j^\ell + s_2)\|^2 < \|a_j^k - \alpha_k(x_j^\ell, y_j^\ell)\|^2$  then
7     set  $x_j^{\ell+1} = x_j^\ell + s_1$ ,  $y_j^{\ell+1} = y_j^\ell + s_2$ ;
8     set  $\ell = \ell + 1$ ;
9     adjust the trust region size  $\delta$ ;
10  else
11    reduce the trust region size  $\delta$ ;
12  end

```

A drawback of this method is that it still has a high computational cost. Besides, since the algorithm obtains x_j and y_j pixel by pixel, a priori information cannot be employed in this algorithm for the process of obtaining temperature and concentration results from absorption coefficients. That means, in traditional approaches for TAS, a priori information is solely employed in the process to solve the absorption coefficients, which is sometimes defective. For example, if the temperature and the concentration satisfy different a priori information, the method to add a priori information on absorption coefficients cannot work. In addition, the calculation of temperature and mole fraction from absorption coefficient can introduce extra errors due to complicated error propagation. In the following sections, we motivate the implementation of our proposed Algorithm 2 for tackling problems of the form given by (4.7), and present a demonstrative example in which it outperforms the NF approach.

4.3. Implementation of the descent pairs algorithm. Let \mathbb{R}_{++}^M denote the positive orthant of \mathbb{R}^M . For every $k \in \{1, 2, \dots, W\}$, denote by $\beta_k : \mathbb{R}_{++}^M \times \mathbb{R}_{++}^M \rightarrow \mathbb{R}_{++}^M$ the operator defined component-wise as

$$(\beta_k(x, y))_j := \alpha_k(x_j, y_j), \quad j = 1, 2, \dots, M, \quad (4.8)$$

where $\alpha_k : \mathbb{R}_{++} \times \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$ are the operators in (4.7). This yields the system of equations

$$\beta_k(x, y) = a^k, \quad k = 1, 2, \dots, W. \quad (4.9)$$

It is a known property of the absorption coefficients that the operators α_k fulfill Assumption 1 [7]. Thus, there exist operators $\tilde{\beta}_k : \mathbb{R}_{++}^M \rightarrow \mathbb{R}_{++}^M$ such that

$$\beta_k(x, y) = \text{diag} \left(\tilde{\beta}_k(x) \right) y, \quad k = 1, 2, \dots, W. \quad (4.10)$$

In order to validate the implementation of the descent pairs algorithm (Algorithm 2) to this problem, we employ a property of the TAS system of equations (4.9) which is known from the physics of the problem and is presented in the next remark.

Remark 2. For a fixed index $t \in \{1, 2, \dots, W\}$, define for each $k \in \{1, 2, \dots, W\} \setminus \{t\}$ the functions

$$f_k(x) := \text{diag} \left(\tilde{\beta}_t(x) \right)^{-1} \tilde{\beta}_k(x). \quad (4.11)$$

Then, it is known from the physics of the problem that the functions f_k are given component-wise by

$$(f_k(x))_j := \frac{S_k}{S_t} \exp \left(- (E_k - E_t) \left(\frac{1}{x_j} - \frac{1}{T_0} \right) \right), \quad j = 1, 2, \dots, M, \quad (4.12)$$

where $S_k, S_j > 0$ and $E_k, E_t \in \mathbb{R}$ are constants which depend on the wavelength, and T_0 is a positive known constant.

The following lemmata assure that Assumption 2 holds for the system of equations (4.9).

