SIAM J. SCI. COMPUT.
Vol. 46, No. 2, pp. S142–S175

VARIATIONAL DATA ASSIMILATION AND ITS DECOUPLED ITERATIVE NUMERICAL ALGORITHMS FOR STOKES–DARCY MODEL*

XUEJIAN LI†, WEI GONG‡, XIAOMING HE†, AND TAO LIN§

Abstract. In this paper we develop and analyze a variational data assimilation method with efficient decoupled iterative numerical algorithms for the Stokes–Darcy equations with the Beavers–Joseph interface condition. By using Tikhonov regularization and formulating the variational data assimilation into an optimization problem, we establish the existence, uniqueness, and stability of the optimal solution. Based on the weak formulation of the Stokes–Darcy equations, the Lagrange multiplier rule is utilized to derive the first order optimality system for both the continuous and discrete variational data assimilation problems, where the discrete data assimilation is based on a finite element discretization in space and the backward Euler scheme in time. By rescaling the optimality system and then analyzing its corresponding bilinear forms, we prove the optimal finite element convergence rate with special attention paid to recovering uncertainties missed in the optimality system. To solve the discrete optimality system efficiently, three decoupled iterative algorithms are proposed to address the computational cost for both well-conditioned and ill-conditioned variational data assimilation problems, respectively. Finally, numerical results are provided to validate the proposed methods.

Key words. data assimilation, Stokes–Darcy equations, optimization, gradient-based iterative method, finite element method

MSC codes. 35Q93, 49M41, 65K10, 76S05, 93C20

DOI. 10.1137/22M1492994

1. Introduction.

Data assimilation seeks to optimally incorporate observations into a physics model for improving performance of the state forecast. This technique is employed in many applications, such as weather prediction [14, 15, 77], ocean state forecasts [2, 41, 79], geoscience [12, 81, 85], and chemistry transport [58, 82, 84], among many others. Currently there are several main categories of data assimilation techniques. One category includes the statistical methods based on the Bayes' theorem and the Kalman filtering approach, which evolve the state vector along with time according to error statistics [1, 3, 36, 65, 69, 76]. Another category includes the variational methods based on the optimal control theory, which minimizes a cost functional measuring the discrepancy between the state variable and the observed data [11, 26, 38, 43, 61, 70]. The nudging method and continuous data assimilation are also important and popular data assimilation techniques; see [7, 8, 42, 68, 72, 75, 90, 92]. In this paper we discuss a variational method to solve the data assimilation problem for a Stokes–Darcy interface model.

*Received by the editors April 27, 2022; accepted for publication (in revised form) January 13, 2023; published electronically July 18, 2023.
https://doi.org/10.1137/22M1492994

Funding: The first and third authors were supported by NSF grants DMS-1722647 and DMS-2152609. The second author was supported in part by the Strategic Priority Research Program of the Chinese Academy of Sciences (grant XDB 41000000) and the NSFC (grant 12071468).

†Department of Mathematics and Statistics, Missouri University of Science and Technology, Rolla, MO 65409 USA (xlcdt@mst.edu, hex@mst.edu).

‡LSEC, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Haidian, Beijing 100190, People's Republic of China (wgong@lsec.cc.ac.cn).

§Department of Mathematics, Virginia Tech, Blacksburg, VA 24061 USA (tlin@vt.edu).

The Stokes–Darcy model has attracted significant attention due to its potential applications to a variety of flow phenomena, for instance, the hydrological system where surface water percolates through rock and sand [21, 24, 27, 30, 33, 45, 46, 56], petroleum extraction [4, 5, 22, 39, 40, 51, 52, 64], and industrial filtration [35]. In recent decades, there has been significant effort to study this sophisticated interface system and its variations both theoretically and numerically [6, 9, 13, 16, 18, 23, 25, 28, 29, 31, 32, 34, 44, 47, 53, 54, 57, 60, 62, 71, 73, 78, 80, 83, 91]. However, these works were dedicated to the idealized model, i.e., the relevant input data, such as initial condition, boundary condition, sink/source term, and diffusion coefficients, are entirely provided for the model prediction. In real implementations, some of these input data literally remain unknown or in uncertainty. Therefore, one of the challenging problems is to identify a set of faithful input data such that the forecast of the target flow can be estimated reliably. This is where the data assimilation comes in.

A primary interest of this paper is to investigate a variational data assimilation (VDA) for the Stokes–Darcy model and develop decoupled iterative numerical algorithms to efficiently solve the VDA problem. Through theoretical derivation and numerical iteration, we focus on identifying a faithful initial condition for the model such that the flow state can be better predicted. To be able to construct reliable efficient iterative numerical algorithms, we first need to mathematically formulate the VDA problem, and then rigorously carry out the corresponding theoretical derivation and analysis as a solid foundation. Our approach to achieve the goal is to incorporate the noisy observation into the Stokes–Darcy model through an appropriately designed cost functional, where the L^2-Tikhonov regularization is utilized, and the discontinuity on initial condition is admitted. The data assimilation problem hereby becomes a minimization problem. Existence and uniqueness of such a minimization problem are established. Stability with respect to the perturbation on observations and the regularization parameter is also proven. The Lagrange multiplier rule is utilized to derive the first order optimality system (OptS) for both the continuous and the discrete VDA, where the discrete VDA is constructed based on a finite element method (FEM) and the backward Euler scheme. Afterward, we analyze the convergence features between the discrete optimal solution and the solution to the continuous data assimilation problem. Particularly, with smooth enough input data, we prove the optimal finite element convergence rate. The analysis is carried out by rescaling the continuous OptS and rigorously proving its essential properties, such as Gårding type coercivity and continuity, for relevant bilinear forms. The necessary auxiliary equations are then delicately introduced in order to fill in the gaps or difference between the classical partial differential equations (PDEs) and the PDEs in the optimality systems. The optimal convergence rate is finally achieved by transferring the convergence behavior of the VDA problem to the convergence results of the classical Stokes–Darcy equation and a backward Stokes-type equation. All of these theoretical works lay a solid foundation for both the numerical computation of the VDA and the further development of efficient numerical algorithms. Hence we finally develop three decoupled iterative algorithms based on the conjugate gradient (CG) method, the BFGS method, and the steepest descent (SD) method, which greatly reduce the computational cost for solving the discrete optimality system.

The outline of this article is as follows. In section 2, we introduce the basic formulation for the Stokes–Darcy model and provide the necessary mathematical preliminaries. In section 3, we prove the well-posedness of the continuous data assimilation problem and derive the optimality system. In section 4, we discuss the finite element approximation to the continuous data assimilation problem and present its convergence analysis. In section 5, three iterative methods are presented to address

the extreme computational cost. In section 6, numerical experiments are presented to verify the proposed methods. In section 7, we draw some conclusions.

2. Mathematics formulation of the Stokes–Darcy model. We consider a free flow in a bounded domain Ω_f and a porous media flow in another bounded domain Ω_p. These two flows are coupled together in the domain Ω through the interface $\Gamma = \bar{\Omega}_p \cap \bar{\Omega}_f$ such that $\bar{\Omega} = \bar{\Omega}_p \cup \bar{\Omega}_f$. We also let $\Gamma_p = \partial\Omega_p \setminus \Gamma$ and $\Gamma_f = \partial\Omega_f \setminus \Gamma$. A Stokes–Darcy model can be used to describe this coupled fluid phenomena, in which the porous media flow is governed by the Darcy equation,

$$\begin{aligned} &\frac{\partial \phi}{\partial t} - \nabla \cdot (\mathbb{K}\nabla\phi) = f_p \quad \text{in}\,\Omega_p \times (0,T], \\ &\phi(\cdot,0) = \phi_0 \quad \text{in}\ \Omega_p, \\ &\phi = 0 \quad \text{on}\ \Gamma_p, \end{aligned} \tag{2.1}$$

where ϕ denotes the hydraulic head, $\mathbb{K}$ is the hydraulic conductivity tensor assumed to be positive definite, f_p is a sink/source term, and the free flow is governed by the Stokes equation,

$$\begin{aligned} &\frac{\partial \boldsymbol{u}}{\partial t} - \nabla \cdot \mathbb{T}(\boldsymbol{u}, p) = \boldsymbol{f}_f \quad \text{in}\,\Omega_f \times (0,T], \\ &\nabla \cdot \boldsymbol{u} = 0 \quad \text{in}\,\Omega_f \times (0,T], \\ &\boldsymbol{u}(\cdot,0) = \boldsymbol{u}_0 \quad \text{in}\ \Omega_f, \\ &\boldsymbol{u} = 0 \quad \text{on}\ \Gamma_f, \end{aligned} \tag{2.2}$$

where $\boldsymbol{u}$ denotes the fluid velocity, $\mathbb{T}(\boldsymbol{u}, p) = 2\nu\mathbb{D}(\boldsymbol{u}) - p\mathbb{I}$ is the stress tensor, $\mathbb{D}(\boldsymbol{u}) = \frac{1}{2}(\nabla\boldsymbol{u} + \nabla^T\boldsymbol{u})$ is the deformation tensor, ν is the kinematic viscosity of the fluid, p is the kinematic pressure, and $\boldsymbol{f}_f$ is a general external forcing term. Systems (2.1) and (2.2) interact on Γ through the Beavers–Joseph interface conditions [10, 17, 19, 20, 48, 63, 74, 88, 89]:

$$\begin{aligned} \boldsymbol{u} \cdot \boldsymbol{n}_f = \mathbb{K}\nabla\phi \cdot \boldsymbol{n}_p, \quad -\boldsymbol{\tau} \cdot (\mathbb{T}(\boldsymbol{u}, p) \cdot \boldsymbol{n}_f) = \alpha\boldsymbol{\tau} \cdot (\boldsymbol{u} + \mathbb{K}\nabla\phi), \\ -\boldsymbol{n}_f \cdot (\mathbb{T}(\boldsymbol{u}, p) \cdot \boldsymbol{n}_f) = g(\phi - z), \end{aligned} \tag{2.3}$$

where $\boldsymbol{n}_f$ and $\boldsymbol{n}_p$ denote the outer normal vectors to the fluid and the porous media regions on the interface Γ, respectively, $\boldsymbol{\tau}$ denotes the unit tangential vector to the interface Γ, α is a permeability constant depending on ν and $\mathbb{K}$, g is the gravitational acceleration, and z is a constant assumed to be 0 from now on.

For the purpose of discussing the data assimilation problem, it is necessary to appropriately understand the Stokes–Darcy model. We first define the Hilbert spaces

$$\begin{aligned} &X_p := \{\psi \in H^1(\Omega_p) :\ \psi = 0 \text{ on } \partial\Omega_p \setminus \Gamma\}, \\ &\boldsymbol{X}_f := \{\boldsymbol{v} \in \boldsymbol{H}^1(\Omega_f) :\ \boldsymbol{v} = (v_1, v_2)^T = \boldsymbol{0} \text{ on } \partial\Omega_f \setminus \Gamma\}, \\ &\boldsymbol{X} := X_p \times \boldsymbol{X}_f, \quad \boldsymbol{X}_{\text{div}}^f := \{\boldsymbol{v} \in \boldsymbol{X}_f :\ \nabla \cdot \boldsymbol{v} = 0\}, \quad \boldsymbol{X}_{\text{div}} := X_p \times \boldsymbol{X}_{\text{div}}^f, \\ &Q := \boldsymbol{L}^2(\Omega_f), \quad \boldsymbol{L}^2(\Omega) := L^2(\Omega_p) \times \boldsymbol{L}^2(\Omega_f) \end{aligned}$$

and the corresponding norms

$$\begin{aligned} &\|\psi\|_{X_p} := \|\psi\|_{H^1(\Omega_p)}, \quad \|\boldsymbol{v}\|_{\boldsymbol{X}_f} := \|\boldsymbol{v}\|_{\boldsymbol{H}^1(\Omega_f)} = \left(\|v_1\|_{H^1(\Omega_f)}^2 + \|v_2\|_{H^1(\Omega_f)}^2\right)^{\frac{1}{2}}, \\ &\|\boldsymbol{V}\|_{\boldsymbol{X}} := \left(\|\boldsymbol{v}\|_{\boldsymbol{X}_f}^2 + \|\psi\|_{X_p}^2\right)^{\frac{1}{2}}, \quad \|\boldsymbol{v}\|_{\boldsymbol{X}_{\text{div}}^f} := \|\boldsymbol{v}\|_{\boldsymbol{H}^1(\Omega_f)}, \\ &\|\boldsymbol{V}\|_{\boldsymbol{X}_{\text{div}}} := \left(\|\boldsymbol{v}\|_{\boldsymbol{X}_f}^2 + \|\psi\|_{X_p}^2\right)^{\frac{1}{2}}, \quad \|\boldsymbol{V}\|_{\boldsymbol{L}^2(\Omega)} := \left(\|\psi\|_{L^2(\Omega_p)}^2 + \|\boldsymbol{v}\|_{\boldsymbol{L}^2(\Omega_f)}^2\right)^{\frac{1}{2}}. \end{aligned}$$

For a domain D, $(\cdot,\cdot)_D$ denotes the L^2 inner product on D, and $(\cdot,\cdot)_H$ denotes the inner product for other Hilbert spaces $H(D)$. Depending on the context, $\langle\cdot,\cdot\rangle$ can represent the inner product on the interface Γ or a general duality between a Banach space and its dual space. For simplicity, let $\|\cdot\|_0$ denote all the L^2 norms and $H^m(D)$ denote the Sobolev space $W^{m,2}(D)$. Besides, considering the temporal-spatial function spaces, let $L^p(0,T;\mathcal{B}) = W^{0,p}(0,T;\mathcal{B})$ and $H^m(0,T;\mathcal{B}) = W^{m,2}(0,T;\mathcal{B})$, where $\mathcal{B}$ is a generic Banach space. We use these notations to define the following bilinear forms and linear functionals:

(2.4)

$$a_p(\phi,\psi) = (K\nabla\phi,\nabla\psi)_{\Omega_p} \ \forall\phi,\psi\in X_p,\ \ a_f(\boldsymbol{u},\boldsymbol{v}) = 2\nu\left(\mathbb{D}(\boldsymbol{u}),\mathbb{D}(\boldsymbol{v})\right)_{\Omega_f} \ \forall\boldsymbol{u},\boldsymbol{v}\in\boldsymbol{X}_f,$$

$$a(\boldsymbol{U},\boldsymbol{V}) = a_f(\boldsymbol{u},\boldsymbol{v}) + a_p(\phi,\psi) + \langle g\phi,\boldsymbol{v}\cdot\boldsymbol{n}_f\rangle - \langle\boldsymbol{u}\cdot\boldsymbol{n}_f,\psi\rangle$$

(2.5)
$$+\alpha\langle P_\tau(\boldsymbol{u}+\mathbb{K}\nabla\phi),P_\tau\boldsymbol{v}\rangle \quad \forall\boldsymbol{U}=(\phi,\boldsymbol{u})^T\in\boldsymbol{X},\ \forall\boldsymbol{V}=(\psi,\boldsymbol{v})^T\in\boldsymbol{X},$$

(2.6)

$$b_f(\boldsymbol{v},p) = -(\nabla\cdot\boldsymbol{v},p)_{\Omega_f},\quad b(\boldsymbol{V},p) = b_f(\boldsymbol{v},p)\ \forall\boldsymbol{v}\in\boldsymbol{X}_f,\ \forall\boldsymbol{V}=(\psi,\boldsymbol{v})^T\in\boldsymbol{X},\ \forall p\in Q,$$

(2.7)

$$\langle\boldsymbol{F},\boldsymbol{V}\rangle = (f_p,\psi)_{\Omega_p} + (\boldsymbol{f}_f,\boldsymbol{v})_{\Omega_f} \quad \forall\boldsymbol{F}=(f_p,\boldsymbol{f}_f)^T\in\boldsymbol{X}',\ \forall\boldsymbol{V}=(\psi,\boldsymbol{v})^T\in\boldsymbol{X},$$

(2.8)

$$\left\langle\frac{\partial\boldsymbol{U}}{\partial t},\boldsymbol{V}\right\rangle = \left\langle\frac{\partial\phi}{\partial t},\psi\right\rangle + \left\langle\frac{\partial\boldsymbol{u}}{\partial t},\boldsymbol{v}\right\rangle \quad \forall\boldsymbol{U}=(\phi,\boldsymbol{u})^T\in\boldsymbol{H}^1(0,T;\boldsymbol{X}'),\ \forall\boldsymbol{V}=(\psi,\boldsymbol{v})^T\in\boldsymbol{X},$$

where P_τ denotes the projection onto the tangent space on Γ, i.e., $P_\tau\boldsymbol{u} = (\boldsymbol{u}\cdot\boldsymbol{\tau})\boldsymbol{\tau}$. For $\langle P_\tau(\mathbb{K}\nabla\phi),P_\tau\boldsymbol{v}\rangle$ in (2.5), we need the trace space defined as $\mathbf{H}_{\mathbf{00}}^{\mathbf{1/2}}(\Gamma) := \boldsymbol{X}_f|_\Gamma$, which is a nonclosed subspace of $\mathbf{H}_{\mathbf{0}}^{\mathbf{1/2}}(\Gamma)$ and has continuous zero extension to $\mathbf{H}_{\mathbf{0}}^{\mathbf{1/2}}(\partial\Omega_f)$; $\langle P_\tau(\mathbb{K}\nabla\phi),P_\tau\boldsymbol{v}\rangle$ is then interpreted as a duality between $(\mathbf{H}_{\mathbf{00}}^{\mathbf{1/2}}(\Gamma))'$ and $\mathbf{H}_{\mathbf{00}}^{\mathbf{1/2}}(\Gamma)$. See [20] and references cited therein for more details.

We use bilinear forms $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ to define linear operators A, A^*, B, and B^*:

$$a(\boldsymbol{U},\boldsymbol{V}) = \langle A\boldsymbol{U},\boldsymbol{V}\rangle = \langle\boldsymbol{U},A^*\boldsymbol{V}\rangle,\quad b(\boldsymbol{V},p) = \langle B\boldsymbol{V},p\rangle = \langle\boldsymbol{V},B^*p\rangle, \tag{2.9}$$

where $A\in\mathscr{L}(\boldsymbol{X},\boldsymbol{X}')$, $B\in\mathscr{L}(\boldsymbol{X},Q')$, $A^*\in\mathscr{L}(\boldsymbol{X},\boldsymbol{X}')$, and $B^*\in\mathscr{L}(Q,\boldsymbol{X}')$ are the adjoint operators of A and B, and $\mathscr{L}$ is a collection of linear and continuous operators in the relevant spaces.