Lemma 4.1. Let t be an index such that $E_t = \min \{E_1, E_2, \dots, E_W\}$ and let $\bar{x} \in \mathbb{R}_{++}^M$ be a fixed arbitrary vector. Then, for any $k \in \{1, 2, \dots, W\} \setminus \{t\}$, the vector given by

$$v^k := \text{diag}(a^t)^{-1} a^k - f_k(\bar{x}) \quad (4.13)$$

is a descent direction at the point \bar{x} for the function $g_k : \mathbb{R}_{++}^M \rightarrow \mathbb{R}_+$ defined as

$$g_k(\cdot) := \frac{1}{2} \|f_k(\cdot) - \text{diag}(a^t)^{-1} a^k\|^2. \quad (4.14)$$

Proof. The partial derivative of the j -th component of f_k is given by

$$\frac{\partial (f_k(x))_j}{\partial x_\ell} = \begin{cases} \frac{S_k}{S_t} (E_k - E_t) \frac{(f_k(x))_j}{x_j^2}, & \text{if } \ell = j, \\ 0, & \text{otherwise.} \end{cases} \quad (4.15)$$

Therefore, the gradient of g_k at the point $\bar{x} \in \mathbb{R}_{++}^M$ is given by

$$\nabla g_k(\bar{x}) = -\frac{S_k}{S_t} (E_k - E_t) \text{diag} \left(\left(\frac{(f_k(\bar{x}))_j}{\bar{x}_j^2} \right)_{j=1}^M \right) v^k. \quad (4.16)$$

Hence, we have

$$\nabla g_k(\bar{x})^T v^k = -\frac{S_k}{S_t} (E_k - E_t) \text{diag} \left(\left(\frac{(f_k(\bar{x}))_j}{\bar{x}_j^2} \right)_{j=1}^M \right) \|v^k\|^2 < 0, \quad (4.17)$$

since $S_k, S_t > 0, E_k - E_t > 0$ for all k and all diagonal elements of the matrix are positive. \square

Lemma 4.2. Let $\bar{x}, \bar{y} \in \mathbb{R}_{++}^M$. The vector w^k given by

$$w^k := a^k - \beta_k(\bar{x}, \bar{y}) = a^k - \text{diag} \left(\tilde{\beta}_k(\bar{x}) \right) \bar{y} \quad (4.18)$$

is a descent direction at the point \bar{y} for the function $h_k : \mathbb{R}_{++}^M \rightarrow \mathbb{R}_+$ defined by

$$h_k(y) := \frac{1}{2} \|a^k - \beta_k(\bar{x}, y)\|^2 = \frac{1}{2} \|a^k - \text{diag} \left(\tilde{\beta}_k(\bar{x}) \right) y\|^2, \quad \text{for all } y \in \mathbb{R}_{++}^M. \quad (4.19)$$

Proof. The gradient of h_k at the point \bar{y} is given by

$$\nabla h_k(\bar{y}) = -\text{diag} \left(\tilde{\beta}_k(\bar{x}) \right) w^k. \quad (4.20)$$

Thus, we have

$$\nabla h_k(\bar{y})^T w^k = -\text{diag} \left(\tilde{\beta}_k(\bar{x}) \right) \|w^k\|^2 < 0, \quad (4.21)$$

where the last strict inequality holds since $\tilde{\beta}_k : \mathbb{R}_{++}^M \rightarrow \mathbb{R}_{++}^M$. \square

The above analysis shows that the system of equations (4.9) fulfills Assumptions 1 and 2, making Algorithm 2 a good choice for handling it. In Section 4.5 we provide a numerical demonstration of its good performance.

4.4. Improving the descent pairs algorithm by superiorization. The iterative nature of Algorithm 2 enables us to introduce a priori conditions into the iterative process. We do this via the superiorization methodology.

4.4.1. The superiorization methodology. The superiorization methodology [29] works by taking an iterative algorithm, investigating its **perturbation resilience**, and then, using proactively such permitted perturbations, it forces the perturbed algorithm to do something useful in addition to what it was originally designed to do. The original unperturbed algorithm is called the **basic algorithm** and the perturbed algorithm is called the **superiorized version of the basic algorithm**.

When the basic algorithm is computationally efficient and useful in terms of the application at hand, and the perturbations are simple and not expensive to calculate, then the advantage of this methodology is that, for essentially the computational cost of the basic algorithm, we are able to get something more by steering its iterates according to the perturbations. A detailed description of the superiorization methodology along with pertinent up-to-date references can be found in several papers, see, e.g., [12], [28]

This general principle has been successfully used in a variety of important practical applications, see the recent papers in the, compiled and continuously updated, bibliography of scientific publications on the superiorization methodology and perturbation resilience of algorithms [10], where many applications oriented works with the method appear, e.g., [23].

In the language of [11], the superiorization methodology allows us to employ a given function, called **target function**, $\varphi : \mathbb{R}^M \rightarrow \mathbb{R}$. It interlaces into the iterations of a basic algorithm steps that perform locally reductions of the target function, these steps are called **perturbations**. The resulting **superiorized version of the basic algorithm** is expected to retain the convergence properties of the basic (unperturbed) algorithm but, additionally, steer the process to an output with reduced, not necessarily minimized, value of the target function.

We use this methodology to superiorize the descent pairs algorithm (Algorithm 2), as we describe below.