Testing systems (2.1) and (2.2) with $(\psi,\boldsymbol{v},q)^T\in\boldsymbol{X}\times Q$ and incorporating the three interface conditions (2.3), we obtain the weak formulation of the Stokes–Darcy model:

(2.10)

$$\begin{cases}\left\langle\dfrac{\partial\phi}{\partial t},\psi\right\rangle + a_p(\phi,\psi) - \langle\boldsymbol{u}\cdot\boldsymbol{n}_f,\psi\rangle = \langle f_p,\psi\rangle \quad \forall\psi\in X_p,\\ \left\langle\dfrac{\partial\boldsymbol{u}}{\partial t},\boldsymbol{v}\right\rangle + a_f(\boldsymbol{u},\boldsymbol{v}) + b_f(\boldsymbol{v},p) + \langle g\phi,\boldsymbol{v}\cdot\boldsymbol{n}_f\rangle + \alpha\langle P_\tau(\boldsymbol{u}+\mathbb{K}\nabla\phi),P_\tau\boldsymbol{v}\rangle = \langle\boldsymbol{f}_f,\boldsymbol{v}\rangle\\ \forall\boldsymbol{v}\in\boldsymbol{X}_f,\\ b_f(\boldsymbol{u},q) = 0 \quad \forall q\in Q,\\ \phi(\cdot,0) = \phi_0 \quad \text{in } L^2(\Omega_p),\ \boldsymbol{u}(\cdot,0) = \boldsymbol{u}_0 \quad \text{in } \boldsymbol{L}^2(\Omega_f).\end{cases}$$

By definitions in (2.4)–(2.9) and denoting $\boldsymbol{U_0} = (\phi_0,\boldsymbol{u}_0)^T$, (2.10) is equivalent to the following expression:

$$\begin{cases} \left\langle \dfrac{\partial \boldsymbol{U}}{\partial t}, \boldsymbol{V} \right\rangle + a(\boldsymbol{U}, \boldsymbol{V}) + b(\boldsymbol{V}, p) = \langle \boldsymbol{F}, \boldsymbol{V} \rangle \qquad \forall \boldsymbol{V} \in \boldsymbol{X}, \\ b(\boldsymbol{U}, q) = 0 \quad \forall q \in Q, \\ \boldsymbol{U}(\cdot, 0) = \boldsymbol{U_0} \quad \text{in } \boldsymbol{L}^2(\Omega). \end{cases} \tag{2.11}$$

An operator form of (2.11) can be written as

$$\begin{cases} \dfrac{\partial \boldsymbol{U}}{\partial t} + A\boldsymbol{U} + B^* p = \boldsymbol{F} \quad \text{in } \boldsymbol{X}', \\ B\boldsymbol{U} = 0 \quad \text{in } Q', \\ \boldsymbol{U}(\cdot, 0) = \boldsymbol{U_0} \quad \text{in } \boldsymbol{L}^2(\Omega). \end{cases} \tag{2.12}$$

If we consider $B\boldsymbol{U} = 0 \in Q'$ in the above as a constraint and restrict the discussion in space $\boldsymbol{X}_{\text{div}}$, a more concise form of (2.12) is

$$\begin{cases} \dfrac{\partial \boldsymbol{U}}{\partial t} + A\boldsymbol{U} = \boldsymbol{F} \quad \text{in } \boldsymbol{X}'_{\text{div}}, \\ \boldsymbol{U}(\cdot, 0) = \boldsymbol{U_0} \quad \text{in } \boldsymbol{L}^2(\Omega). \end{cases} \tag{2.13}$$

Recalling that (2.10), (2.11), (2.12), and (2.13) are all equivalent, the well-posedness of each is further guaranteed by the continuous inf-sup condition [20]:

$$\inf_{0 \neq q \in Q} \sup_{\mathbf{0} \neq \boldsymbol{V} \in \mathbf{X}} \frac{b(\boldsymbol{V}, q)}{\|q\|_Q \|\boldsymbol{V}\|_{\boldsymbol{X}}} \geq \beta, \ \beta \text{ is a positive constant.} \tag{2.14}$$

For each $\boldsymbol{U_0} \in \boldsymbol{L}^2(\Omega)$ and $\boldsymbol{F} \in \boldsymbol{L}^2(0,T;\boldsymbol{X}')$, the coupled Stokes–Darcy system (2.12) admits a unique solution $(\boldsymbol{U}, p) \in \boldsymbol{L}^2(0,T;\boldsymbol{X}) \cap \boldsymbol{H}^1(0,T;\boldsymbol{X}') \times L^2(0,T;Q)$ (cf. [18]). We use formulation (2.12) to define the operator $\mathcal{F} : \boldsymbol{L}^2(0,T;\boldsymbol{X}) \cap \boldsymbol{H}^1(0,T;\boldsymbol{X}') \times L^2(0,T;Q) \times \boldsymbol{L}^2(\Omega) \to \boldsymbol{L}^2(0,T;\boldsymbol{X}') \times L^2(0,T;Q') \times \boldsymbol{L}^2(\Omega)$,

$$\mathcal{F}\begin{pmatrix} \boldsymbol{U} \\ p \\ \boldsymbol{U_0} \end{pmatrix} = \begin{pmatrix} \frac{\partial \boldsymbol{U}}{\partial t} + A\boldsymbol{U} + B^* p - \boldsymbol{F} \\ B\boldsymbol{U} \\ \boldsymbol{U}(\cdot, 0) - \boldsymbol{U_0} \end{pmatrix}.$$

With a simple calculus of variation, one can see that the Fréchet derivative operator $\mathcal{F}'$ is a bijective mapping, and the surjective of $\mathcal{F}'$ is thereafter self-contained. These basics later will allow us to use the Lagrange multiplier rule to find the optimal solution for cost functionals constrained by the Stokes–Darcy equation. In addition, throughout this paper, C, C_i, C_i^j, and $C_{i,j,k}$ are generic positive constants that are independent of the mesh parameter h and the time step τ and are not necessarily the same at each occurrence.

3. Data assimilation of the Stokes–Darcy model. Letting $\boldsymbol{Y}_{ad}$ be an admissible set for the initial value that could be either $\boldsymbol{L}^2(\Omega)$ or a closed convex subset of $\boldsymbol{L}^2(\Omega)$, we look for a solution to our data assimilation problem stated as follows: given $T > 0$, $\gamma > 0$, and the distributed observation $\widehat{\boldsymbol{U}} = (\widehat{\phi}, \widehat{\boldsymbol{u}})^T \in L^2(0,T;\boldsymbol{L}^2(\Omega))$, the VDA of the Stokes–Darcy model is

$$\min_{\boldsymbol{U_0} \in \boldsymbol{Y}_{ad}} J(\boldsymbol{U_0}) = \frac{1}{2} \int_0^T \|\widehat{\boldsymbol{U}} - \boldsymbol{U}(\boldsymbol{U_0})\|_0^2 dt + \frac{\gamma}{2} \|\boldsymbol{U_0}\|_0^2 \quad \text{subject to (2.12)}, \tag{3.1}$$

where the mapping $\boldsymbol{U}(\boldsymbol{U_0}) : \boldsymbol{L}^2(\Omega) \to \boldsymbol{W}(0,T) = \boldsymbol{L}^2(0,T;\boldsymbol{X}) \cap \boldsymbol{H}^1(0,T;\boldsymbol{X}')$ is defined as the solution of (2.12) with initial condition $\boldsymbol{U_0}$. The minimization of

$\frac{1}{2}\int_0^T \|\widehat{\boldsymbol{U}} - \boldsymbol{U}(\boldsymbol{U_0})\|_0^2 dt$ in (3.1) is the primary goal, which drives the state variable $\boldsymbol{U}(\boldsymbol{U_0})$ close to the distributed observation $\widehat{\boldsymbol{U}}$ via adjusting the initial data $\boldsymbol{U_0}$. The second term $\frac{\gamma}{2}\|\boldsymbol{U_0}\|_0^2$ is a $\boldsymbol{L}^2$-Tikhonov regularization. The parameter γ balances the relative importance of the minimization between terms $\int_0^T \|\widehat{\boldsymbol{U}} - \boldsymbol{U}(\boldsymbol{U_0})\|_0^2 dt$ and $\|\boldsymbol{U_0}\|_0^2$.

Provided that $\boldsymbol{F}, \widehat{\boldsymbol{U}} \in L^2(0,T;\boldsymbol{L}^2(\Omega))$ and $\partial\Omega$ and Γ are regular enough, we have the following existence and uniqueness result.

THEOREM 3.1. *There exists a unique solution $\boldsymbol{U_0^\star} \in \boldsymbol{Y}_{ad}$ for the data assimilation problem* (3.1). *Furthermore, the solution $\boldsymbol{U_0^\star}$ can be characterized by the variational inequality:*

$$
\begin{aligned}
J'(\boldsymbol{U_0^\star})(\boldsymbol{Z}_0 - \boldsymbol{U_0^\star}) = &\int_0^T \int_\Omega (\boldsymbol{U}(\boldsymbol{U_0^\star}) - \widehat{\boldsymbol{U}})(\boldsymbol{U}(\boldsymbol{Z}_0) - \boldsymbol{U}(\boldsymbol{U_0^\star}))dxdydt \\
&+ \gamma \int_\Omega \boldsymbol{U_0^\star}(\boldsymbol{Z}_0 - \boldsymbol{U_0^\star})dxdy \geq 0 \quad \forall \boldsymbol{Z}_0 \in \boldsymbol{Y}_{ad}.
\end{aligned} \tag{3.2}
$$

Proof. Since $J(\boldsymbol{U_0})$ is nonnegative and thus bounded from below, the infimum exists. Then let $\{\boldsymbol{U}_0^n\} \in \boldsymbol{Y}_{ad}$ be a minimizing sequence such that $J(\boldsymbol{U}_0^n) \to \inf_{\boldsymbol{U_0} \in \boldsymbol{Y}_{ad}} J(\boldsymbol{U_0})$. Note that $J(\boldsymbol{U_0}) \geq \frac{\gamma}{2}\|\boldsymbol{U_0}\|_0^2$ gives the boundedness of the sequence $\{\boldsymbol{U}_0^n\}$ in $\boldsymbol{L}^2$. By the Eberlin–Šmulian theorem, a bounded sequence in Hilbert space has a weakly convergent subsequence. Therefore, we have a subsequence $\boldsymbol{U}_0^{n^k} \to \boldsymbol{U_0^\star}$ weakly as $k \to \infty$. Since the closed and convex set in $\boldsymbol{L}^2$ is weakly closed, we have $\boldsymbol{U_0^\star} \in \boldsymbol{Y}_{ad}$. We know that the norm is a continuous functional, and the mapping $\boldsymbol{U}(\boldsymbol{U}_0) : \boldsymbol{L}^2(\Omega) \mapsto \boldsymbol{W}(0,T)$ is continuous because of the well-posedness results of Stokes–Darcy equations [18]. Hence the composite cost functional $J(\boldsymbol{U_0})$ is continuous. Combining with the convexity of $J(\boldsymbol{U_0})$ that can be seen in (3.3), we claim that $J(\boldsymbol{U_0})$ is weakly lower semicontinuous. Hence,

$$
\begin{aligned}
J(\boldsymbol{U_0^\star}) = \frac{\gamma}{2}\|\boldsymbol{U}_0^\star\|_0^2 + \frac{1}{2}\int_0^T \|\boldsymbol{U}(\boldsymbol{U_0^\star}) - \widehat{\boldsymbol{U}}\|_0^2 dt &\leq \liminf_{k\to\infty} J(\boldsymbol{U}_0^{n^k}) \\
&= \inf_{\boldsymbol{U}_0 \in \boldsymbol{Y}_{ad}} J(\boldsymbol{U}_0) \leq J(\boldsymbol{U}_0^\star),
\end{aligned}
$$

which implies that $\boldsymbol{U}_0^\star$ is a minimizer we need.

By the linear property of the Stokes–Darcy equation, one can find out that $J(\boldsymbol{U}_0)$ is Fréchet differentiable and its second order derivative can be calculated as follows:

$$
J''(\boldsymbol{U}_0)(\boldsymbol{Z}_0, \boldsymbol{Z}_0) = \int_0^T \int_\Omega \boldsymbol{U}^2(\boldsymbol{Z}_0)dxdydt + \gamma \int_\Omega \boldsymbol{Z}_0^2 dxdy \geq \gamma \|\boldsymbol{Z}_0\|_0^2 \quad \forall \boldsymbol{Z}_0 \in \boldsymbol{Y}_{ad}. \tag{3.3}
$$

Based on (3.3) and a standard argument for convex minimization we know the minimizer $\boldsymbol{U}_0^\star$ is unique. Further, a calculus of variation of $J(\boldsymbol{U}_0)$ with respect to $\boldsymbol{U}_0$ characterizes $\boldsymbol{U}_0^\star$ as

$$
\begin{aligned}
J'(\boldsymbol{U}_0^\star)(\boldsymbol{Z}_0 - \boldsymbol{U}_0^\star) = &\int_0^T \int_\Omega (\boldsymbol{U}(\boldsymbol{U}_0^\star) - \widehat{\boldsymbol{U}})(\boldsymbol{U}(\boldsymbol{Z}_0) - \boldsymbol{U}(\boldsymbol{U}_0^\star))dxdydt \\
&+ \gamma \int_\Omega \boldsymbol{U}_0^\star(\boldsymbol{Z}_0 - \boldsymbol{U}_0^\star)dxdy \geq 0 \, \forall \boldsymbol{Z}_0 \in \boldsymbol{Y}_{ad}.
\end{aligned}
$$

This finishes the proof. ∎

Next, we show that the solution of problem (3.1) is stable regarding the perturbation of the observational data $\widehat{\boldsymbol{U}}$ and the regularization parameter γ.

THEOREM 3.2. *The solution of problem* (3.1) *continuously depends on the observational data* $\widehat{\boldsymbol{U}}$ *and the parameter* γ.

Proof. Introducing perturbations $\epsilon_1 \in \mathbb{R}$ on γ and $\boldsymbol{\epsilon}_2 \in L^2(0,T;\boldsymbol{L}^2(\Omega))$ on $\widehat{\boldsymbol{U}}$, respectively, and letting $\bar{\boldsymbol{U}}_0$ denote the perturbed optimal solution, we have

$$\begin{aligned}(3.4)\qquad &\int_0^T\int_\Omega (\boldsymbol{U}(\bar{\boldsymbol{U}}_0)-\widehat{\boldsymbol{U}}-\boldsymbol{\epsilon}_2)(\boldsymbol{U}(\boldsymbol{Z}_0)-\boldsymbol{U}(\bar{\boldsymbol{U}}_0))dxdydt\\ &\quad+(\gamma+\epsilon_1)\int_\Omega \bar{\boldsymbol{U}}_0(\boldsymbol{Z}_0-\bar{\boldsymbol{U}}_0)dxdy\geq 0 \qquad \forall \boldsymbol{Z}_0\in \boldsymbol{Y}_{ad}.\end{aligned}$$

Taking $\boldsymbol{Z}_0=\boldsymbol{U}_0^\star$ in (3.4) and $\boldsymbol{Z}_0=\bar{\boldsymbol{U}}_0$ in (3.2), we obtain

$$\int_0^T\int_\Omega (\boldsymbol{U}(\bar{\boldsymbol{U}}_0)-\widehat{\boldsymbol{U}}-\boldsymbol{\epsilon}_2)(\boldsymbol{U}(\boldsymbol{U}_0^\star)-\boldsymbol{U}(\bar{\boldsymbol{U}}_0))dxdydt+(\gamma+\epsilon_1)\int_\Omega \bar{\boldsymbol{U}}_0(\boldsymbol{U}_0^\star-\bar{\boldsymbol{U}}_0)dxdy\geq 0,$$

$$\int_0^T\int_\Omega (\boldsymbol{U}(\boldsymbol{U}_0^\star)-\widehat{\boldsymbol{U}})(\boldsymbol{U}(\bar{\boldsymbol{U}}_0)-\boldsymbol{U}(\boldsymbol{U}_0^\star))dxdydt+\gamma\int_\Omega \boldsymbol{U}_0^\star(\bar{\boldsymbol{U}}_0-\boldsymbol{U}_0^\star)dxdy\geq 0.$$

Adding the two inequalities together, we have

$$\begin{aligned}(3.5)\qquad &\int_0^T\int_\Omega (\boldsymbol{U}(\boldsymbol{U}_0^\star)-\boldsymbol{U}(\bar{\boldsymbol{U}}_0))^2dxdydt+(\gamma+\epsilon_1)\int_\Omega(\boldsymbol{U}_0^\star-\bar{\boldsymbol{U}}_0)^2dxdy\\ &\leq\int_0^T\int_\Omega \boldsymbol{\epsilon}_2(\boldsymbol{U}(\bar{\boldsymbol{U}}_0)-\boldsymbol{U}(\boldsymbol{U}_0^\star))dxdydt+\epsilon_1\int_\Omega \boldsymbol{U}_0^\star(\boldsymbol{U}_0^\star-\bar{\boldsymbol{U}}_0)dxdy.\end{aligned}$$

Applying the Cauchy–Schwarz and Young's inequalities for the right-hand-side terms in (3.5), we have

$$\begin{aligned}(3.6)\qquad \int_0^T\int_\Omega \epsilon_2(\boldsymbol{U}(\bar{\boldsymbol{U}}_0)-\boldsymbol{U}(\boldsymbol{U}_0^\star))dxdydt&\leq\frac{1}{2}\int_0^T\int_\Omega(\boldsymbol{U}(\boldsymbol{U}_0^\star)-\boldsymbol{U}(\bar{\boldsymbol{U}}_0))^2dxdydt\\ &\quad+\frac{1}{2}\|\boldsymbol{\epsilon}_2\|^2_{L^2(0,T;\boldsymbol{L}^2(\Omega))},\end{aligned}$$

$$(3.7)\qquad \epsilon_1\int_\Omega \boldsymbol{U}_0(\boldsymbol{U}_0^\star-\bar{\boldsymbol{U}}_0)dxdy\leq\frac{|\epsilon_1|}{2}\|\boldsymbol{U}_0^\star\|_0^2+\frac{|\epsilon_1|}{2}\int_\Omega(\boldsymbol{U}_0^\star-\bar{\boldsymbol{U}}_0)^2dxdy.$$

Combining (3.5)–(3.7) and setting $|\epsilon_1|\leq\frac{\gamma}{3}$, we obtain the inequality

$$\begin{aligned}(3.8)\qquad &\frac{1}{2}\int_0^T\int_\Omega(\boldsymbol{U}(\boldsymbol{U}_0^\star)-\boldsymbol{U}(\bar{\boldsymbol{U}}_0))^2dxdydt\\ &\quad+\frac{\gamma}{2}\int_\Omega(\boldsymbol{U}_0^\star-\bar{\boldsymbol{U}}_0)^2dxdy\leq\frac{1}{2}\|\boldsymbol{\epsilon}_2\|^2_{L^2(0,T;\boldsymbol{L}^2(\Omega))}+\frac{|\epsilon_1|}{2}\|\boldsymbol{U}_0^\star\|_0^2,\end{aligned}$$

which implies that the solution of problem (3.1) continuously depends on the observational data $\widehat{\boldsymbol{U}}$ and the regularization parameter γ. ∎

Remark 3.3. Continuing on (3.5), another treatment of the term $\int_0^T\int_\Omega \boldsymbol{\epsilon}_2(\boldsymbol{U}(\bar{\boldsymbol{U}}_0)-\boldsymbol{U}(\boldsymbol{U}_0^\star))dxdydt$ in (3.5) will produce a different stability estimation:

$$(3.9)\qquad \frac{\gamma}{2}\int_\Omega(\boldsymbol{U}_0^\star-\bar{\boldsymbol{U}}_0)^2dxdy\leq\frac{1}{4}\|\boldsymbol{\epsilon}_2\|^2_{L^2(0,T;\boldsymbol{L}^2(\Omega))}+\frac{|\epsilon_1|}{2}\|\boldsymbol{U}_0^\star\|_0^2.$$

Inequality (3.9) provides perspective on how parameter γ affects the stability of solution, i.e., small γ will cause ill-conditioning of the data assimilation system.