4.4.2. Implementation of the superiorized version of the descent pairs algorithm. In our implementations, priors are applied to Algorithm 2 according to the superiorization methodology. Corresponding target functions φ , mentioned above, to lead the function reduction perturbations in the superiorization process are the priors defined by [38]. These are the **total variation (TV)** function and the **smoothness** (Tikhonov, Tik for short) function [20], defined, respectively, by

$$\psi_{TV}(Z) := \sum_{i,j} \sqrt{(Z(i,j) - Z(i+1,j))^2 + (Z(i,j) - Z(i,j+1))^2}, \quad (4.22)$$

and

$$\psi_{Tik}(Z) := \sum_{i,j} \left(Z(i,j) - \frac{1}{rn} \sum_{ii,jj} (Z(ii,jj) - Z(i,j)) \right)^2, \quad (4.23)$$

where Z represents one of the properties x or y , $\{(i,j)\}_{i=1,j=1}^{n,n}$ is a set of pairs of indices of pixels in an exact grid in the ROI (see Figure 1), rn is the number of pixels that surround pixel (i,j) , and (ii,jj) are the indices of these pixels. The pseudo-code of this method applied to (4.9) is shown in Algorithm 4. In the experimental runs the function φ in the algorithm will be either $\psi_{TV}(Z)$ or $\psi_{Tik}(Z)$ of (4.22) or (4.23), respectively. Although in the experiments only the above two functions are chosen as target functions, and thus gradients are involved in the pseudo-code below, it should be emphasized that in general any target function can be considered and function reduction is not necessarily via derivatives [12, 13, 24]. To avoid zero as the divisor when taking gradient for the TV function, we add a small constant 10^{-5} under the square root sign in our application.

In this algorithm, λ is the step length of the DPA iteration and η is the step length for superiorization, which is contracted according to γ as the iteration processes. The value of γ and the initial value of η influence the impact of the superiorization together. We denote by v the normalized gradient of the function φ at the current iteration and z is an intermediate variable between the superiorization and the DPA iteration.

4.5. A numerical experiment. In this subsection, Algorithms 2, 3 and 4 are nicknamed as Algorithm DPA, NF and SUP-DPA, respectively. We conducted simulations on numerous temperature and concentration phantoms, two of which are shown here as demonstrative examples. Phantom 1 mimics the temperature and H₂O concentration in a two-dimensional McKenna flame. Phantom 2 imitates smooth temperature and concentration distributions consisting of two Gaussian peaks.

The square ROI was divided into 40×40 pixels grids and 160 beams were employed, which were distributed uniformly and parallelly in four directions, 0°, 45°, 90° and 135°. Ten discrete wavelengths were selected from the H₂O absorption spectrum.

In order to observe the progress and behavior of the DPA algorithm (Algorithm 2) for solving the second stage of the multi-spectral TAS problem, as explained in Subsection 4.2 above, we compared its performance with the NF algorithm (Algorithm 3). In our implementation with MATLAB, the nonlinear least-squares solver, i.e., the function “lsqnonlin”, was applied as the NF algorithm, wherein MATLAB’s trust-region algorithm was employed in this function to find the solution. For the solution of the first stage we employed the superiorized ART algorithm [5, 21].

Algorithm 4: The Superiorized version of the descent pairs algorithm (Algorithm 2), nicknamed SUP-DPA, applied in multi-spectral TAS.

```

1 Initialization: Set  $\lambda_x, \lambda_y > 0$ ,  $\eta_x$  and  $\eta_y$ . Set the index  $t$  such that
    $E_t = \min \{E_1, E_2, \dots, E_W\}$  and choose  $x^0$  and  $y^0$  in  $\mathbb{R}_+^M$ ;
2 set  $\ell = 0$ ;
3 repeat
4   set  $x^{\ell,0} = x^\ell$ ;
5   for  $q = 1, 2, \dots, W$  do
6     set  $v_x := -\frac{\nabla\varphi(x^{\ell,q-1})}{\|\nabla\varphi(x^{\ell,q-1})\|}$ ;
7     while  $\varphi(x^{\ell,q-1} + \beta_x v_x) > \varphi(x^{\ell,q-1})$  do
8       set  $\eta_x = \gamma\eta_x$ ;
9     end
10    set  $z_x = x^{\ell,q-1} + \eta_x v_x$ ;
11    set  $x^{\ell,q} = z_x + \lambda_x \left( \text{diag}(a^t)^{-1} a^q - \text{diag} \left( \tilde{\beta}_t(z_x) \right)^{-1} \tilde{\beta}_q(z_x) \right)$ ;
12  end
13  set  $x^{\ell+1} = x^{\ell,W}$ ;
14  set  $y^{\ell,0} = y^\ell$ ;
15  for  $q = 1, 2, \dots, W$  do
16    set  $v_y := -\frac{\nabla\varphi(y^{\ell,q-1})}{\|\nabla\varphi(y^{\ell,q-1})\|}$ ;
17    while  $\varphi(y^{\ell,q-1} + \beta_y v_y) > \varphi(y^{\ell,q-1})$  do
18      set  $\eta_y = \gamma\eta_y$ ;
19    end
20    set  $z_y = y^{\ell,q-1} + \eta_y v_y$ ;
21    set  $y^{\ell,q} = z_y + \lambda_y (a^q - \beta_q(x^{\ell+1}, z_y))$ ;
22  end
23  set  $y^{\ell+1} = y^{\ell,W}$ ;

```

To mimic practical situations, noise was added to the absorbance measurements. Uniform noise was used in our simulations, defined as

$$b_{i;mes}^k := b_{i;ori}^k \cdot (1 + rand \cdot u), \quad (4.24)$$

where the subscript *ori* refers to the original absorbance calculated from the phantom without noise, while *mes* refers to the practical absorbance with noise added; u is the noise level; *rand* is a random number in the interval $(-1, 1)$.