To find out the unique optimal solution that minimizes the objective functional in (3.1), we apply the Lagrange multiplier rule which is apparently available due to the property of the operator $\mathcal{F}$ (surjective of $\mathcal{F}'$) shown in section 2. The Lagrange functional is formed as

$$\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{U}, p, \boldsymbol{U_0}) = \frac{1}{2}\int_0^T \|\widehat{\boldsymbol{U}} - \boldsymbol{U}\|_0^2 dt + \frac{\gamma}{2}\|\boldsymbol{U_0}\|_0^2 + \langle \boldsymbol{\lambda}, \mathcal{F}(\boldsymbol{U}, p, \boldsymbol{U_0})^T\rangle, \tag{3.10}$$

where $\boldsymbol{\lambda} \in \boldsymbol{L}^2(0,T;\boldsymbol{X}'') \cap \boldsymbol{H}^1(0,T;\boldsymbol{X}') \times L^2(0,T;Q'') \times \boldsymbol{L}^2(\Omega)'$ is a Lagrange multiplier. Since Hilbert space is reflexive, $\boldsymbol{X} \times Q$ and $\boldsymbol{X}'' \times Q''$ are therefore isometric. The element in dual space of a Hilbert space can be identified by the element in the Hilbert space itself. Hence, using the definition of operator $\mathcal{F}$, (3.10) then can be rewritten as

(3.11)

$$\begin{aligned}
\mathcal{L}(\boldsymbol{U^*}, p^*, \boldsymbol{U^*}(\cdot,0), \boldsymbol{U}, p, \boldsymbol{U_0}) &= \frac{1}{2}\int_0^T \|\widehat{\boldsymbol{U}} - \boldsymbol{U}\|_0^2 dt \\
&\quad + \frac{\gamma}{2}\|\boldsymbol{U_0}\|_0^2 + \int_0^T \left\langle \frac{\partial \boldsymbol{U}}{\partial t} + A\boldsymbol{U} + B^* p - \boldsymbol{F}, \boldsymbol{U^*} \right\rangle dt \\
&\quad + \int_0^T \langle B\boldsymbol{U}, p^*\rangle dt + (\boldsymbol{U}(\cdot,0) - \boldsymbol{U}_0, \boldsymbol{U^*}(\cdot,0)) \\
&= \frac{1}{2}\int_0^T \|\widehat{\boldsymbol{U}} - \boldsymbol{U}\|_0^2 dt + \frac{\gamma}{2}\|\boldsymbol{U_0}\|_0^2 + \int_0^T \left\langle \frac{\partial \boldsymbol{U}}{\partial t}, \boldsymbol{U^*} \right\rangle dt \\
&\quad + \int_0^T a(\boldsymbol{U}, \boldsymbol{U^*})\, dt + \int_0^T b(\boldsymbol{U^*}, p)\, dt \\
&\quad + \int_0^T b(\boldsymbol{U}, p^*)\, dt + (\boldsymbol{U}(\cdot,0) - \boldsymbol{U_0}, \boldsymbol{U^*}(\cdot,0)) \\
&\quad - \int_0^T \langle \boldsymbol{F}, \boldsymbol{U^*}\rangle dt.
\end{aligned}$$

Variations in the Lagrange multipliers $\boldsymbol{U^*}$, p^*, and $\boldsymbol{U^*}(\cdot,0)$ recover the constraint equation (2.12). Variations with respect to $\boldsymbol{U}$, p, and $\boldsymbol{U_0}$ yield

$$\begin{aligned}
&\int_0^T (\widehat{\boldsymbol{U}} - \boldsymbol{U}, -\boldsymbol{V}) dt + \int_0^T \left\langle \frac{\partial \boldsymbol{V}}{\partial t}, \boldsymbol{U^*} \right\rangle dt + \int_0^T a(\boldsymbol{V}, \boldsymbol{U^*})\, dt + \int_0^T b(\boldsymbol{V}, p^*)\, dt \\
&\qquad + (\boldsymbol{V}(\cdot,0), \boldsymbol{U^*}(\cdot,0)) = 0 \quad \forall \boldsymbol{V} \in \boldsymbol{L}^2(0,T;\boldsymbol{X}) \times \boldsymbol{H}^1(0,T;\boldsymbol{X}'), \\
&\int_0^T b(\boldsymbol{U^*}, q)\, dt = 0 \quad \forall q \in L^2(0,T;Q), \\
&\gamma(\boldsymbol{U_0}, \boldsymbol{Z}_0) - (\boldsymbol{Z}_0, \boldsymbol{U^*}(\cdot,0)) = 0 \quad \forall \boldsymbol{Z}_0 \in \boldsymbol{L}^2(\Omega).
\end{aligned} \tag{3.12}$$

Taking integration by parts with respect to time for $\int_0^T \langle \frac{\partial \boldsymbol{V}}{\partial t}, \boldsymbol{U^*}\rangle dt$ in the first equation of (3.12), we obtain

$$\begin{aligned}
&\int_0^T (\widehat{\boldsymbol{U}} - \boldsymbol{U}, -\boldsymbol{V}) dt + (\boldsymbol{V}, \boldsymbol{U^*})|_0^T - \int_0^T \left\langle \frac{\partial \boldsymbol{U^*}}{\partial t}, \boldsymbol{V} \right\rangle dt + \int_0^T a(\boldsymbol{V}, \boldsymbol{U^*})\, dt \\
&\qquad + \int_0^T b(\boldsymbol{V}, p^*)\, dt + (\boldsymbol{V}(\cdot,0), \boldsymbol{U^*}(\cdot,0)) = 0.
\end{aligned} \tag{3.13}$$

Choosing $\boldsymbol{U}^*(\cdot,T)=\mathbf{0}$ and simplifying (3.13), we have

$$\begin{aligned}\boldsymbol{U}^*(\cdot,T)=\mathbf{0},\ &\int_0^T(\widehat{\boldsymbol{U}}-\boldsymbol{U},-\boldsymbol{V})dt-\int_0^T\left\langle\frac{\partial\boldsymbol{U}^*}{\partial t},\boldsymbol{V}\right\rangle dt\\&+\int_0^T a^*\left(\boldsymbol{U}^*,\boldsymbol{V}\right)dt+\int_0^T b\left(\boldsymbol{V},p^*\right)dt=0,\end{aligned}\tag{3.14}$$

where $a^*(\boldsymbol{U}^*,\boldsymbol{V})$ is given as

$$\begin{aligned}a^*\left(\boldsymbol{U}^*,\boldsymbol{V}\right)=&2\nu\left(\mathbb{D}\left(\boldsymbol{u}^*\right),\mathbb{D}\left(\boldsymbol{v}\right)\right)_{\Omega_f}+\left(K\nabla\phi^*,\nabla\psi\right)_{\Omega_p}+\langle g\boldsymbol{u}^*\cdot\boldsymbol{n_f},\psi\rangle\\&-\langle\phi^*,\boldsymbol{v}\cdot\boldsymbol{n}_f\rangle+\alpha\langle P_\tau\boldsymbol{u}^*,P_\tau\boldsymbol{v}\rangle+\alpha\langle P_\tau\boldsymbol{u}^*,P_\tau\left(\mathbb{K}\nabla\psi\right)\rangle.\end{aligned}\tag{3.15}$$

Note that (3.15) is essentially a consequence of swapping terms related to $\boldsymbol{V}$ and $\boldsymbol{U}^*$ of $a(\boldsymbol{V},\boldsymbol{U}^*)$; this is because $\langle A\boldsymbol{V},\boldsymbol{U}^*\rangle=a(\boldsymbol{V},\boldsymbol{U}^*)=a^*(\boldsymbol{U}^*,\boldsymbol{V})=\langle A^*\boldsymbol{U}^*,\boldsymbol{V}\rangle$.

Summarizing all operations from (3.10)–(3.15) and using the definition of dualities and bilinear forms in (2.4)–(2.8), the optimal solution $\boldsymbol{U}_0$ is attained by solving the following coupled equation systems in a weak form:

the forward state equation

$$\begin{cases}\left\langle\dfrac{\partial\phi}{\partial t},\psi\right\rangle+a_p\left(\phi,\psi\right)-\langle\boldsymbol{u}\cdot\boldsymbol{n_f},\psi\rangle=\langle f_p,\psi\rangle\quad\forall\psi\in X_p,\\ \left\langle\dfrac{\partial\boldsymbol{u}}{\partial t},\boldsymbol{v}\right\rangle+a_f\left(\boldsymbol{u},\boldsymbol{v}\right)+b_f\left(\boldsymbol{v},p\right)+\langle g\phi,\boldsymbol{v}\cdot\boldsymbol{n}_f\rangle+\alpha\langle P_\tau\left(\boldsymbol{u}+\mathbb{K}\nabla\phi\right),P_\tau\boldsymbol{v}\rangle\\ \qquad=\langle\boldsymbol{f}_f,\boldsymbol{v}\rangle\quad\forall\boldsymbol{v}\in\boldsymbol{X}_f,\\ b_f\left(\boldsymbol{u},q\right)=0\quad\forall q\in Q,\\ \phi(\cdot,0)=\phi_0\quad\boldsymbol{\phi}_0\in L^2(\Omega_p),\quad\boldsymbol{u}(\cdot,0)=\boldsymbol{u_0}\quad\boldsymbol{u}_0\in\boldsymbol{L}^2(\Omega_f),\end{cases}\tag{3.16}$$

the backward adjoint equation

$$\begin{cases}-\left\langle\dfrac{\partial\phi^*}{\partial t},\psi\right\rangle+a_p\left(\phi^*,\psi\right)+\langle g\boldsymbol{u}^*\cdot\boldsymbol{n_f},\psi\rangle+\alpha\langle P_\tau\boldsymbol{u}^*,P_\tau\left(\mathbb{K}\nabla\psi\right)\rangle\\ \qquad=(\hat{\phi}-\phi,\psi)\quad\forall\psi\in X_p,\\ -\left\langle\dfrac{\partial\boldsymbol{u}^*}{\partial t},\boldsymbol{v}\right\rangle+a_f\left(\boldsymbol{u}^*,\boldsymbol{v}\right)+b_f\left(\boldsymbol{v},p^*\right)-\langle\phi^*,\boldsymbol{v}\cdot\boldsymbol{n}_f\rangle+\alpha\langle P_\tau\boldsymbol{u}^*,P_\tau\boldsymbol{v}\rangle\\ \qquad=(\hat{\boldsymbol{u}}-\boldsymbol{u},\boldsymbol{v})\quad\forall\boldsymbol{v}\in\boldsymbol{X}_f,\\ b_f\left(\boldsymbol{u}^*,q\right)=0\quad\forall q\in Q,\\ \phi^*(\cdot,T)=0,\quad\boldsymbol{u}^*(\cdot,T)=\mathbf{0},\end{cases}\tag{3.17}$$

and

$$\phi_0=\frac{1}{\gamma}\phi^*(\cdot,0),\quad\boldsymbol{u_0}=\frac{1}{\gamma}\boldsymbol{u}^*(\cdot,0).\tag{3.18}$$

The coupled system (3.16)–(3.18) is the first order necessary optimality system. The minimization problem (3.1) is strictly convex, thus the first order necessary condition is also sufficient.

4. Numerical approximation. In this section, we propose a fully discrete approximation of the data assimilation problem (3.1), which is based on a finite element discretization in space and the backward Euler scheme in time.

For spatial discretization, we consider $\boldsymbol{X}^h = X_p^h \times \boldsymbol{X}_f^h$ and Q^h being pairwise well-defined finite element subspaces of $\boldsymbol{X} = X_p \times \boldsymbol{X}_f$ and Q, respectively. These families of spaces are parameterized by the mesh size h that tends to 0, and we assume these finite element spaces satisfy the inf-sup condition, i.e., there exists a positive constant β such that

$$\inf_{0\neq q\in Q^h} \sup_{\mathbf{0}\neq \boldsymbol{V}\in \boldsymbol{X}^{\mathbf{h}}} \frac{b(\boldsymbol{V},q)}{\|q\|_Q \|\boldsymbol{V}\|_{\boldsymbol{X}}} \geq \beta. \tag{4.1}$$

For the time discretization we uniformly construct a time grid $0 = t_0 < t_1 < \cdots < t_n < \cdots < t_N = T$ with time step $\tau = \frac{T}{N}$. Letting $I_n = (t_{n-1}, t_n]$ denote the nth subinterval, we use the finite-dimensional space

$$\boldsymbol{X}_{\tau,h} = \{\boldsymbol{V} : [0,T] \to \boldsymbol{X}^h : \boldsymbol{V}|_{I_n} \in \boldsymbol{X}^h \quad \text{is constant in time}\}.$$

Let $\boldsymbol{V}_h^n$ be the value of $\boldsymbol{V}_h \in \boldsymbol{X}_{\tau,h}$ at t_n and $\boldsymbol{X}_{\tau,h}^n$ be the restriction to I_n of the functions in $\boldsymbol{X}_{\tau,h}$.

Given specific $h, \tau, \gamma > 0$ and an admissible set $\boldsymbol{Y}_{ad}^h = \boldsymbol{X}^h \cap \boldsymbol{Y}_{ad}$ for the possible initial values, the fully discrete approximation of problem (3.1) is stated as

$$\min_{\boldsymbol{U}_{0,h}\in \boldsymbol{Y}_{ad}^h} J_h(\boldsymbol{U}_{0,h}) = \frac{1}{2}\tau \sum_{n=1}^{N} \|\widehat{\boldsymbol{U}}^n - \boldsymbol{U}_h^n\|_0^2 + \frac{\gamma}{2}\|\boldsymbol{U}_{0,h}\|_0^2 \tag{4.2}$$

subject to

$$\begin{cases} \dfrac{\boldsymbol{U}_h^{n+1} - \boldsymbol{U}_h^n}{\tau} + A\boldsymbol{U}_h^{n+1} + B^* p_h^{n+1} = \boldsymbol{F}^{n+1} \quad \text{in} \quad (\boldsymbol{X}^h)', \\ B\boldsymbol{U}_h^{n+1} = 0 \quad \text{in} \quad (Q^h)', \\ \boldsymbol{U}_h^0 = \boldsymbol{U}_{0,h} \quad \text{in} \quad \boldsymbol{X}^h. \end{cases} \tag{4.3}$$

Similar to the proof for the well-posedness of the continuous data assimilation problem (3.1), one can prove the well-posedness of the fully discrete data assimilation problem (4.2)–(4.3).

THEOREM 4.1. *Given $\tau = \frac{T}{N}$ and mesh size h, for every fixed regularization parameter γ, there exists a unique optimal solution $\boldsymbol{U}_{0,h}^\star \in \boldsymbol{Y}_{ad}^h$ such that the cost functional* (4.2) *is minimized. The optimal solution continuously depends on the observation data $\widehat{\boldsymbol{U}}$ and the parameter γ.*

Furthermore, one can also observe that small γ will reduce the stability of the discrete data assimilation problem.

We expect that the optimal discrete solution of (4.2)–(4.3) converges to the solution of (3.1). That is, given a fixed γ, $\boldsymbol{U}_{0,h}^\star \to \boldsymbol{U}_0^\star$ should be attained when the time step τ and finite element mesh size h tend to 0.

THEOREM 4.2. *For a fixed regularization parameter γ, let $\{\boldsymbol{U}_{0,h}^\star\}_{h,\tau>0}$ be the corresponding sequence of minimizers of the discrete data assimilation problems* (4.2)–(4.3). *Then $\{\boldsymbol{U}_{0,h}^\star\}_{h,\tau>0}$ converges to the continuous optimal solution $\boldsymbol{U}_0^\star$ as $h \to 0$ and $\tau \to 0$.*

Proof. The proof is a general extension from results in [59, 87]. It is not difficult to see $J_h(\boldsymbol{U}_{0,h}^\star) \leq C$ for some constant C independent of τ and h. Then the coercevity of $J_h(\boldsymbol{U}_{0,h}^\star)$ implies the boundedness of $\{\boldsymbol{U}_{0,h}^\star\}$ in $\boldsymbol{L}^2(\Omega)$. Hence we can extract a

subsequence $\{\boldsymbol{U}^\star_{0,h'}\}$ from $\{\boldsymbol{U}^\star_{0,h}\}$ such that $\{\boldsymbol{U}^\star_{0,h'}\}$ converges weakly to $\boldsymbol{\mu}^\star$ in $\boldsymbol{L}^2(\Omega)$. Recalling the Lemma 4.4 in [87], we have

$$\lim_{h',\tau\to 0}\frac{1}{2}\tau\sum_{n=1}^{N}\|\widehat{\boldsymbol{U}}^n-\boldsymbol{U}^n_{h'}(\boldsymbol{U}^\star_{0,h'})\|_0^2\to\frac{1}{2}\int_0^T\|\widehat{\boldsymbol{U}}-\boldsymbol{U}(\boldsymbol{\mu}^\star)\|_0^2dt.$$

Thus, for any $\boldsymbol{V}\in\boldsymbol{Y}_{ad}$, by the weakly lower semicontinuity

$$\begin{aligned}J(\boldsymbol{\mu}^\star)&\le\liminf_{h',\tau\to 0}\frac{1}{2}\tau\sum_{n=1}^{N}\|\widehat{\boldsymbol{U}}^n-\boldsymbol{U}^n_{h'}(\boldsymbol{U}^\star_{0,h'})\|_0^2+\frac{\gamma}{2}\liminf_{h',\tau\to 0}\|\boldsymbol{U}^\star_{0,h'}\|_0^2\\&\le\liminf_{h',\tau\to 0}J_{h'}(\boldsymbol{U}^\star_{0,h'})\le\liminf_{h',\tau\to 0}J_{h'}(\boldsymbol{\pi}_{h'}(\boldsymbol{V}))\\&=\frac{1}{2}\int_0^T\|\widehat{\boldsymbol{U}}-\boldsymbol{U}(\boldsymbol{V})\|_0^2dt+\frac{\gamma}{2}\|\boldsymbol{V}\|_0^2=J(\boldsymbol{V}),\end{aligned}\tag{4.4}$$

where $\boldsymbol{\pi}_h$ is the $\boldsymbol{L}^2$ projection operator from $\boldsymbol{Y}_{ad}$ to $\boldsymbol{X}^h$. Then (4.4) and the uniqueness result in Theorem 3.1 imply $\boldsymbol{\mu}^\star$ is the optimal solution of the problem (3.1) and thus the theorem is proved. ∎

In order to derive the discrete optimality system and solve for the optimal solution, we formulate the discrete Lagrange functional:

$$\begin{aligned}&\mathcal{L}(\bar{\boldsymbol{U}}_h,\bar{p}_h,\boldsymbol{U}_{0,h},\bar{\boldsymbol{U}}^*_h,\bar{p}^*_h,\boldsymbol{U}^{*0}_h)\\&\quad=\frac{1}{2}\tau\sum_{n=1}^{N}\|\widehat{\boldsymbol{U}}^n-\boldsymbol{U}^n_h\|_0^2+\frac{\gamma}{2}\|\boldsymbol{U}_{0,h}\|_0^2+\tau\sum_{n=0}^{N-1}\langle B\boldsymbol{U}^{n+1}_h,p^{*n}_h\rangle\\&\quad\quad+\tau\sum_{n=0}^{N-1}\left\langle\frac{\boldsymbol{U}^{n+1}_h-\boldsymbol{U}^n_h}{\tau}+A\boldsymbol{U}^{n+1}_h+B^*p^{n+1}_h-\boldsymbol{F}_{n+1},\boldsymbol{U}^{*n}_h\right\rangle\\&\quad\quad+(\boldsymbol{U}^0_h-\boldsymbol{U}_{0,h},\boldsymbol{U}^{*0}_h),\end{aligned}\tag{4.5}$$

where $\bar{\boldsymbol{U}}_h=(\boldsymbol{U}^0_h,\boldsymbol{U}^1_h,\boldsymbol{U}^2_h,\ldots,\boldsymbol{U}^N_h)$, $\bar{\boldsymbol{U}}^*_h=(\boldsymbol{U}^{*1}_h,\boldsymbol{U}^{*2}_h,\ldots,\boldsymbol{U}^{*N-1}_h)$, $\bar{p}_h=(p^1_h,p^2_h,p^3_h,\ldots,p^N_h)$, and $\bar{p}^*_h=(p^{*0}_h,p^{*1}_h,p^{*2}_h,p^{*3}_h,\ldots,p^{*N-1}_h)$. By a few manipulations on $\boldsymbol{U}^n_h$, $\boldsymbol{U}^{*n}_h$ and using the adjoint notation $a(\boldsymbol{U},\boldsymbol{V})=\langle A\boldsymbol{U},\boldsymbol{V}\rangle=\langle\boldsymbol{U},A^*\boldsymbol{V}\rangle=a^*(\boldsymbol{V},\boldsymbol{U})$, we reorganize (4.5) as

$$\begin{aligned}&\mathcal{L}(\bar{\boldsymbol{U}}_h,\bar{p}_h,\boldsymbol{U}_{0,h},\bar{\boldsymbol{U}}^*_h,\bar{p}^*_h,\boldsymbol{U}^{*0}_h)\\&\quad=\frac{1}{2}\tau\sum_{n=1}^{N}\|\hat{\boldsymbol{U}}^n-\boldsymbol{U}^n_h\|_0^2+\frac{\gamma}{2}\|\boldsymbol{U}_{0,h}\|_0^2+\tau\sum_{n=0}^{N-1}\langle B\boldsymbol{U}^{n+1}_h,p^{*n}_h\rangle+\tau\sum_{n=0}^{N-1}\langle B^*p^{n+1}_h,\boldsymbol{U}^{*n}_h\rangle\\&\quad\quad+\tau\sum_{n=0}^{N-1}\langle A\boldsymbol{U}^{n+1}_h,\boldsymbol{U}^{*n}_h\rangle-\tau\sum_{n=0}^{N-1}\langle\boldsymbol{F}_{n+1},\boldsymbol{U}^{*n}_h\rangle+\tau\sum_{n=0}^{N-1}\left\langle\frac{\boldsymbol{U}^{n+1}_h-\boldsymbol{U}^n_h}{\tau},\boldsymbol{U}^{*n}_h\right\rangle\\&\quad\quad+(\boldsymbol{U}^N_h,\boldsymbol{U}^{*N}_h)-(\boldsymbol{U}^N_h,\boldsymbol{U}^{*N}_h)+(\boldsymbol{U}^0_h-\boldsymbol{U}_{0,h},\boldsymbol{U}^{*0}_h)\\&\quad=\frac{1}{2}\tau\sum_{n=1}^{N}\|\hat{\boldsymbol{U}}^n-\boldsymbol{U}^n_h\|_0^2+\frac{\gamma}{2}\|\boldsymbol{U}_{0,h}\|_0^2+\tau\sum_{n=1}^{N}\langle B\boldsymbol{U}^n_h,p^{*n-1}_h\rangle+\tau\sum_{n=1}^{N}\langle B^*p^n_h,\boldsymbol{U}^{*n-1}_h\rangle\\&\quad\quad+\tau\sum_{n=1}^{N}\langle A^*\boldsymbol{U}^{*n-1}_h,\boldsymbol{U}^n_h\rangle-\tau\sum_{n=1}^{N}\langle\boldsymbol{F}_n,\boldsymbol{U}^{*n-1}_h\rangle+\tau\sum_{n=1}^{N}\left\langle\frac{\boldsymbol{U}^{*n-1}_h-\boldsymbol{U}^{*n}_h}{\tau},\boldsymbol{U}^n_h\right\rangle\\&\quad\quad+(\boldsymbol{U}^N_h,\boldsymbol{U}^{*N}_h)-(\boldsymbol{U}_{0,h},\boldsymbol{U}^{*0}_h).\end{aligned}\tag{4.6}$$

Variations in the Lagrange multipliers $\bar{\boldsymbol{U}}_h^*, \bar{p}_h^*$, and $\boldsymbol{U}_h^{*0}$ recover the constraint equation (4.3). Variations with respect to $\boldsymbol{U}_{0,h}$, $\boldsymbol{U}_h^n$, and p_h^n yield

$$\frac{\partial \mathcal{L}(\bar{\boldsymbol{U}}_h, \bar{p}_h, \boldsymbol{U}_{0,h}, \bar{\boldsymbol{U}}_h^*, \bar{p}_h^*, \boldsymbol{U}_h^{*0})}{\partial \boldsymbol{U}_{0,h}} \boldsymbol{Z}_0^h = (\gamma \boldsymbol{U}_{0,h}, \boldsymbol{Z}_0^h) - (\boldsymbol{U}_h^{*0}, \boldsymbol{Z}_0^h) = 0 \quad \forall \boldsymbol{Z}_0^h \in \boldsymbol{Y}_{ad}^h, \tag{4.7}$$

$$\begin{aligned} \frac{\partial \mathcal{L}(\bar{\boldsymbol{U}}_h, \bar{p}_h, \boldsymbol{U}_{0,h}, \bar{\boldsymbol{U}}_h^*, \bar{p}_h^*, \boldsymbol{U}_h^{*0})}{\partial \boldsymbol{U}_h^n} \boldsymbol{V}_h &= \tau \left\langle \frac{\boldsymbol{U}_h^{*n-1} - \boldsymbol{U}_h^{*n}}{\tau}, \boldsymbol{V}_h \right\rangle + \tau \langle A^* \boldsymbol{U}_h^{*n-1}, \boldsymbol{V}_h \rangle \\ &\quad + \tau \langle B \boldsymbol{V}_h, p_h^{*n-1} \rangle - \tau (\widehat{\boldsymbol{U}}^n - \boldsymbol{U}_h^n, \boldsymbol{V}_h) \\ &= 0 \qquad \forall \boldsymbol{V}_h \in \boldsymbol{X}^h, \ n = 1, \ldots, N-1, \end{aligned} \tag{4.8}$$

$$\begin{aligned} \frac{\partial \mathcal{L}(\bar{\boldsymbol{U}}_h, \bar{p}_h, \boldsymbol{U}_{0,h}, \bar{\boldsymbol{U}}_h^*, \bar{p}_h^*, \boldsymbol{U}_h^{*0})}{\partial \boldsymbol{U}_h^N} \boldsymbol{V}_h &= \tau \left\langle \frac{\boldsymbol{U}_h^{*N-1}}{\tau}, \boldsymbol{V}_h \right\rangle + \tau \langle A^* \boldsymbol{U}_h^{*N-1}, \boldsymbol{V}_h \rangle \\ &\quad + \tau \langle B \boldsymbol{V}_h, p_h^{*N-1} \rangle - \tau (\hat{\boldsymbol{U}}^N - \boldsymbol{U}_h^N, \boldsymbol{V}_h) \\ &= 0 \qquad \forall \boldsymbol{V}_h \in \boldsymbol{X}^h, \end{aligned} \tag{4.9}$$

$$\frac{\partial \mathcal{L}(\bar{\boldsymbol{U}}_h, \bar{p}_h, \boldsymbol{U}_{0,h}, \bar{\boldsymbol{U}}_h^*, \bar{p}_h^*, \boldsymbol{U}_h^{*0})}{\partial p_h^n} q_h = \langle B^* q_h, \boldsymbol{U}_h^{*n-1} \rangle = 0 \ \forall q_h \in Q^h, \ n = 1, \ldots, N. \tag{4.10}$$

Using (4.7)–(4.10) and the fact $\langle B \boldsymbol{V}_h, p_h^{*n-1} \rangle = \langle B^* p_h^{*n-1}, \boldsymbol{V}_h \rangle$, we obtain the discrete optimality system, $n = 0, \ldots, N-1$,

$$\begin{cases} \dfrac{\boldsymbol{U}_h^{n+1} - \boldsymbol{U}_h^n}{\tau} + A \boldsymbol{U}_h^{n+1} + B^* p_h^{n+1} = \boldsymbol{F}^{n+1}, \ B \boldsymbol{U}_h^{n+1} = 0, \ \boldsymbol{U}_h^0 = \boldsymbol{U}_{0,h}, \\ -\dfrac{\boldsymbol{U}_h^{*n+1} - \boldsymbol{U}_h^{*n}}{\tau} + A^* \boldsymbol{U}_h^{*n} + B^* p_h^{*n} = \hat{\boldsymbol{U}}^{n+1} - \boldsymbol{U}_h^{n+1}, \ B \boldsymbol{U}_h^{*n} = 0, \ \boldsymbol{U}_h^{*N} = 0, \\ \boldsymbol{U}_{0,h} = \dfrac{1}{\gamma} \boldsymbol{U}_h^{*0}. \end{cases} \tag{4.11}$$

In addition to the general convergence result in Theorem 4.2, one may be more interested in how the convergence behaves in practical simulations since it will help us properly set up discretization parameters for different scenarios. In the rest of this section, we focus on proving that under enough smoothness assumptions on $\boldsymbol{U}_0, \boldsymbol{U}$, and $\boldsymbol{U}^*$, the optimal finite element convergence rate is preserved for each of them.

Recall that the discrete optimality system (4.11) coincides with the direct full discretization of (3.16)–(3.18) in the sense of the operator form. (4.11) thereby shares lots of similarities with the discretization of classical PDEs except for a few special terms. Therefore, instead of directly investigating the error equation between (4.11) and (3.16)–(3.18), we can utilize the FEM results from classical PDEs to study the convergence behavior in the data assimilation problem.

Before doing so, we need to rescale (3.16)–(3.18) such that the rescaled formulations possess crucial features for our analysis. The rescaling is achieved by multiplying the second equation in (3.16) and the first equation in (3.18) with η, respectively, the corresponding rescaled bilinear forms are as follows:

$$\begin{aligned} a_\eta(\boldsymbol{U}, \boldsymbol{V}) &= \eta a_f(\boldsymbol{u}, \boldsymbol{v}) + a_p(\phi, \psi) + \eta \langle g\phi, \boldsymbol{v} \cdot \boldsymbol{n}_{\boldsymbol{f}} \rangle - \langle \boldsymbol{u} \cdot \boldsymbol{n}_{\boldsymbol{f}}, \psi \rangle \\ &\quad + \eta \alpha \langle P_\tau(\boldsymbol{u} + \mathbb{K} \nabla \phi), P_\tau \boldsymbol{v} \rangle, \\ a_\eta^*(\boldsymbol{U}^*, \boldsymbol{V}) &= a_f(\boldsymbol{u}^*, \boldsymbol{v}) + \eta a_p(\phi^*, \psi) + \eta \langle g \boldsymbol{u}^* \cdot \boldsymbol{n}_{\boldsymbol{f}}, \psi \rangle - \langle \phi^*, \boldsymbol{v} \cdot \boldsymbol{n}_f \rangle \\ &\quad + \alpha \langle P_\tau \boldsymbol{u}^*, P_\tau \boldsymbol{v} \rangle + \eta \alpha \langle P_\tau \boldsymbol{u}^*, P_\tau(\mathbb{K} \nabla \psi) \rangle. \end{aligned}$$

As stated in the following lemma, both $a_\eta(\boldsymbol{U},\boldsymbol{V})$ and $a_\eta^*(\boldsymbol{U^*},\boldsymbol{V})$ are coercive in the sense of a Gårding type inequality, and this property will be frequently used in the convergence analysis.

LEMMA 4.3. *For appropriately chosen positive rescaling parameter η, there exist constants $C_{1,\eta}$, $C_{2,\eta}$, $C_{3,\eta}$, and $C_{4,\eta}$ such that $a_\eta(\boldsymbol{U},\boldsymbol{V})$ and $a_\eta^*(\boldsymbol{U^*},\boldsymbol{V})$ are coercive in sense of the Gårding type inequality:*

$$a_\eta(\boldsymbol{U},\boldsymbol{U}) + C_{1,\eta}\|\boldsymbol{U}\|_0^2 \geq C_{2,\eta}\|\boldsymbol{U}\|_{\boldsymbol{X}}^2, \tag{4.12}$$

$$a_\eta^*(\boldsymbol{U^*},\boldsymbol{U^*}) + C_{3,\eta}\|\boldsymbol{U^*}\|_0^2 \geq C_{4,\eta}\|\boldsymbol{U^*}\|_{\boldsymbol{X}}^2. \tag{4.13}$$