We tested our algorithms on two different phantoms, Phantom 1 and Phantom 2. For both phantoms, the DPA algorithm was initialized with $\lambda_x = 1000$, $\lambda_y = 2$, and x^0 and y^0

randomly obtained by picking their components between some lower and upper bounds. These bounds were set, for all $j = 1, 2, \dots, M$, as $x_j^0 \in (400, 2000)$ and $y_j^0 \in (0.005, 0.2)$ in the algorithmic runs for the reconstruction of Phantom 1, while for all $j = 1, 2, \dots, M$, $x_j^0 \in (800, 2400)$ and $y_j^0 \in (0.005, 0.2)$ in the algorithmic runs for the reconstruction of Phantom 2.

For the SUP-DPA algorithm (Algorithm 4), in the reconstruction of Phantom 1, TV was employed as the prior, i.e., as the function φ in the algorithm. The SUP-DPA algorithm was initialized with $\lambda_x = 1000$, $\lambda_y = 2$, $\beta_x = 5 \times 10^6$, $\beta_y = 10$ and $\gamma = 0.999$. Initial guesses of x^0 and y^0 were vectors randomly chosen in the intervals described above for the DPA algorithm.

For the reconstruction of Phantom 2 with the SUP-DPA algorithm, smoothness was regarded as the prior, thus, the function φ in the algorithm was chosen as ψ_{Tik} of (4.23) and the algorithm was initialized with $\lambda_x = 1000$, $\lambda_y = 2$, $\beta_x = 5 \times 10^4$, $\beta_y = 10$ and $\gamma = 0.999$. Initial guesses of x^0 and y^0 were vectors randomly picked in the intervals described above for the DPA algorithm.

All runs of the DPA and the SUP-DPA algorithms were stopped when either the number of iterations exceeded 50 or when $\sum_{k=1}^W \|b^k - \beta_k(x^{\ell+1}, y^{\ell+1})\| < 10^{-3}$. Figure 2 shows the relation between relative error in x and the stopping criterion function $\sum_{k=1}^W \|b^k - \beta_k(x^{\ell+1}, y^{\ell+1})\|$ as the iterations proceed for DPA and SUP-DPA. The relative errors are defined as

$$\begin{cases} Error_x := \frac{\|x_{rec} - x_{act}\|_2}{\|x_{act}\|_2}, \\ Error_y := \frac{\|y_{rec} - y_{act}\|_2}{\|y_{act}\|_2}, \end{cases} \quad (4.25)$$

where the subscripts “rec” and “act” stand for reconstructed value and actual value, respectively.

It can be seen that after the stopping criterion function value fell below 10^{-3} , the relative error hardly changes, indicating that a convergence is reached. The relative error of SUP-DPA had a slight fluctuation at the left end of the curve because of the perturbation introduced. On the other hand, 50 iterations is a large enough number for this testing case to assume convergence when the stopping criterion function value cannot decrease to 10^{-3} . For the NF algorithm, the function tolerance and optimality tolerance were both set to 5×10^{-10} . When implementing the NF algorithm, δ was initially set as the 2-norm of the difference between the upper and lower bounds, $((800, 2400)$ for x and $(0.005, 0.2)$ for y), and the adjustment of δ followed the default settings in the MATLAB function “lsqnonlin”. The algorithm was terminated when $\delta < 5 \times 10^{-10}$.