Proof. We first prove the coercivity of the adjoint bilinear form $a_\eta^*(\boldsymbol{U}^*,\boldsymbol{U}^*)$. According to the Korn, Cauchy–Schwarz, Poincaré, Young, and trace inequalities, we deduce (cf. [20, Lemma 3.2])

$$\begin{aligned}
&a_\eta^*(\boldsymbol{U^*},\boldsymbol{U^*}) + C_{3,\eta}\|\boldsymbol{U^*}\|_0^2\\
&= C_{3,\eta}\|\boldsymbol{U^*}\|_0^2 + 2\nu\left(\mathbb{D}(\boldsymbol{u^*}),\mathbb{D}(\boldsymbol{u^*})\right)_{\Omega_f} + \eta(\mathbb{K}\nabla\phi^*,\nabla\phi^*)_{\Omega_p} + \eta\langle g\boldsymbol{u^*}\cdot\boldsymbol{n}_f,\phi^*\rangle\\
&\quad - \langle\phi^*,\boldsymbol{u^*}\cdot\boldsymbol{n}_f\rangle + \alpha\langle P_\tau\boldsymbol{u^*},P_\tau\boldsymbol{u^*}\rangle + \eta\alpha\langle P_\tau\boldsymbol{u^*},P_\tau\left(\mathbb{K}\nabla\phi^*\right)\rangle\\
&\geq C_{3,\eta}\|\boldsymbol{U^*}\|_0^2 + 2\nu\|\mathbb{D}(\boldsymbol{u^*})\|_0^2 + \eta\lambda_{\min}(\mathbb{K})\|\nabla\phi^*\|_0^2 - \eta g C_1\|\nabla\boldsymbol{u^*}\|_0^{\frac{1}{2}}\|\boldsymbol{u^*}\|_0^{\frac{1}{2}}\|\nabla\phi^*\|_0^{\frac{1}{2}}\|\phi^*\|_0^{\frac{1}{2}}\\
&\quad + \alpha\|P_\tau\boldsymbol{u^*}\|_{\boldsymbol{L}^2(\Gamma)}^2 - C_1\|\nabla\boldsymbol{u^*}\|_0^{\frac{1}{2}}\|\boldsymbol{u^*}\|_0^{\frac{1}{2}}\|\nabla\phi^*\|_0^{\frac{1}{2}}\|\phi^*\|_0^{\frac{1}{2}} - \eta\alpha\lambda_{\max}(\mathbb{K})\|\nabla\boldsymbol{u^*}\|_0\|\nabla\phi^*\|_0\\
&\geq C_{3,\eta}\|\boldsymbol{U^*}\|_0^2 + 2C_0\nu\|\nabla\boldsymbol{u^*}\|_0^2 + \eta\lambda_{\min}(\mathbb{K})\|\nabla\phi^*\|_0^2 - \frac{\eta\lambda_{\min}(\mathbb{K})}{4}\|\nabla\phi^*\|_0^2\\
&\quad - \frac{\left(C_1(\eta g+1)\right)^4}{4\eta\lambda_{\min}(\mathbb{K})}\|\phi^*\|_0^2 - C_0\nu\|\nabla\boldsymbol{u^*}\|_0^2 - \frac{1}{16C_0\nu}\|\boldsymbol{u^*}\|_0^2\\
&\quad - \frac{\eta\alpha^2\lambda_{\max}^2(\mathbb{K})}{\lambda_{\min}(\mathbb{K})}\|\nabla\boldsymbol{u^*}\|_0^2 - \frac{\eta\lambda_{\min}(\mathbb{K})}{4}\|\nabla\phi^*\|_0^2\\
&= (C_{3,\eta} - \frac{1}{16C_0\nu})\|\boldsymbol{u^*}\|_0^2 + (C_{3,\eta} - \frac{\left(C_1(\eta g+1)\right)^4}{4\eta\lambda_{\min}(\mathbb{K})})\|\phi^*\|_0^2\\
&\quad + (C_0\nu - \frac{\eta\alpha^2\lambda_{\max}^2(\mathbb{K})}{\lambda_{\min}(\mathbb{K})})\|\nabla\boldsymbol{u^*}\|_0^2 + \frac{\eta\lambda_{\min}(\mathbb{K})}{2}\|\nabla\phi^*\|_0^2,
\end{aligned}$$

where C_i are generic constants depending on Ω, or Γ, or both Ω and Γ, $\lambda_{\min}(\mathbb{K})$, $\lambda_{\max}(\mathbb{K})$ are the smallest and largest eigenvalues of matrix $\mathbb{K}$, and $\langle P_\tau\boldsymbol{u},P_\tau(\mathbb{K}\nabla\phi)\rangle$ is understood as the duality between $\mathbf{H}_{\mathbf{00}}^{\mathbf{1/2}}(\Gamma)$ and $(\mathbf{H}_{\mathbf{00}}^{\mathbf{1/2}}(\Gamma))'$. In addition, the boundedness of $-(\eta g+1)C_1\|\nabla\boldsymbol{u^*}\|_0^{\frac{1}{2}}\|\boldsymbol{u^*}\|_0^{\frac{1}{2}}\|\nabla\phi^*\|_0^{\frac{1}{2}}\|\phi^*\|_0^{\frac{1}{2}}$ in the above is considered as follows:

$$\begin{aligned}
&-(\eta g+1)C_1\|\nabla\boldsymbol{u^*}\|_0^{\frac{1}{2}}\|\boldsymbol{u^*}\|_0^{\frac{1}{2}}\|\nabla\phi^*\|_0^{\frac{1}{2}}\|\phi^*\|_0^{\frac{1}{2}}\\
&\quad\geq -\frac{\|\nabla\boldsymbol{u^*}\|_0\|\boldsymbol{u^*}\|_0}{2} - \frac{\left((\eta g+1)C_1\right)^2\|\nabla\phi^*\|_0\|\phi^*\|_0}{2}\\
&\quad\geq -C_0\nu\|\nabla\boldsymbol{u^*}\|_0^2 - \frac{1}{16C_0\nu}\|\boldsymbol{u^*}\|_0^2 - \frac{\eta\lambda_{\min}(\mathbb{K})}{4}\|\nabla\phi^*\|_0^2 - \frac{\left(C_1(\eta g+1)\right)^4}{4\eta\lambda_{\min}(\mathbb{K})}\|\phi^*\|_0^2.
\end{aligned}$$

Once one chooses η and $C_{3,\eta}$ satisfying

$$\eta < \frac{C_0\nu\lambda_{\min}(\mathbb{K})}{\alpha^2\lambda_{\max}^2(\mathbb{K})} \text{ and } C_{3,\eta} \geq \max\{\frac{1}{16C_0\nu}, \frac{\left(C_1(\eta g+1)\right)^4}{4\eta\lambda_{\min}(\mathbb{K})}\},$$

then there exists a positive constant $C_{4,\eta}$ such that

$$a_\eta^*(\boldsymbol{U^*},\boldsymbol{U^*}) + C_{3,\eta}\|\boldsymbol{U^*}\|_0^2 \geq C_{4,\eta}\|\boldsymbol{U^*}\|_{\boldsymbol{X}}^2.$$

Furthermore, one can proceed with an argument similar to the above and identify η and $C_{1,\eta}$ satisfying $\eta < \frac{C_0\nu\lambda_{\min}(\mathbb{K})}{\alpha^2\lambda_{\max}^2(\mathbb{K})}$ and $C_{1,\eta} \geq \max\{\frac{(C_1(\eta g+1))^4}{8\eta C_0\nu}, \frac{1}{8\lambda_{\min}(\mathbb{K})}\}$; then there exists a constant $C_{2,\eta}$ such that

$$a_\eta(\boldsymbol{U},\boldsymbol{U}) + C_{1,\eta}\|\boldsymbol{U}\|_0^2 \geq C_{2,\eta}\|\boldsymbol{U}\|_{\boldsymbol{X}}^2.$$

The proof is finally completed by choosing $0 < \eta < \min\{\frac{C_0\nu\lambda_{\min}(\mathbb{K})}{\alpha^2\lambda_{\max}^2(\mathbb{K})}, \frac{C_0\nu\lambda_{\min}(\mathbb{K})}{\alpha^2\lambda_{\max}^2(\mathbb{K})}\}$. ∎

Remark 4.4. In this paper, we consider α, $\mathbb{K}$, ν, and g as given constant parameters. Then $C_{i,\eta}$ above, $i = 1,2,3,4$, are constants depending on Ω and Γ. We may later emerge $C_{i,\eta}$ to a generic coefficient $C_{\Omega,\Gamma}$.

The following lemma is for the continuity of $a_\eta(\boldsymbol{U},\boldsymbol{V})$ and $a_\eta^*(\boldsymbol{U^*},\boldsymbol{V})$, which follow naturally from a group of standard inequalities, such as the trace, Korn's, the Cauchy–Schwarz, and the Poincaré inequalities.

LEMMA 4.5. *$a_\eta(\boldsymbol{U},\boldsymbol{V})$ and $a_\eta^*(\boldsymbol{U^*},\boldsymbol{V})$ are continuous, i.e., there exist constants C depending on Ω, Γ, η, g, α, $\mathbb{K}$ such that*

$$a_\eta(\boldsymbol{U},\boldsymbol{V}) \leq C\|\boldsymbol{U}\|_{\boldsymbol{X}}\|\boldsymbol{V}\|_{\boldsymbol{X}}, \tag{4.14}$$

$$a_\eta^*(\boldsymbol{U^*},\boldsymbol{V}) \leq C\|\boldsymbol{U^*}\|_{\boldsymbol{X}}\|\boldsymbol{V}\|_{\boldsymbol{X}}. \tag{4.15}$$

Proof. We provide only a brief proof for the adjoint bilinear form $a_\eta^*(\boldsymbol{U}^*,\boldsymbol{V})$, since the proof of the other one is similar. We have

$$\begin{aligned}
a_\eta^*(\boldsymbol{U}^*,\boldsymbol{V}) \leq\ & 2C_3\nu\|\boldsymbol{u}^*\|_{\boldsymbol{X}_f}\|\boldsymbol{v}\|_{\boldsymbol{X}_f} + C_7\eta\lambda_{\max}(\mathbb{K})\|\phi^*\|_{X_p}\|\psi\|_{X_p} \\
& + C_4(1+\eta g)\|\boldsymbol{u}^*\|_{X_f}\|\psi\|_{X_p} + C_5\alpha\|\boldsymbol{u}^*\|_{\boldsymbol{X}_f}\|\boldsymbol{v}\|_{\boldsymbol{X}_f} \\
& + C_6\eta\alpha\lambda_{\max}(\mathbb{K})\|\boldsymbol{u}^*\|_{\boldsymbol{X}_f}\|\psi\|_{X_p} \\
\leq\ & C_{3,5}(\|\boldsymbol{u}^*\|_{\boldsymbol{X}_f}^2 + \|\phi^*\|_{X_p}^2)^{\frac{1}{2}}(\|\boldsymbol{v}\|_{\boldsymbol{X}_f}^2 + \|\psi\|_{X_p}^2)^{\frac{1}{2}} \\
& + C_{4,6}(\|\boldsymbol{u}^*\|_{\boldsymbol{X}_f}^2 + \|\phi^*\|_{X_p}^2)^{\frac{1}{2}}(\|\boldsymbol{v}\|_{\boldsymbol{X}_f}^2 + \|\psi\|_{X_p}^2)^{\frac{1}{2}} \\
& + C_7\eta\lambda_{\max}(\mathbb{K})(\|\boldsymbol{u}^*\|_{\boldsymbol{X}_f}^2 + \|\phi^*\|_{X_p}^2)^{\frac{1}{2}}(\|\boldsymbol{v}\|_{\boldsymbol{X}_f}^2 + \|\psi\|_{X_p}^2)^{\frac{1}{2}} \\
=\ & C\|\boldsymbol{U}^*\|_{\boldsymbol{X}}\|\boldsymbol{V}\|_{\boldsymbol{X}},
\end{aligned}$$

where C_i and $C_{i,j}$ are generic constants depending on Ω, Γ, and

$$\begin{aligned}
C_{3,5} &= \max\{2C_3\nu, C_5\alpha\}, \quad C_{4,6} = \max\{C_4(1+\eta g), C_6\eta\alpha\lambda_{\max}(\mathbb{K})\}, \\
C &= \max\{C_{3,5}, C_{4,6}, C_7\eta\lambda_{\max}(\mathbb{K})\}.
\end{aligned}$$

∎

Remark 4.6. A consequence of Lemmas 4.3 and 4.5 is the optimal FEM convergence of the backward adjoint equation if equipped with a regular, nonvariable force term. This can be shown by an extension of the proof in [19, Theorem 4.4].

In the following, we will use the rescaled norms which are naturally defined as

$$\|\boldsymbol{V}\|_{0,\eta} = (\eta\|\boldsymbol{v}\|_0^2 + \|\phi\|_0^2)^{\frac{1}{2}}, \|\boldsymbol{V}\|_{L^2(0,T;\boldsymbol{L}_\eta^2(\Omega))} = \left(\int_0^T \|\boldsymbol{V}\|_{0,\eta}^2 dt\right)^{\frac{1}{2}},$$

$$\|\boldsymbol{V}\|_{0,\eta^*} = (\|\boldsymbol{v}\|_0^2 + \eta\|\phi\|_0^2)^{\frac{1}{2}},\ \|\boldsymbol{V}\|_{L^2(0,T;\boldsymbol{L}_{\eta^*}^2(\Omega))} = \left(\int_0^T \|\boldsymbol{V}\|_{0,\eta^*}^2 dt\right)^{\frac{1}{2}}.$$

One can verify that these rescaled norms are well-defined and the standard triangle and Cauchy–Schwarz inequalities are applied. Furthermore, by definition, one can establish the norm equivalences for $\|\cdot\|_0$, $\|\cdot\|_{0,\eta}$ and $\|\cdot\|_{0,\eta^*}$ stated in the following lemma.

LEMMA 4.7. *Norms $\|\cdot\|_0$, $\|\cdot\|_{0,\eta}$, and $\|\cdot\|_{0,\eta^*}$ are connected with each other as*

$$C_\eta^1\|\boldsymbol{U}\|_{0,\eta} \le \|\boldsymbol{U}\|_0 \le C_\eta^2\|\boldsymbol{U}\|_{0,\eta}, \tag{4.16}$$

$$C_\eta^1\|\boldsymbol{U}\|_{0,\eta^*} \le \|\boldsymbol{U}\|_0 \le C_\eta^2\|\boldsymbol{U}\|_{0,\eta^*}, \tag{4.17}$$

$$C_\eta^3\|\boldsymbol{U}\|_{0,\eta^*} \le \|\boldsymbol{U}\|_{0,\eta} \le C_\eta^4\|\boldsymbol{U}\|_{0,\eta^*}, \tag{4.18}$$

where

$$C_\eta^1 = \min\left\{1, \frac{1}{\sqrt{\eta}}\right\}, C_\eta^2 = \max\left\{1, \frac{1}{\sqrt{\eta}}\right\}, C_\eta^3 = \min\left\{\sqrt{\eta}, \frac{1}{\sqrt{\eta}}\right\}, C_\eta^4 = \max\left\{\sqrt{\eta}, \frac{1}{\sqrt{\eta}}\right\}.$$

Define notations

$$\left\langle \frac{\partial \boldsymbol{U}}{\partial t}, \boldsymbol{V} \right\rangle_\eta = \left\langle \frac{\partial \phi}{\partial t}, \psi \right\rangle + \eta \left\langle \frac{\partial \boldsymbol{u}}{\partial t}, \boldsymbol{v} \right\rangle, \langle \boldsymbol{F}, \boldsymbol{V}\rangle_\eta = \langle f_p, \psi\rangle + \eta\langle \boldsymbol{f}_f, \boldsymbol{v}\rangle,$$

$$\left\langle \frac{\partial \boldsymbol{U}}{\partial t}, \boldsymbol{V} \right\rangle_\eta^* = \eta \left\langle \frac{\partial \phi}{\partial t}, \psi \right\rangle + \left\langle \frac{\partial \boldsymbol{u}}{\partial t}, \boldsymbol{v} \right\rangle, \langle \boldsymbol{F}, \boldsymbol{V}\rangle_\eta^* = \eta\langle f_p, \psi\rangle + \langle \boldsymbol{f}_f, \boldsymbol{v}\rangle.$$

Then, using equivalent arguments similar to those used for (2.10), (2.11), (2.12), and (2.13), we can rewrite the continuous optimality system (3.16)–(3.18) as

$$\begin{cases} \left\langle \frac{\partial \boldsymbol{U}}{\partial t}, \boldsymbol{V} \right\rangle_\eta + a_\eta(\boldsymbol{U}, \boldsymbol{V}) = \langle \boldsymbol{F}, \boldsymbol{V}\rangle_\eta \quad \forall \boldsymbol{V} \in \boldsymbol{X}_{\mathrm{div}}, \\ \boldsymbol{U}(\cdot, 0) = \boldsymbol{U_0} \quad \text{in } \boldsymbol{L}^2(\Omega), \end{cases} \tag{4.19}$$

$$\begin{cases} -\left\langle \frac{\partial \boldsymbol{U^*}}{\partial t}, \boldsymbol{V} \right\rangle_\eta^* + a_\eta^*(\boldsymbol{U^*}, \boldsymbol{V}) = \langle \widehat{\boldsymbol{U}} - \boldsymbol{U}, \boldsymbol{V}\rangle_\eta^* \quad \forall \boldsymbol{V} \in \boldsymbol{X}_{\mathrm{div}}, \\ \boldsymbol{U^*}(\cdot, T) = \boldsymbol{0} \quad \text{in } \boldsymbol{L}^2(\Omega), \end{cases} \tag{4.20}$$

$$\boldsymbol{U_0} = \frac{1}{\gamma}\boldsymbol{U^*}(\cdot, 0). \tag{4.21}$$

As mentioned previously, we intend to carry out the convergence analysis for the data assimilation problem by the finite element convergence results from classical PDEs. A key step is to introduce the following auxiliary equations:

$$\begin{cases} \left\langle \frac{\partial \boldsymbol{U}(\boldsymbol{U}_{0,h})}{\partial t}, \boldsymbol{V} \right\rangle_\eta + a_\eta(\boldsymbol{U}(\boldsymbol{U}_{0,h}), \boldsymbol{V}) = \langle \boldsymbol{F}, \boldsymbol{V}\rangle_\eta \quad \forall \boldsymbol{V} \in \boldsymbol{X}_{\mathrm{div}}, \\ \boldsymbol{U}(\boldsymbol{U}_{0,h})(\cdot, 0) = \boldsymbol{U}_{0,h} \quad \text{in } \boldsymbol{L}^2(\Omega), \end{cases} \tag{4.22}$$

$$\begin{cases} -\left\langle \frac{\partial \boldsymbol{U^*}(\boldsymbol{U}_{0,h})}{\partial t}, \boldsymbol{V} \right\rangle_\eta^* + a_\eta^*(\boldsymbol{U^*}(\boldsymbol{U}_{0,h}), \boldsymbol{V}) = \langle \widehat{\boldsymbol{U}} - \boldsymbol{U}(\boldsymbol{U}_{0,h}), \boldsymbol{V}\rangle_\eta^* \quad \forall \boldsymbol{V} \in \boldsymbol{X}_{\mathrm{div}}, \\ \boldsymbol{U^*}(\boldsymbol{U}_{0,h})(\cdot, T) = \boldsymbol{0} \quad \text{in } \boldsymbol{L}^2(\Omega), \end{cases} \tag{4.23}$$

$$\begin{cases} -\left\langle \frac{\partial \boldsymbol{U^*}(\boldsymbol{U}_h)}{\partial t}, \boldsymbol{V} \right\rangle_\eta^* + a_\eta^*(\boldsymbol{U^*}(\boldsymbol{U}_h), \boldsymbol{V}) = \langle \widehat{\boldsymbol{U}} - \boldsymbol{U}_h, \boldsymbol{V}\rangle_\eta^* \quad \forall \boldsymbol{V} \in \boldsymbol{X}_{\mathrm{div}}, \\ \boldsymbol{U^*}(\boldsymbol{U}_h)(\cdot, T) = \boldsymbol{0} \quad \text{in } \boldsymbol{L}^2(\Omega). \end{cases} \tag{4.24}$$

Equations (4.22) and (4.23) are used to remove the concern from the initial condition in (4.19). Equation (4.24) basically recovers the Galerkin orthogonality we lost between the continuous and discrete adjoint equations (4.20) and (4.11). Analyzing the equations in the continuous optimality system and the auxiliary equations in pair, we can establish the inequalities stated in the following lemma.