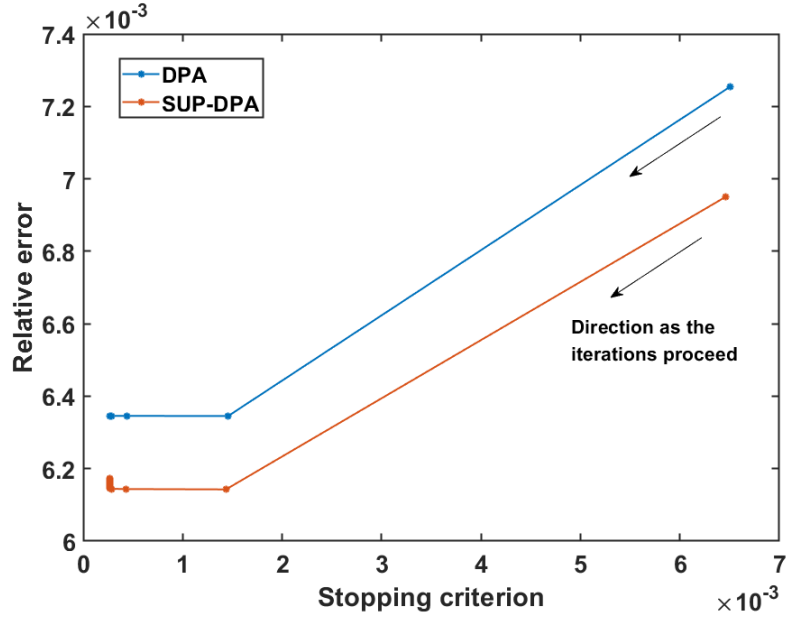


FIGURE 2. Relation between relative error and stopping criterion as the iterations proceed.

Results are shown in Figures 3 and 4 and Table 1. The name of the algorithms is shown at the top of the corresponding reconstructed profiles. The relative reconstruction errors are also labeled, which are defined by (4.25). To be specific, the unit of temperature x was K , which is labeled beneath its color-bar in the figures, and for the mole fraction y , the unit is one, thus, no unit was labeled.

TABLE 1. Computation times of the algorithms

Computation time (sec)	DPA	SUP-DPA	NF
Phantom 1	0.48	0.56	9.16
Phantom 2	0.48	0.54	7.67

Figure 3 shows Phantom 1 and the results recovered by the DPA, the SUP-DPA and the NF algorithms. Temperature and concentration profiles generated by the three algorithms show similar shapes to those of Phantom 1, although there are some defects at the edges of the profiles. The reconstruction errors of SUP-DPA is smaller than those of the DPA and the NF algorithm. Besides, it takes much less time for the DPA and the SUP-DPA algorithms to finish the reconstruction than it took for the NF algorithm. Overall, the SUP-DPA algorithm can be regarded as the best among the three algorithms, for the particular experiments that we conducted and report here.

Figure 4 shows Phantom 2 and the results recovered by the DPA, the SUP-DPA and the NF algorithms. Results by the three methods show a similar shape to Phantom 2.

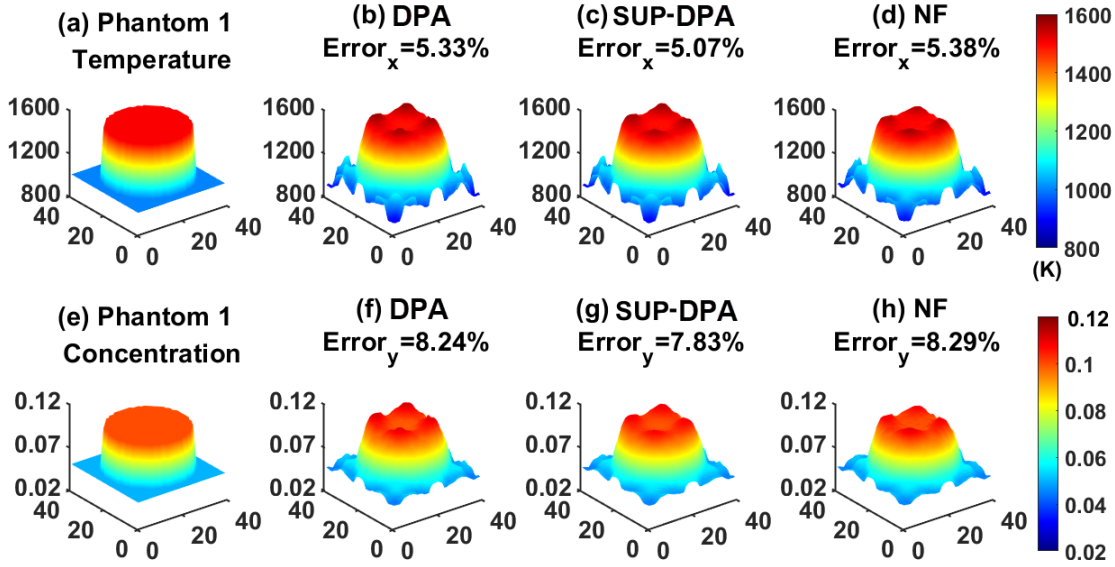


FIGURE 3. (a) Temperature profile of Phantom 1. Temperature profiles recovered by (b) the DPA algorithm, (c) the SUP-DPA algorithm, and (d) the NF algorithm. (e) Concentration profile of Phantom 1. Concentration profiles recovered by (f) the DPA algorithm, (g) the SUP-DPA algorithm, and (h) the NF algorithm.

The error of the DPA algorithm is slightly smaller than that of the NF algorithm, while the SUP-DPA algorithm improves the error of both other algorithms. Besides, the computational time of the SUP-DPA algorithm is much smaller than that of the NF algorithm. Though the DPA algorithm has the lowest computation time, its reconstruction effect is worse than that of the SUP-DPA algorithm. Therefore, the SUP-DPA algorithm can still be regarded advantageous over the DPA and the NF algorithms.

To further investigate the performance of the algorithms, simulation studies were conducted under different spatial resolutions. For Phantom 2, reconstructions were conducted when it was divided into 20×20 , 40×40 , 60×60 and 80×80 pixel grids, denoted as “gridding scale” 20, 40, 60 and 80, respectively. To ensure the comparability of the results, the measurement of each case was conducted in four directions, while the number of beams in each direction increased correspondingly to the gridding scale, which means that for a $G \times G$ grid, $4 \times G$ measurement beams were applied. Other conditions, including noise level, parameters and stopping criteria, were the same as those described in Figure 4. The results are shown in Figure 5.

As the pixel grid was made finer, the computation time grew exponentially, indicating an increasing computational efficiency improvement of DPA and SUP-DPA compared with NF. In each simulation condition, DPA and SUP-DPA had better performance than

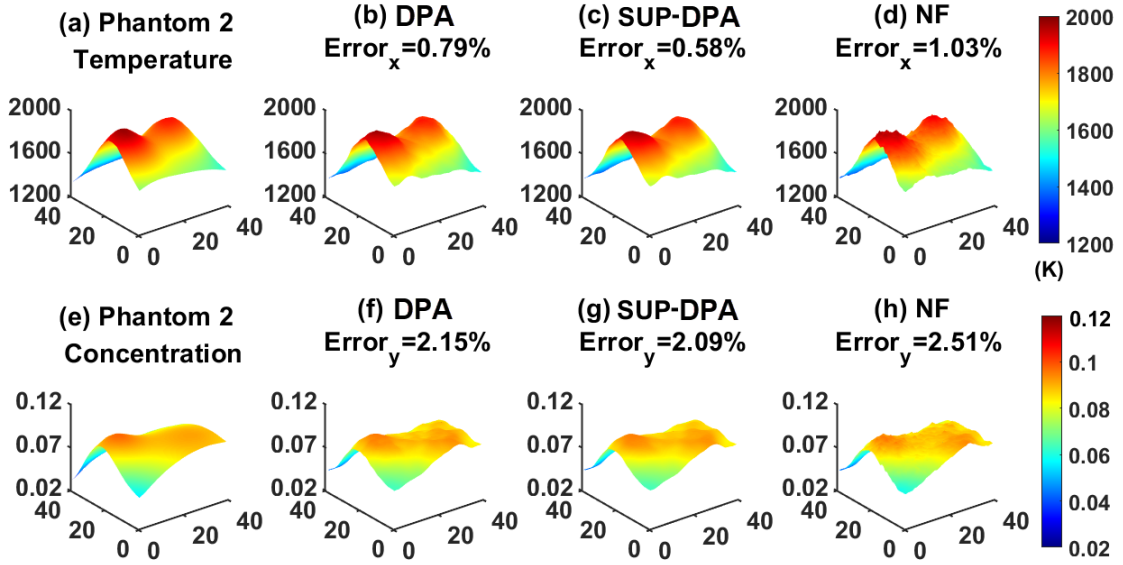


FIGURE 4. (a) Temperature profile of Phantom 2. Temperature profiles recovered by (b) the DPA algorithm, (c) the SUP-DPA algorithm, and (d) the NF algorithm. (e) Concentration profile of Phantom 2. Concentration profiles recovered by (f) the DPA algorithm, (g) the SUP-DPA algorithm, and (h) the NF algorithm.

NF, with SUP-DPA smaller than DPA in relative error and DPA slightly smaller in computation time. At gridding scale 20, there were the largest errors. This might be due to less measurement beams. Coarser gridding made the phantom less smooth, which could result in difficulties for the reconstruction.