LEMMA 4.8. *Let* $\boldsymbol{U}(\boldsymbol{U}_{0,h}), \boldsymbol{U}^*(\boldsymbol{U}_{0,h}), \boldsymbol{U}^*(\boldsymbol{U}_h)$ *be solutions of* (4.22), (4.23), *and* (4.24), *respectively, and let* $(\boldsymbol{U}, \boldsymbol{U}^*, \boldsymbol{U_0})$ *and* $(\boldsymbol{U}_h, \boldsymbol{U}_h^*, \boldsymbol{U}_{0,h})$ *be solutions of the continuous and discrete optimality systems* (4.19)–(4.21) *and* (4.11)*; then the following estimates hold:*

$$\|\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h})\|_{L^2(0,T;\boldsymbol{L}^2_\eta(\Omega))} \le C_{\Omega,\Gamma,T}\|\boldsymbol{U_0} - \boldsymbol{U}_{0,h}\|_{0,\eta}, \tag{4.25}$$

$$\|\boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h})\|_{L^2(0,T;\boldsymbol{L}^2_{\eta^*}(\Omega))} \le C_{\Omega,\Gamma,T}\|\boldsymbol{U}(\boldsymbol{U}_{0,h}) - \boldsymbol{U}\|_{L^2(0,T;\boldsymbol{L}^2_\eta(\Omega))}, \tag{4.26}$$

$$\sup_{0\le t\le T}\|\boldsymbol{U}^*(\boldsymbol{U}_h) - \boldsymbol{U}^*(\boldsymbol{U}_{0,h})\|_{0,\eta^*} \le C_{\Omega,\Gamma,T}\|\boldsymbol{U}(\boldsymbol{U}_{0,h}) - \boldsymbol{U}_h\|_{L^2(0,T;\boldsymbol{L}^2_\eta(\Omega))}, \tag{4.27}$$

$$\|\boldsymbol{U}^*(\boldsymbol{U}_h) - \boldsymbol{U}^*(\boldsymbol{U}_{0,h})\|_{L^2(0,T;\boldsymbol{L}^2_{\eta^*}(\Omega))} \le C_{\Omega,\Gamma,T}\|\boldsymbol{U}(\boldsymbol{U}_{0,h}) - \boldsymbol{U}_h\|_{L^2(0,T;\boldsymbol{L}^2_\eta(\Omega))}. \tag{4.28}$$

Proof. Subtracting (4.22) from (4.19), we have

$$\begin{cases} \left\langle \dfrac{\partial(\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}))}{\partial t}, \boldsymbol{V} \right\rangle_\eta + a_\eta(\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}), \boldsymbol{V}) = \langle \boldsymbol{0}, \boldsymbol{V}\rangle_\eta \quad \forall \boldsymbol{V} \in \boldsymbol{X}_{\mathrm{div}}, \\ (\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}))(\cdot, 0) = \boldsymbol{U_0} - \boldsymbol{U}_{0,h} \quad \text{in } \boldsymbol{L}^2(\Omega). \end{cases} \tag{4.29}$$

Taking $\boldsymbol{V} = \boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h})$ on (4.29), using the coercive inequality (4.12) and norm relation (4.16), we obtain

$$\frac{d\|\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h})\|^2_{0,\eta}}{2dt} + C_{2,\eta}\|\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h})\|^2_{\boldsymbol{X}} \le C_{1,\eta}C_\eta^2\|\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h})\|^2_{0,\eta}. \tag{4.30}$$

Applying the Gronwall inequality on (4.30) leads to

$$\|(\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}))(\cdot,t)\|^2_{0,\eta} + C_{\Omega,\Gamma,T}\int_0^t \|\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h})\|^2_{\boldsymbol{X}} \le C_{\Omega,\Gamma,T}\|\boldsymbol{U_0} - \boldsymbol{U}_{0,h}\|^2_{0,\eta}, \tag{4.31}$$

where $C_{\Omega,\Gamma,T}$ is a generic emerged constant that depends on Ω, Γ, T. Inequality (4.31) leads to

$$\|\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h})\|_{L^2(0,T;\boldsymbol{L}^2_\eta(\Omega))} \le C_{\Omega,\Gamma,T}\|\boldsymbol{U_0} - \boldsymbol{U}_{0,h}\|_{0,\eta}. \tag{4.32}$$

Again, we subtract (4.23) from (4.20) to obtain

$$\begin{cases} -\left\langle \dfrac{\partial(\boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h}))}{\partial t}, \boldsymbol{V} \right\rangle^*_\eta + a^*_\eta(\boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h}), \boldsymbol{V}) \\ \qquad = \langle \boldsymbol{U}(\boldsymbol{U}_{0,h}) - \boldsymbol{U}, \boldsymbol{V}\rangle^*_\eta \quad \forall \boldsymbol{V} \in \boldsymbol{X}_{\mathrm{div}}, \\ (\boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h}))(\cdot, T) = \boldsymbol{0} \quad \text{in } \boldsymbol{L}^2(\Omega). \end{cases} \tag{4.33}$$

Testing (4.33) with $\boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h})$ and using the coercive inequality (4.13), the Cauchy–Schwarz and Young's inequalities, and (4.17)–(4.18), we deduce

$$\begin{aligned} &-\frac{d\|\boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h})\|^2_{0,\eta^*}}{dt} + 2C_{4,\eta}\|\boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h})\|^2_{\boldsymbol{X}} \\ &\le 2C_{3,\eta}\|\boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h})\|^2_0 + 2\|\boldsymbol{U}(\boldsymbol{U}_{0,h}) - \boldsymbol{U}\|_{0,\eta^*}\|\boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h})\|_{0,\eta^*} \\ &\le (1 + 2C_{3,\eta}(C_\eta^2)^2)\|\boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h})\|^2_{0,\eta^*} + \frac{1}{(C_\eta^3)^2}\|\boldsymbol{U}(\boldsymbol{U}_{0,h}) - \boldsymbol{U}\|^2_{0,\eta}. \end{aligned} \tag{4.34}$$

Applying the Gronwall's inequality on (4.34), we find out

$$\|\boldsymbol{U}^{\boldsymbol{*}}-\boldsymbol{U}^{\boldsymbol{*}}(\boldsymbol{U}_{0,h})\|_{L^2(0,T;\boldsymbol{L}^2_{\eta^*}(\Omega))}\le C_{\Omega,\Gamma,T}\|\boldsymbol{U}(\boldsymbol{U}_{0,h})-\boldsymbol{U}\|_{L^2(0,T;\boldsymbol{L}^2_{\eta}(\Omega))}. \tag{4.35}$$

Finally, subtracting (4.24) from (4.20), we have

$$\begin{cases} -\left\langle \dfrac{\partial(\boldsymbol{U}^{\boldsymbol{*}}(\boldsymbol{U}_h)-\boldsymbol{U}^{\boldsymbol{*}}(\boldsymbol{U}_{0,h}))}{\partial t},\boldsymbol{V}\right\rangle_{\eta}^{*}+a^*_{\eta}(\boldsymbol{U}^{\boldsymbol{*}}(\boldsymbol{U}_h)-\boldsymbol{U}^{\boldsymbol{*}}(\boldsymbol{U}_{0,h}),\boldsymbol{V})\\ \qquad =\langle \boldsymbol{U}(\boldsymbol{U}_{0,h})-\boldsymbol{U}_h,\boldsymbol{V}\rangle^*_{\eta}\quad \forall \boldsymbol{V}\in \boldsymbol{X}_{\mathrm{div}},\\ (\boldsymbol{U}^{\boldsymbol{*}}(\boldsymbol{U}_h)-\boldsymbol{U}^{\boldsymbol{*}}(\boldsymbol{U}_{0,h}))(\cdot,T)=\boldsymbol{0}\quad \text{in } \boldsymbol{L}^2(\Omega). \end{cases} \tag{4.36}$$

Similarly, by choosing test function $\boldsymbol{V}=\boldsymbol{U}^{\boldsymbol{*}}(\boldsymbol{U}_h)-\boldsymbol{U}^{\boldsymbol{*}}(\boldsymbol{U}_{0,h})$ in (4.36) and applying the coercive inequality (4.13), the norm relations (4.17)–(4.18), the Cauchy–Schwarz inequality, the Young's inequality, and the Gronwall's inequality, the following estimates hold:

$$\sup_{0\le t<T}\|\boldsymbol{U}^{\boldsymbol{*}}(\boldsymbol{U}_h)-\boldsymbol{U}^{\boldsymbol{*}}(\boldsymbol{U}_{0,h})\|_{0,\eta^*}\le C_{\Omega,\Gamma,T}\|\boldsymbol{U}(\boldsymbol{U}_{0,h})-\boldsymbol{U}_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta}(\Omega))}, \tag{4.37}$$

$$\|\boldsymbol{U}^{\boldsymbol{*}}(\boldsymbol{U}_h)-\boldsymbol{U}^{\boldsymbol{*}}(\boldsymbol{U}_{0,h})\|_{L^2(0,T;\boldsymbol{L}^2_{\eta^*}(\Omega))}\le C_{\Omega,\Gamma,T}\|\boldsymbol{U}(\boldsymbol{U}_{0,h})-\boldsymbol{U}_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta}(\Omega))}. \tag{4.38}$$

The proof is completed by putting (4.32), (4.35), (4.37), and (4.38) together. ∎

By using the triangle inequality and inequality (4.25), $\|\boldsymbol{U}-\boldsymbol{U}_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta}(\Omega))}$ can be estimated as

$$\begin{aligned}\|\boldsymbol{U}-\boldsymbol{U}_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta}(\Omega))}&\le \|\boldsymbol{U}-\boldsymbol{U}(\boldsymbol{U}_{0,h})\|_{L^2(0,T;\boldsymbol{L}^2_{\eta}(\Omega))}+\|\boldsymbol{U}(\boldsymbol{U}_{0,h})-\boldsymbol{U}_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta}(\Omega))}\\ &\le C_{\Omega,\Gamma,T}\|\boldsymbol{U}_0-\boldsymbol{U}_{0,h}\|_{0,\eta}+\|\boldsymbol{U}(\boldsymbol{U}_{0,h})-\boldsymbol{U}_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta}(\Omega))}.\end{aligned} \tag{4.39}$$

Using inequalities (4.26), (4.28), and (4.25), one can also bound $\|\boldsymbol{U}^*-\boldsymbol{U}^*_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta^*}(\Omega))}$ as follows:

$$\begin{aligned}&\|\boldsymbol{U}^*-\boldsymbol{U}^*_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta^*}(\Omega))}\\ &\quad\le \|\boldsymbol{U}^*-\boldsymbol{U}^*(\boldsymbol{U}_{0,h})\|_{L^2(0,T;\boldsymbol{L}^2_{\eta^*}(\Omega))}+\|\boldsymbol{U}^*(\boldsymbol{U}_{0,h})-\boldsymbol{U}^*(\boldsymbol{U}_h)\|_{L^2(0,T;\boldsymbol{L}^2_{\eta^*}(\Omega))}\\ &\qquad+\|\boldsymbol{U}^*(\boldsymbol{U}_h)-\boldsymbol{U}^*_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta^*}(\Omega))}\\ &\quad\le C_{\Omega,\Gamma,T}(\|\boldsymbol{U}(\boldsymbol{U}_{0,h})-\boldsymbol{U}\|_{L^2(0,T;\boldsymbol{L}^2_{\eta}(\Omega))}+\|\boldsymbol{U}(\boldsymbol{U}_{0,h})-\boldsymbol{U}_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta}(\Omega))})\\ &\qquad+\|\boldsymbol{U}^*(\boldsymbol{U}_h)-\boldsymbol{U}^*_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta^*}(\Omega))}\\ &\quad\le C_{\Omega,\Gamma,T}(\|\boldsymbol{U}_0-\boldsymbol{U}_{0,h}\|_{0,\eta}+\|\boldsymbol{U}(\boldsymbol{U}_{0,h})-\boldsymbol{U}_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta}(\Omega))})\\ &\qquad+\|\boldsymbol{U}^*(\boldsymbol{U}_h)-\boldsymbol{U}^*_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta^*}(\Omega))}.\end{aligned} \tag{4.40}$$

Note that $\boldsymbol{U}_h$ and $\boldsymbol{U}^*_h$ are the classical finite element approximations of $\boldsymbol{U}(\boldsymbol{U}_{0,h})$ and $\boldsymbol{U}^*(\boldsymbol{U}_h)$. From (4.39) and (4.40), we observe that the bounds for the finite element approximations depend on $\|\boldsymbol{U}_0-\boldsymbol{U}_{0,h}\|_{0,\eta}$, which is estimated in the following lemma through two given equalities $\boldsymbol{U}_0=\frac{1}{\gamma}\boldsymbol{U}^*(\cdot,0)$ and $\boldsymbol{U}_{0,h}=\frac{1}{\gamma}\boldsymbol{U}^{*0}_h$.

LEMMA 4.9. *Let* $\boldsymbol{U}_0$, $\boldsymbol{U}_{0,h}$, $\boldsymbol{U}^{*0}_h$, $\boldsymbol{U}^*(\boldsymbol{U}_{0,h})(\cdot,0)$ *be functions defined in* (3.16), (4.11), *and* (4.23). *Then the following error estimate holds:*

$$\|\boldsymbol{U}_0-\boldsymbol{U}_{0,h}\|_{0,\eta}\le\frac{C_{\Omega,\Gamma}}{\gamma}\|\boldsymbol{U}^*(\boldsymbol{U}_{0,h})(\cdot,0)-\boldsymbol{U}^{*0}_h\|_{0,\eta^*}. \tag{4.41}$$

Proof. Using $\boldsymbol{U}_0 = \frac{1}{\gamma}\boldsymbol{U}^*(\cdot,0)$ and $\boldsymbol{U}_{0,h} = \frac{1}{\gamma}\boldsymbol{U}_h^{*0}$ we have

$$\begin{aligned}\|\boldsymbol{U}_0 - \boldsymbol{U}_{0,h}\|_0^2 &= \frac{1}{\gamma}(\boldsymbol{U}^*(\cdot,0) - \boldsymbol{U}_h^{*0}, \boldsymbol{U}_0 - \boldsymbol{U}_{0,h}) \\ &= \frac{1}{\gamma}(\boldsymbol{U}^*(\cdot,0) - \boldsymbol{U}^*(\boldsymbol{U}_{0,h})(\cdot,0), \boldsymbol{U}_0 - \boldsymbol{U}_{0,h}) \\ &\quad + \frac{1}{\gamma}(\boldsymbol{U}^*(\boldsymbol{U}_{0,h})(\cdot,0) - \boldsymbol{U}_h^{*0}, \boldsymbol{U}_0 - \boldsymbol{U}_{0,h}).\end{aligned} \tag{4.42}$$

Taking $\boldsymbol{V} = \boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h})$ on (4.29) without the scalar η, we obtain

$$\begin{aligned}&\int_0^T \left\langle \frac{\partial(\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}))}{\partial t}, \boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h}) \right\rangle dt \\ &\quad + \int_0^T a(\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}), \boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h}))dt = 0.\end{aligned} \tag{4.43}$$

Taking integration by parts with respect to t on (4.43) results in

$$\begin{aligned}&((\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}))(\cdot,T), (\boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h}))(\cdot,T)) \\ &\quad - ((\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}))(\cdot,0), (\boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h}))(\cdot,0)) \\ &\quad - \int_0^T \left\langle \frac{\partial(\boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h}))}{\partial t}, \boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}) \right\rangle dt \\ &\quad + \int_0^T a(\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}), \boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h}))dt = 0.\end{aligned}$$

Using (4.33) without the scalar η and the fact $a(\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}), \boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h})) = a^*(\boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h}), \boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}))$, we simplify the previous equation as

$$((\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}))(\cdot,0), (\boldsymbol{U}^* - \boldsymbol{U}^*(\boldsymbol{U}_{0,h}))(\cdot,0)) = -\int_0^T (\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}), \boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}))dt.$$

Based on the nonnegativity of $\int_0^T (\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}), \boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}_{0,h}))dt$, (4.16), (4.17), and (4.42), we have

$$\|\boldsymbol{U}_0 - \boldsymbol{U}_{0,h}\|_{0,\eta} \le \frac{C_{\Omega,\Gamma}}{\gamma}\|\boldsymbol{U}^*(\boldsymbol{U}_{0,h})(\cdot,0) - \boldsymbol{U}_h^{*0}\|_{0,\eta^*},$$

where $C_{\Omega,\Gamma} = \frac{C_\eta^2}{C_\eta^1}$. ∎

Using (4.41) and the triangle inequality, $\|\boldsymbol{U}_0 - \boldsymbol{U}_{0,h}\|_{0,\eta}$ can be bounded as below:

$$\begin{aligned}&\|\boldsymbol{U}_0 - \boldsymbol{U}_{0,h}\|_{0,\eta} \le \frac{C_{\Omega,\Gamma}}{\gamma}\|\boldsymbol{U}_h^{*0} - \boldsymbol{U}^*(\boldsymbol{U}_{0,h})(\cdot,0)\|_{0,\eta^*} \\ &\le \frac{C_{\Omega,\Gamma}}{\gamma}\|\boldsymbol{U}_h^{*0} - \boldsymbol{U}^*(\boldsymbol{U}_h)(\cdot,0)\|_{0,\eta^*} + \frac{C_{\Omega,\Gamma}}{\gamma}\|\boldsymbol{U}^*(\boldsymbol{U}_h)(\cdot,0) - \boldsymbol{U}^*(\boldsymbol{U}_{0,h})(\cdot,0)\|_{0,\eta^*} \\ &\le \frac{C_{\Omega,\Gamma}}{\gamma}\max_{0\le i\le N-1}\|\boldsymbol{U}_h^{*i} - \boldsymbol{U}^*(\boldsymbol{U}_h)(\cdot,t_i)\|_{0,\eta^*} + \frac{C_{\Omega,\Gamma}}{\gamma}\sup_{0\le t<T}\|\boldsymbol{U}^*(\boldsymbol{U}_h) - \boldsymbol{U}^*(\boldsymbol{U}_{0,h})\|_{0,\eta^*} \\ &\le \frac{C_{\Omega,\Gamma}}{\gamma}\max_{0\le i\le N-1}\|\boldsymbol{U}_h^{*i} - \boldsymbol{U}^*(\boldsymbol{U}_h)(\cdot,t_i)\|_{0,\eta^*} + \frac{C_{\Omega,\Gamma,T}}{\gamma}\|\boldsymbol{U}_h - \boldsymbol{U}(\boldsymbol{U}_{0,h})\|_{L^2(0,T;\boldsymbol{L}^2_\eta(\Omega))}.\end{aligned} \tag{4.44}$$

Summarizing (4.39), (4.40), (4.44), and the classical FEM error estimates [19, Theorem 4.4] and Remark 4.6, we finally arrive at the estimation

$$\begin{aligned}
&\|\boldsymbol{U}_0-\boldsymbol{U}_{0,h}\|_{0,\eta}+\|\boldsymbol{U}-\boldsymbol{U}_h\|_{L^2(0,T;\boldsymbol{L}^2_\eta(\Omega))}+\|\boldsymbol{U}^*-\boldsymbol{U}^*_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta^*}(\Omega))}\\
&\quad\le C_{\Omega,\Gamma,T}\|\boldsymbol{U}_0-\boldsymbol{U}_{0,h}\|_{0,\eta}+(C_{\Omega,\Gamma,T}+1)\|\boldsymbol{U}(\boldsymbol{U}_{0,h})-\boldsymbol{U}_h\|_{L^2(0,T;\boldsymbol{L}^2_\eta(\Omega)}\\
&\qquad+\|\boldsymbol{U}^*(\boldsymbol{U}_h)-\boldsymbol{U}^*_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta^*}(\Omega))}\\
&\quad\le\frac{C_{\Omega,\Gamma,T}}{\gamma}\max_{0\le i\le N-1}\|\boldsymbol{U}^{*i}_h-\boldsymbol{U}^*(\boldsymbol{U}_h)(\cdot,t_i)\|_{0,\eta}+\frac{C_{\Omega,\Gamma,T}}{\gamma}\|\boldsymbol{U}_h-\boldsymbol{U}(\boldsymbol{U}_{0,h})\|_{L^2(0,T;\boldsymbol{L}^2_\eta(\Omega))}\\
&\qquad+(C_{\Omega,\Gamma,T}+1)\|\boldsymbol{U}(\boldsymbol{U}_{0,h})-\boldsymbol{U}_h\|_{L^2(0,T;\boldsymbol{L}^2_\eta(\Omega)}+\|\boldsymbol{U}^*(\boldsymbol{U}_h)-\boldsymbol{U}^*_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta^*}(\Omega)}\\
&\quad\le C_{\gamma,\Omega,\Gamma,T}(h^{r+1}+\tau),
\end{aligned}$$

where r is the polynomial degree of the finite element basis function.

THEOREM 4.10. *Let* $(\boldsymbol{U}_0,\boldsymbol{U},\boldsymbol{U}^*)$ *and* $(\boldsymbol{U}_{0,h},\boldsymbol{U}_h,\boldsymbol{U}^*_h)$ *be solutions of the continuous optimality system* (3.16)–(3.18) *and the discrete optimality system* (4.11)*, respectively. Assuming the input data are smooth enough; then the following error estimate holds:*

$$\begin{aligned}
&\|\boldsymbol{U}_0-\boldsymbol{U}_{0,h}\|_{0,\eta}+\|\boldsymbol{U}-\boldsymbol{U}_h\|_{L^2(0,T;\boldsymbol{L}^2_\eta(\Omega))}+\|\boldsymbol{U}^*-\boldsymbol{U}^*_h\|_{L^2(0,T;\boldsymbol{L}^2_{\eta^*}(\Omega))}\\
&\qquad\le C_{\gamma,\Omega,\Gamma,T}(h^{r+1}+\tau),
\end{aligned}\tag{4.45}$$

where $C_{\gamma,\Omega,\Gamma,T}$ *is a constant proportional to* $\frac{1}{\gamma}$ *and also depends on* Ω, Γ, *and* T.

The inequality in (4.45) indicates that very small regularization parameter γ may have a negative impact on the numerical accuracy. Therefore, in practice, refined h and τ are needed to offset the impact from a small γ.

5. Iterative methods for solving the discrete optimality system. Due to the complex structure of the Stokes–Darcy model and the forward-backward coupled temporal nature in the optimality system, solving (4.11) directly results in an extremely large coupled linear system [66], thereby being very computationally expensive. Hence we propose three iterative algorithms, the CG method, the BFGS method, and the SD method, to decouple the discrete optimality system.

For a clear implementation of these gradient-based methods, we first show a matrix-vector calculation of the gradient at each iteration based on the finite element assembling. In the rest of the presentation, we define $\vec{\Phi}\in\mathbb{R}^m$ as the vector representation of an element $\Phi\in H$ with respect to the given finite element base. Specifically, we define $\|\vec{\Phi}\|_0^2=(\vec{\Phi},\vec{\Phi})_M=\vec{\Phi}^TM\vec{\Phi}=\|\Phi\|_0^2$, where M is the corresponding mass matrix. Recalling (4.7)–(4.10) or the OptS (4.11), a vector form of the gradient

$$\nabla\vec{J}_h(\boldsymbol{U}^{(k)}_{0,h})=\gamma\vec{\boldsymbol{U}}^{(k)}_{0,h}-\vec{\boldsymbol{U}}^{*0(k)}_h=\gamma\vec{\boldsymbol{U}}^{(k)}_{0,h}-\begin{pmatrix}\vec{\phi}^{*0(k)}_h\\ \vec{\boldsymbol{u}}^{*0(k)}_h\end{pmatrix}\tag{5.1}$$

can be obtained by sequentially solving the following equations forward and backward for $n=1,2,\dots,N-1$:

$$\frac{M_a}{\tau}\left(\begin{pmatrix}\vec{\phi}^{n+1(k)}_h\\ \vec{\boldsymbol{u}}^{n+1(k)}_h\\ \vec{p}^{n+1(k)}_h\end{pmatrix}-\begin{pmatrix}\vec{\phi}^{n(k)}_h\\ \vec{\boldsymbol{u}}^{n(k)}_h\\ \vec{p}^{n(k)}_h\end{pmatrix}\right)+S\begin{pmatrix}\vec{\phi}^{n+1(k)}_h\\ \vec{\boldsymbol{u}}^{n+1(k)}_h\\ \vec{p}^{n+1(k)}_h\end{pmatrix}=\begin{pmatrix}\vec{f}^{n+1(k)}_{p,h}\\ \vec{\boldsymbol{f}}^{n+1(k)}_{f,h}\\ \vec{0}\end{pmatrix},\ \vec{\boldsymbol{U}}^0_h=\begin{pmatrix}\vec{\phi}^{0(k)}_h\\ \vec{\boldsymbol{u}}^{0(k)}_h\end{pmatrix},\tag{5.2}$$

$$-\frac{M_a}{\tau}\left(\begin{pmatrix}\vec{\phi}_h^{*n+1(k)}\\ \vec{\boldsymbol{u}}_h^{*n+1(k)}\\ \vec{p}_h^{*n+1(k)}\end{pmatrix}-\begin{pmatrix}\vec{\phi}_h^{*n(k)}\\ \vec{\boldsymbol{u}}_h^{*n(k)}\\ \vec{p}_h^{*n(k)}\end{pmatrix}\right)+S^*\begin{pmatrix}\vec{\phi}_h^{*n(k)}\\ \vec{\boldsymbol{u}}_h^{*n(k)}\\ \vec{p}_h^{*n(k)}\end{pmatrix}$$
$$=\begin{pmatrix}\vec{\widehat{\phi}}_h^{n+1(k)}\\ \vec{\widehat{\boldsymbol{u}}}_h^{n+1(k)}\\ \vec{0}\end{pmatrix}-M_a\begin{pmatrix}\vec{\phi}_h^{n+1(k)}\\ \vec{\boldsymbol{u}}_h^{n+1(k)}\\ \vec{0}\end{pmatrix},\ \vec{\boldsymbol{U}}_h^{*N(k)}=\begin{pmatrix}\vec{0}\\ \vec{\boldsymbol{0}}\end{pmatrix}. \tag{5.3}$$

Recalling the definition of the operators A, A^*, B, B^*, the matrices mentioned above are assembled as

$$M_a=\begin{pmatrix}M_{a\phi}&0&0\\0&M_{a\boldsymbol{u}}&0\\0&0&0\end{pmatrix},\quad M_{a\boldsymbol{u}}=\begin{pmatrix}M_{au}&0\\0&M_{au}\end{pmatrix},\quad M_{fp}=\begin{pmatrix}M_{a\phi}&0\\0&M_{a\boldsymbol{u}}\end{pmatrix}, \tag{5.4}$$

$$S=\begin{pmatrix}S_{a\phi}&S_{\boldsymbol{u}\phi}&0\\S_{\phi\boldsymbol{u}}&S_{a\boldsymbol{u}}+S_{\boldsymbol{uu}}&S_{p\boldsymbol{u}}\\0&S_{\boldsymbol{u}p}&0\end{pmatrix},\quad S^*=\begin{pmatrix}S_{a\phi}&S^*_{\boldsymbol{u}\phi}&0\\S^*_{\phi\boldsymbol{u}}&S_{a\boldsymbol{u}}+S^*_{\boldsymbol{uu}}&S_{p\boldsymbol{u}}\\0&S_{\boldsymbol{u}p}&0\end{pmatrix}, \tag{5.5}$$

and the related matrices $M_{a\phi}, M_{au}, S_{a\phi}, S_{p\boldsymbol{u}}, S_{\boldsymbol{u}p}, S_{a\boldsymbol{u}}, S_{\phi\boldsymbol{u}}, S_{\boldsymbol{u}\phi}, S_{\boldsymbol{uu}}, S^*_{\boldsymbol{u}\phi}, S^*_{\boldsymbol{uu}}, S^*_{\phi\boldsymbol{u}}$ and other vectors $\vec{f}_{p,h}^{n+1}, \vec{\boldsymbol{f}}_{f,h}^{n+1}$, $\vec{\widehat{\phi}}_h^{n+1}$, and $\vec{\widehat{\boldsymbol{u}}}_h^{n+1}$ are further assembled as follows:

$$M_{a\phi}=\left[\int_{\Omega_p}\psi_j\psi_i dxdy\right], M_{au}=\left[\int_{\Omega_f}v_jv_i dxdy\right], S_{a\phi}=\left[\int_{\Omega_p}\mathbb{K}\nabla\psi_j\nabla\psi_i dxdy\right],$$
$$S_{p\boldsymbol{u}}=\left[-\int_{\Omega_f}q_j\nabla\cdot\boldsymbol{v}_i dxdy\right], S_{\boldsymbol{u}p}=S^T_{p\boldsymbol{u}}, S_{a\boldsymbol{u}}=\left[\int_{\Omega_f}2\nu\mathbb{D}(\boldsymbol{v}_j):\mathbb{D}(\boldsymbol{v}_i)\,dxdy\right],$$
$$S_{\phi\boldsymbol{u}}=\left[\int_\Gamma g\psi_j\boldsymbol{v}_i\cdot\boldsymbol{n_f}dS\right]+\left[\int_\Gamma\alpha P_\tau\left(\mathbb{K}\nabla\psi_j\right)P_\tau\boldsymbol{v}_i dS\right], S_{\boldsymbol{u}\phi}=-\left[\int_\Gamma\boldsymbol{v_j}\cdot\boldsymbol{n}_f\psi_i dS\right],$$
$$S_{\boldsymbol{uu}}=\left[\int_\Gamma\alpha P_\tau\boldsymbol{v}_jP_\tau\boldsymbol{v}_i dS\right], S^*_{\boldsymbol{u}\phi}=\left[\int_\Gamma g\boldsymbol{v}_j\cdot\boldsymbol{n_f}\psi_i dS\right]+\left[\int_\Gamma\alpha P_\tau\boldsymbol{v}_jP_\tau\left(\mathbb{K}\nabla\psi_i\right)dS\right],$$
$$S^*_{\boldsymbol{uu}}=\left[\int_\Gamma\alpha P_\tau\boldsymbol{v}_jP_\tau\boldsymbol{v}_i dS\right], S^*_{\phi\boldsymbol{u}}=-\left[\int_\Gamma\psi_j\boldsymbol{v}_i\cdot\boldsymbol{n}_f dS\right], \vec{f}_{p,h}^{n+1}=\left[\int_{\Omega_p}f_p(t_{n+1})\psi_i dxdy\right],$$
$$\vec{\boldsymbol{f}}_{f,h}^{n+1}=\left[\int_{\Omega_f}\boldsymbol{f}_f(t_{n+1})\boldsymbol{v}_i dxdy\right], \vec{\widehat{\phi}}_h^{n+1}=\left[\int_{\Omega_p}\widehat{\phi}(t_{n+1})\psi_i dxdy\right],$$
$$\vec{\widehat{\boldsymbol{u}}}_h^{n+1}=\left[\int_{\Omega_f}\widehat{\boldsymbol{u}}(t_{n+1})\boldsymbol{v}_i dxdy\right],$$

where $\{\psi_i\}$, $\{\boldsymbol{v}_i\}=(v_i,v_i)^T$, and $\{q_i\}$ are basis functions of the finite element spaces X_p^h, $\boldsymbol{X}_f^h$, and Q^h, respectively.

5.1. The conjugate gradient method. The CG method is a powerful algorithm for solving the VDA problem. It can achieve a linear or superlinear convergence rate only considering the first order derivative. The main feature of CG is that the current descent direction $\boldsymbol{d}^k$ is conjugate orthogonal to all previous descent directions $\boldsymbol{d}^0, \boldsymbol{d}^1, \boldsymbol{d}^2, \ldots \boldsymbol{d}^{k-1}$, which allows a finite iteration convergence for the finite-dimensional optimization. A standard CG algorithm update is stated as follows:

- Initialize $\boldsymbol{U}_{0,h}^{(0)}$ and $\boldsymbol{d}^0=-\nabla J_h(\boldsymbol{U}_{0,h}^{(0)})$.

- Update $\boldsymbol{U}_{0,h}^{(k+1)} = \boldsymbol{U}_{0,h}^{(k)} + \eta^k \boldsymbol{d}^k$ with

$$\boldsymbol{d}^k = \begin{cases} -\nabla J_h(\boldsymbol{U}_{0,h}^{(0)}) \; k = 0, \\ -\nabla J_h(\boldsymbol{U}_{0,h}^{(k)}) + \beta^k \boldsymbol{d}^{k-1} \; k \geq 1, \end{cases} \quad \beta^k = \frac{\left(\nabla J_h(\boldsymbol{U}_{0,h}^{(k)}), \nabla J_h(\boldsymbol{U}_{0,h}^{(k)})\right)}{\left(\nabla J_h(\boldsymbol{U}_{0,h}^{(k-1)}), \nabla J_h(\boldsymbol{U}_{0,h}^{(k-1)})\right)}.$$

- η^k can be simply chosen as $\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots\}$, or determined by using exact line search ($\eta_k = \operatorname{argmin}_{\eta \in \mathbb{R}} J(\boldsymbol{U}_{0,h}^{(k)} + \eta \boldsymbol{d}^k)$), or using inexact line search methods, such as the Armijo or Wolfe condition [50, 55].

The CG method [66, 67] is also interpreted as an accelerated SD method based on the fixed point theorem, which is naturally stated as follows:

- Initialize $\boldsymbol{U}_{0,h}^{(0)}$ and $\boldsymbol{U}_{0,h}^{(1)}$.
- Update $\boldsymbol{U}_{0,h}^{(k+1)} = \boldsymbol{U}_{0,h}^{(k)} - \zeta^k \nabla J_h(\boldsymbol{U}_{0,h}^{(k)}) + \vartheta^k (\boldsymbol{U}_{0,h}^{(k)} - \boldsymbol{U}_{0,h}^{(k-1)})$.

$\zeta^k$and ϑ^k are constants to be determined, and $\boldsymbol{U}_{0,h}^{(k)} - \boldsymbol{U}_{0,h}^{(k-1)}$ is a momentum term accounting for the acceleration.