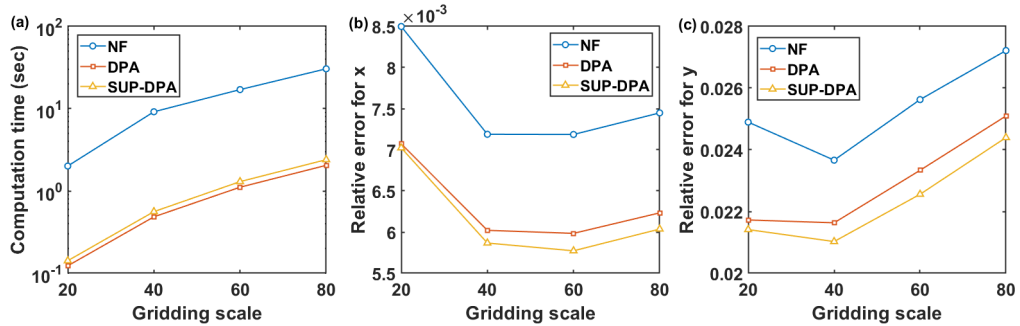


FIGURE 5. (a) Computation time, (b) reconstruction error for x , and (c) reconstruction error for y , with respect to the gridding scale.

Acknowledgements. F.J.A.A. and D.T.B. were partially supported by Grants PGC2018-097960-B-C22 and PID2022-136399NB-C21 funded by ERDF/EU and by MICIU/AEI/10.13039/501100011033. D.T.B. was supported by Grant PRE2019-09075 funded by MICIU/AEI/10.13039/501100011033 and by “ESF Investing in your future”. F.J.A.A. was partially supported by the Generalitat Valenciana (AICO/2021/165). The work of Y.C. was supported by the ISF-NSFC joint research program grant No. 2874/19, by the U.S. National Institutes of Health (NIH) Grant No. R01CA266467 and by the Cooperation Program between the German Cancer Research Center (DKFZ) and Israel’s Ministry of Innovation, Science and Technology (MOST). W. C. and C. S. were partially supported by the National Natural Science Foundation of China (Grant No. 51976122, 52061135108).

DECLARATIONS

Conflict of interest. The authors declare no competing interests.

REFERENCES

- [1] T.O. Alakoya, V.A. Uzor, and O.T. Mewomo. A new projection and contraction method for solving split monotone variational inclusion, pseudomonotone variational inequality, and common fixed point problems. *Computational and Applied Mathematics*, 42, 3, 2023.
- [2] F.J. Aragón, M.A. Goberna, M.A. López, and M.M.L. Rodríguez. *Nonlinear Optimization*. Springer Nature Switzerland AG, 2019.
- [3] V. Berindee. *Iterative Approximation of Fixed Points*. Springer-Verlag, Berlin, Heidelberg, Germany, 2nd revised and enlarged edition, 2007.
- [4] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [5] D. Butnariu, R. Davidi, G.T. Herman, and I.G. Kazantsev. Stable convergence behavior under summable perturbations of a class of projection methods for convex feasibility and optimization problems. *IEEE Journal of Selected Topics in Signal Processing*, 1:540–547, 2007.
- [6] W. Cai, D.J. Ewing, and L. Ma. Application of simulated annealing for multispectral tomography. *Computer Physics Communications*, 179:250–255, 2008.
- [7] W. Cai and C.F. Kaminski. Tomographic absorption spectroscopy for the study of gas dynamics and reactive flows. *Progress in Energy and Combustion Science*, 59:1–31, 2017.
- [8] W. Cai and L. Ma. Applications of critical temperature in minimizing functions of continuous variables with simulated annealing algorithm. *Computer Physics Communications*, 181:11–16, 2010.
- [9] A. Cegielski. *Iterative Methods for Fixed Point Problems in Hilbert Spaces*. Lecture Notes in Mathematics 2057, Springer-Verlag, Berlin, Heidelberg, Germany, 2012.
- [10] Y. Censor. Superiorization and perturbation resilience of algorithms: A bibliography compiled and continuously updated. <http://math.haifa.ac.il/yair/bib-superiorization-censor.html>.
- [11] Y. Censor. Weak and strong superiorization: Between feasibility-seeking and minimization. *Analele Stiintifice ale Universitatii "Ovidius" Constanta-Seria Matematica*, 23:41–54, 2015.
- [12] Y. Censor, E. Garduño, E.S. Helou, and G.T. Herman. Derivative-free superiorization: Principle and algorithm. *Numerical Algorithms*, 88:227–248, 2021.
- [13] Y. Censor, H. Heaton, and R. Schulte. Derivative-free superiorization with component-wise perturbations. *Numerical algorithms*, 80:1219–1240, 2019.
- [14] Y. Censor and A. Segal. On the string averaging method for sparse common fixed-point problems. *International Transactions in Operational Research*, 16:481–494, 2009.
- [15] T.F. Coleman and Y. Li. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, 6:418–445, 1996.
- [16] P.L. Combettes. Quasi-Fejérian analysis of some optimization algorithms. In D. Butnariu, Y. Censor, and S. Reich, editors, *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*, pages 115–152. Elsevier Science Publishers, Amsterdam, The Netherlands, 2001.
- [17] P.L. Combettes and J.-C. Pesquet. Linear convergence of stochastic block-coordinate fixed point algorithms. pages 742–746. 2018 26th European Signal Processing Conference (EUSIPCO). <https://hal.archives-ouvertes.fr/hal-01964580/document>.
- [18] P.L. Combettes and Z.C. Woodstock. A fixed point framework for recovering signals from nonlinear transformations. pages 2120–2124. 2020 28th European Signal Processing Conference (EUSIPCO). <https://www.eurasip.org/Proceedings/Eusipco/Eusipco2020/pdfs/0002120.pdf>.
- [19] J. Dai, T. Yu, L. Xu, and W. Cai. On the regularization for nonlinear tomographic absorption spectroscopy. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 206:233–241, 2018.
- [20] K.J. Daun. Infrared species limited data tomography through Tikhonov reconstruction. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 111:105–115, 2010.

- [21] R. Davidi, G.T. Herman, and Y. Censor. Perturbation-resilient block-iterative projection methods with application to image reconstruction from projections. *International Transactions in Operational Research*, 16:505–524, 2009.
- [22] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39:1–22, 1977.
- [23] J. Fink, R.L.G. Cavalcante, and S. Stańczak. Multi-group multicast beamforming by superiorized projections onto convex sets. *IEEE Transactions on Signal Processing*, 69:5708–5722, 2021.
- [24] A. Gibali and S. Petra. DC-programming versus ℓ -superiorization for discrete tomography. *Analele Universitatii “Ovidius” Constanta-Seria Matematica*, 26:105–133, 2018.
- [25] E.C. Godwin, O.T. Mewomo, and T.O. Alakoya. A strongly convergent algorithm for solving multiple set split equality equilibrium and fixed point problems in Banach spaces. *Proceedings of the Edinburgh Mathematical Society*, 66:475–515, 2023.
- [26] R. Gordon, R. Bender, and G.T. Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography. *Journal of Theoretical Biology*, 29:471–481, 1970.
- [27] S.J. Grauer, J. Emmert, S.T. Sanders, S. Wagner, and K.J. Daun. Multiparameter gas sensing with linear hyperspectral absorption tomography. *Measurement Science and Technology*, 30:105401, 2019.
- [28] G.T. Herman. Iterative reconstruction techniques and their superiorization for the inversion of the radon transform. In R. Ramlau and O. Scherzer, editors, *The Radon Transform: The First 100 Years and Beyond*, pages 217–238. De Gruyter, 2019.
- [29] G.T. Herman, E. Garduño, R. Davidi, and Y. Censor. Superiorization: An optimization heuristic for medical physics. *Medical Physics*, 39:5532–5546, 2012.
- [30] D. Jinghang, Y. Tao, X. Lijun, and C. Weiwei. On the regularization for nonlinear tomographic absorption spectroscopy. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 206:233–241, 2018.
- [31] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.
- [32] J.P.B. Leite and B.H.V. Topping. Parallel simulated annealing for structural optimization. *Computers & Structures*, 73:545–564, 1999.
- [33] M. Lin, X. Li, W. Cai, S. Roy, J.R. Gord, and S.T. Sanders. Selection of multiple optimal absorption transitions for nonuniform temperature sensing. *Applied Spectroscopy*, 64:1274–1282, 2010.
- [34] L. Ma, W. Cai, A.W. Caswell, T. Kraetschmer, S.T. Sanders, S. Roy, and J.R. Gord. Tomographic imaging of temperature and chemical species based on hyperspectral absorption spectroscopy. *Optics Express*, 17:8602–8613, 2009.
- [35] J.J. Moré and D.C. Sorensen. Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing*, 4:553–572, 1983.
- [36] J.M. Ortega and W.C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. SIAM, Philadelphia, PA, USA, 2000.
- [37] W.C. Rheinboldt. *Methods for Solving Systems of Nonlinear Equations*. SIAM, Philadelphia, PA, USA, 1998.
- [38] C. Shui, Y. Wang, W. Cai, and B. Zhou. Linear multispectral absorption tomography based on regularized iterative methods. *Optics Express*, 29:20889–20912, 2021.
- [39] A. Taiwo, L.O. Jolaoso, and O.T. Mewomo. Viscosity approximation method for solving the multiple-set split equality common fixed-point problems for quasi-pseudocontractive mappings in Hilbert spaces. *Journal of Industrial and Management Optimization*, 17:2733–2759, 2021.
- [40] A.J. Zaslavski. *Approximate Solutions of Common Fixed-Point Problems*. Springer International Publishing Switzerland, 2016.