We will adopt the first CG version as our presentation and provide an exact line search method to optimally determine the step sizes η^k and β^k. To begin with, we recall that the CG method [49] was originally developed to solve the linear system

$$\mathcal{A}x = b, \, \mathcal{A} \text{ is a positive definite operator (or matrix).} \tag{5.6}$$

Our task is to rewrite the optimality condition of the data assimilation problem in (5.6). For convenience, we will keep the discussion in the continuous level, and solving the discrete data assimilation will be a straightforward discretization of the continuous one. We revisit the optimization problem (3.1) as a reduced form

$$\min_{\boldsymbol{U_0} \in \boldsymbol{Y}_{ad}} J(\boldsymbol{U_0}) = \frac{1}{2} \int_0^T \|\widehat{\boldsymbol{U}} - \boldsymbol{U}\|_0^2 dt + \frac{\gamma}{2} \|\boldsymbol{U_0}\|_0^2 \quad \text{subject to } \boldsymbol{U} = \mathcal{S}_{\boldsymbol{F}} \boldsymbol{U}_0, \tag{5.7}$$

where the operator $\mathcal{S}_{\boldsymbol{F}} : \boldsymbol{L}^2(\Omega) \mapsto \boldsymbol{W}(0,T)$ is defined by the Stokes–Darcy equation (2.12) and the subscript $\boldsymbol{F}$ corresponds to the source term. Since (2.12) is a linear PDE, the operator $\mathcal{S}_{\boldsymbol{F}}$ is an affine mapping. Hence, $\mathcal{S}'_{\boldsymbol{F}} \boldsymbol{U}_0$, the derivative of $\mathcal{S}_{\boldsymbol{F}}$ at $\boldsymbol{U}_0$, does not depend on $\boldsymbol{U}_0$ and $\boldsymbol{F}$, i.e.,

$$(\mathcal{S}'_{\boldsymbol{F}} \boldsymbol{U}_0)\boldsymbol{z} = \mathcal{S}_{\boldsymbol{0}} \boldsymbol{z} \; \forall \boldsymbol{z} \in \boldsymbol{L}^2(\Omega), \text{ or } \mathcal{S}'_{\boldsymbol{F}} \boldsymbol{U}_0 = \mathcal{S}_{\boldsymbol{0}}. \tag{5.8}$$

We then denote by $\mathcal{S}_{\boldsymbol{0}}^* = (\mathcal{S}'_{\boldsymbol{F}} \boldsymbol{U}_0)^* : \boldsymbol{W}(0,T)' \mapsto \boldsymbol{L}^2(\Omega)'$ the adjoint operator of $\mathcal{S}_{\boldsymbol{0}}$ or $\mathcal{S}'_{\boldsymbol{F}} \boldsymbol{U}_0$. That is,

$$\langle \boldsymbol{q}, \mathcal{S}_{\boldsymbol{F}} \boldsymbol{z} \rangle_{(\boldsymbol{W}(0,T)', \boldsymbol{W}(0,T))} = \langle \mathcal{S}_{\boldsymbol{0}}^* \boldsymbol{q}, \boldsymbol{z} \rangle_{(\boldsymbol{L}^2(\Omega)', \boldsymbol{L}^2(\Omega))} \; \forall (\boldsymbol{q}, \boldsymbol{z}) \in \boldsymbol{W}(0,T)' \times \boldsymbol{L}^2(\Omega). \tag{5.9}$$

Again, we derive the optimality condition of (5.7) by doing a calculus of variation

$$\begin{aligned} \langle J'(\boldsymbol{U}_0), \boldsymbol{z} \rangle &= (\gamma \boldsymbol{U}_0, \boldsymbol{z}) - \left(\widehat{\boldsymbol{U}} - \mathcal{S}_{\boldsymbol{F}} \boldsymbol{U}_0, \mathcal{S}'_{\boldsymbol{F}} \boldsymbol{U}_0 \boldsymbol{z}\right) \\ &= (\gamma \boldsymbol{U}_0, \boldsymbol{z}) - \left((\mathcal{S}'_{\boldsymbol{F}} \boldsymbol{U}_0)^* (\widehat{\boldsymbol{U}} - \mathcal{S}_{\boldsymbol{F}} \boldsymbol{U}_0), \boldsymbol{z}\right) \\ &= (\gamma \boldsymbol{U}_0, \boldsymbol{z}) - \left(\mathcal{S}_{\boldsymbol{0}}^* (\widehat{\boldsymbol{U}} - \mathcal{S}_{\boldsymbol{F}} \boldsymbol{U}_0), \boldsymbol{z}\right) = 0 \, \forall \, \boldsymbol{z} \in \boldsymbol{L}^2(\Omega). \end{aligned} \tag{5.10}$$

We remind readers that the operator $\mathcal{S}_{\boldsymbol{0}}^*$ acting on $\widehat{\boldsymbol{U}} - \mathcal{S}_{\boldsymbol{F}} \boldsymbol{U}_0$ is equivalent to solving the backward adjoint equation in (3.17) with source term $\widehat{\boldsymbol{U}} - \mathcal{S}_{\boldsymbol{F}} \boldsymbol{U}_0$, which exactly

gives $\boldsymbol{U}^*(\cdot,\mathbf{0})$. Also note that $\mathcal{S}_{\mathbf{0}}^*\widehat{\boldsymbol{U}}$ is a known variable; therefore, we can write (5.10) temporarily as

$$\gamma \boldsymbol{U}_0 + \mathcal{S}_{\mathbf{0}}^*\mathcal{S}_{\boldsymbol{F}}\boldsymbol{U}_0 = \mathcal{S}_{\mathbf{0}}^*\widehat{\boldsymbol{U}}. \tag{5.11}$$

For $\mathcal{S}_{\boldsymbol{F}}$, we decompose it as

$$\mathcal{S}_{\boldsymbol{F}}\boldsymbol{U}_0 = \mathcal{S}_{\boldsymbol{F}}\mathbf{0} + \mathcal{S}_{\mathbf{0}}\boldsymbol{U}_0. \tag{5.12}$$

Apparently $\mathcal{S}_{\mathbf{0}}$ is a linear operator and $\mathcal{S}_{\boldsymbol{F}}\mathbf{0}$ is a known variable, which finally allows us to rewrite (5.11) as

$$\gamma \boldsymbol{U}_0 + \mathcal{S}_{\mathbf{0}}^*\mathcal{S}_{\mathbf{0}}\boldsymbol{U}_0 = \mathcal{S}_{\mathbf{0}}^*\widehat{\boldsymbol{U}} - \mathcal{S}_{\mathbf{0}}^*\mathcal{S}_{\boldsymbol{F}}\mathbf{0}. \tag{5.13}$$

It is easy to show that the operator $\gamma + \mathcal{S}_{\mathbf{0}}^*\mathcal{S}_{\mathbf{0}}$ is positive definite. First, $(\gamma + \mathcal{S}_{\mathbf{0}}^*\mathcal{S}_{\mathbf{0}})^* = \gamma + \mathcal{S}_{\mathbf{0}}^*\mathcal{S}_{\mathbf{0}}$ is obviously true. Second, $\forall\, 0 \neq \boldsymbol{z} \in \boldsymbol{L}^2(\Omega)$, we have

$$((\gamma + \mathcal{S}_{\mathbf{0}}^*\mathcal{S}_{\mathbf{0}})\boldsymbol{z}, \boldsymbol{z}) = (\gamma\boldsymbol{z}, \boldsymbol{z}) + (\mathcal{S}_{\mathbf{0}}^*\mathcal{S}_{\mathbf{0}}\boldsymbol{z}, \boldsymbol{z}) = \gamma(\boldsymbol{z}, \boldsymbol{z}) + (\mathcal{S}_{\mathbf{0}}\boldsymbol{z}, \mathcal{S}_{\mathbf{0}}\boldsymbol{z}) > \gamma\|\boldsymbol{z}\|_0^2. \tag{5.14}$$

Hence we are able to write the optimality condition (5.10) in form of (5.6),

$$\mathcal{A}\boldsymbol{U}_0 = (\gamma + \mathcal{S}_{\mathbf{0}}^*\mathcal{S}_{\mathbf{0}})\boldsymbol{U}_0,\ b = \mathcal{S}_{\mathbf{0}}^*\widehat{\boldsymbol{U}} - \mathcal{S}_{\mathbf{0}}^*\mathcal{S}_{\boldsymbol{F}}\mathbf{0}. \tag{5.15}$$

We here clarify the operation of $\mathcal{A}$ acting on an element $\boldsymbol{z} \in \boldsymbol{L}^2(\Omega)$. First, recall that $\mathcal{S}_{\mathbf{0}}\boldsymbol{z}$ is the state solution of the Stokes–Darcy equation solved with initial $\boldsymbol{z}$ and source term $\mathbf{0}$. Second, the operator $\mathcal{S}_{\mathbf{0}}^*$ acting on $\mathcal{S}_{\mathbf{0}}\boldsymbol{z}$ gives the solution of the backward adjoint equation at $t = 0$ solved with initial $\mathbf{0}$ and source term $\mathcal{S}_{\mathbf{0}}\boldsymbol{z}$. Therefore, $\mathcal{A}\boldsymbol{z}$ is obtained by sequentially solving the following forward and backward equations:

$$\left\langle \frac{\partial \boldsymbol{\Phi}}{\partial t}, \boldsymbol{V} \right\rangle + a(\boldsymbol{\Phi}, \boldsymbol{V}) + b(\boldsymbol{V}, p) = \langle \mathbf{0}, \boldsymbol{V} \rangle \quad b(\boldsymbol{\Phi}, q) = 0,\ \boldsymbol{\Phi}(\cdot, 0) = \boldsymbol{z}, \tag{5.16}$$

$$-\left\langle \frac{\partial \boldsymbol{\Phi}^*}{\partial t}, \boldsymbol{V} \right\rangle + a^*(\boldsymbol{\Phi}^*, \boldsymbol{V}) + b(\boldsymbol{V}, p^*) = (\boldsymbol{\Phi}, \boldsymbol{V}) \quad b(\boldsymbol{\Phi}^*, q) = 0,\ \boldsymbol{\Phi}^*(\cdot, T) = \mathbf{0}, \tag{5.17}$$

$$\mathcal{A}\boldsymbol{z} = \boldsymbol{\Phi}^*(\cdot, 0) + \gamma\boldsymbol{z}. \tag{5.18}$$

At the discrete level, (5.16)–(5.18) are written in matrix-vector form as follows:

$$\frac{M_a}{\tau}\left(\begin{pmatrix}\vec{\psi}_h^{n+1}\\ \vec{\boldsymbol{v}}_h^{n+1}\\ \vec{q}_h^{n+1}\end{pmatrix} - \begin{pmatrix}\vec{\psi}_h^{n}\\ \vec{\boldsymbol{v}}_h^{n}\\ \vec{q}_h^{n}\end{pmatrix}\right) + S\begin{pmatrix}\vec{\psi}_h^{n+1}\\ \vec{\boldsymbol{v}}_h^{n+1}\\ \vec{q}_h^{n+1}\end{pmatrix} = \mathbf{0},\ \vec{\boldsymbol{\Phi}}_h^{\,0} = \vec{\boldsymbol{z}}_h = \begin{pmatrix}\vec{\psi}_h^{0(k)}\\ \vec{\boldsymbol{v}}_h^{0}\end{pmatrix}, \tag{5.19}$$

$$-\frac{M_a}{\tau}\left(\begin{pmatrix}\vec{\psi}_h^{*n+1}\\ \vec{\boldsymbol{v}}_h^{*n+1}\\ \vec{q}_h^{*n+1}\end{pmatrix} - \begin{pmatrix}\vec{\psi}_h^{*n}\\ \vec{\boldsymbol{v}}_h^{*n}\\ \vec{q}_h^{*n(k)}\end{pmatrix}\right) + S^*\begin{pmatrix}\vec{\psi}_h^{*n}\\ \vec{\boldsymbol{v}}_h^{*n}\\ \vec{q}_h^{*n}\end{pmatrix} = \begin{pmatrix}\vec{b}_{\phi_h^{n+1}}\\ \vec{b}_{\boldsymbol{u}_h^{n+1}}\\ \mathbf{0}\end{pmatrix},\ \vec{\boldsymbol{\Phi}}_h^{\,*N} = \begin{pmatrix}\vec{\psi}_h^{*N}\\ \vec{\boldsymbol{v}}_h^{*N}\end{pmatrix} = \mathbf{0}, \tag{5.20}$$

$$\overrightarrow{\mathcal{A}\boldsymbol{z}}_h = \vec{\boldsymbol{\Phi}}_h^{\,*0} + \gamma\vec{\boldsymbol{z}}_h, \tag{5.21}$$

where $(\vec{b}_{\psi_h^{n+1}}\, \vec{b}_{\boldsymbol{v}_h^{n+1}}\, \mathbf{0})^T = ([\int_{\Omega_p} \psi_h^{n+1}\psi_i dxdy]\,[\int_{\Omega_f} \boldsymbol{v}_h^{n+1}\boldsymbol{v}_i dxdy]\,\mathbf{0})^T$.

(5.16)–(5.18) or (5.19)–(5.21) provide the key information to update the CG iteration solving our data assimilation problem, which can be summarized in the following algorithm.

Remark 5.1. Note that for the CG method with exact line search, $\vec{\boldsymbol{r}}^{k+1}$ is basically the gradient of J_h at $\boldsymbol{U}_{0,h}^{(k)}$, i.e., $\vec{\boldsymbol{r}}^{k+1} = \nabla\vec{J}_h(\boldsymbol{U}_{0,h}^{(k)})$.

Algorithm 1 Conjugate gradient algorithm

Input: $\vec{\boldsymbol{U}}_{0,h}^{(0)}$ and ϵ;
Compute $\nabla \vec{J}_h(\boldsymbol{U}_{0,h}^{(0)})$, initialize $\vec{\boldsymbol{r}}_0 = -\nabla \vec{J}_h(\boldsymbol{U}_{0,h}^{(0)})$ and $\vec{\boldsymbol{d}}^0 = -\nabla \vec{J}_h(\boldsymbol{U}_{0,h}^{(0)})$, set error $= 1$, and start the iteration step $k = 0$;
while error $> \epsilon$ **do**
  Compute $\overrightarrow{\boldsymbol{d}}^k = \overrightarrow{\mathcal{A}\boldsymbol{d}}^k$ by solving (5.19)–(5.21) sequentially;
  Compute $\eta^k = \dfrac{\|\vec{\boldsymbol{r}}^k\|_0^2}{(\vec{\boldsymbol{d}}^k, \overrightarrow{\boldsymbol{d}}^k)_{M_{fp}}}$;
  Compute $\vec{\boldsymbol{r}}^{k+1} = \vec{\boldsymbol{r}}^k - \eta^k \overrightarrow{\mathcal{A}\boldsymbol{d}}^k$;
  Update $\vec{\boldsymbol{U}}_{0,h}^{(k+1)} = \vec{\boldsymbol{U}}_{0,h}^{(k)} + \eta^k \vec{\boldsymbol{d}}^k$;
  Compute $\beta^k = \dfrac{\|\vec{\boldsymbol{r}}^{k+1}\|_0^2}{\|\vec{\boldsymbol{r}}^k\|_0^2}$;
  Compute $\vec{\boldsymbol{d}}^{k+1} = \vec{\boldsymbol{r}}^{k+1} + \beta^k \vec{\boldsymbol{d}}^k$;
  Set $k = k+1$ and error $= \|\vec{\boldsymbol{r}}^{k+1}\|_0^2$;
end while
Output: $\vec{\boldsymbol{U}}_{0,h}^{(k+1)}$;

5.2. The BFGS method. The BFGS method is a type of quasi-Newton algorithm, which uses an approximated inverse Hessian operator D_k to determine the descent direction $\boldsymbol{d}^k$, i.e., $\boldsymbol{d}^k = -D_k J_h'(\boldsymbol{U}_{0,h})$, where $D_k \approx (J_h''(\boldsymbol{U}_{0,h}))^{-1}$. The iterates of the BFGS method behave similarly to or slightly less efficiently than CG for linear-quadratic optimization but outperform CG for the nonlinear cases. We introduce the BFGS method in this section as an additional option to solve the data assimilation of the Stokes–Darcy equation.

As mentioned, the inverse Hessian operator D_k is approximated in the BFGS method; this is because the calculation of the Hessian operator and its inverse are usually challenging or even impossible in practice due to the constraint complexity and high dimensional unknowns. We here present the following way to find out D_k pertaining to our problem:

$$D_k = \left(I - \theta^k(\boldsymbol{s}^k \otimes \boldsymbol{g}^k)\right) D_{k-1} \left(I - \theta^k(\boldsymbol{g}^k \otimes \boldsymbol{s}^k)\right) + \theta^k(\boldsymbol{s}^k \otimes \boldsymbol{s}^k), \tag{5.22}$$

where $\boldsymbol{s}^k = \boldsymbol{U}_{0,h}^{(k)} - \boldsymbol{U}_{0,h}^{(k-1)}$, $\boldsymbol{g}^k = \nabla J_h(\boldsymbol{U}_{0,h}^{(k)}) - \nabla J_h(\boldsymbol{U}_{0,h}^{(k-1)})$, and $\theta^k = \frac{1}{(\boldsymbol{s}^k, \boldsymbol{g}^k)}$. The operator $\otimes$ is defined as

$$(\boldsymbol{a} \otimes \boldsymbol{b})\boldsymbol{c} = (\boldsymbol{b}, \boldsymbol{c})\boldsymbol{a}, \quad \text{for } \boldsymbol{a}, \boldsymbol{b} \in \boldsymbol{L}^2(\Omega),\ \forall \boldsymbol{c} \in \boldsymbol{L}^2(\Omega). \tag{5.23}$$

The update of D_k in (5.22) is based on a continuity assumption of the second order derivative of the cost functional. We try to search a bounded operator D_k that is close to the previous D_{k-1} in the sense of a weighted Hilbert–Schmidt norm [50, 86]:

$$\min_{D \in \mathscr{L}(\boldsymbol{L}^2(\Omega), \boldsymbol{L}^2(\Omega))} \frac{1}{2} \left\| G^{\frac{1}{2}} (D - D_{k-1}) G^{\frac{1}{2}} \right\|_{HS} \quad \text{subject to } D\boldsymbol{g}^k = \boldsymbol{s}^k.$$

Here, G is a weighted operator satisfying $G\boldsymbol{s}^k = \boldsymbol{g}^k$, and the constraint $D\boldsymbol{g}^k = \boldsymbol{s}^k$ comes from a secant approximation of the second order derivative of $J_h(\boldsymbol{U}_{0,h})$ at $\boldsymbol{U}_{0,h}^{(k)}$.

Algorithm 2 BFGS algorithm

Input: $\vec{\boldsymbol{U}}_{0,h}^{(0)}$, a positive definite matrix D_0, and ϵ;
Compute $\nabla \vec{J}_h(\boldsymbol{U}_{0,h}^{(0)})$, the first descent direction $-D_0\nabla \vec{J}_h(\boldsymbol{U}_{0,h}^{(0)})$, and the first update $\vec{\boldsymbol{U}}_{0,h}^{(1)} = \vec{\boldsymbol{U}}_{0,h}^{(0)} - D_0\nabla \vec{J}(\boldsymbol{U}_{0,h}^{(0)})$. Set error $= 1$ and start the iteration step $k = 1$;

while error $> \epsilon$ **do**
Compute $\nabla \vec{J}_h(\boldsymbol{U}_{0,h}^{(k)})$;
Compute $\vec{\boldsymbol{s}}^k = \vec{\boldsymbol{U}}_{0,h}^{(k)} - \vec{\boldsymbol{U}}_{0,h}^{(k-1)}$, $\vec{\boldsymbol{g}}^k = \nabla \vec{J}_h(\boldsymbol{U}_{0,h}^{(k)}) - \nabla \vec{J}_h(\boldsymbol{U}_{0,h}^{(k-1)})$;
Compute $D_k = \left(I - \frac{\vec{\boldsymbol{s}}^k (M_{fp}^T \vec{\boldsymbol{g}}^k)^T}{\vec{\boldsymbol{s}}^{kT} M_{fp} \vec{\boldsymbol{g}}^k}\right) D_{k-1} \left(I - \frac{\vec{\boldsymbol{g}}^k (M_{fp}^T \vec{\boldsymbol{s}}^k)^T}{\vec{\boldsymbol{s}}^{kT} M_{fp} \vec{\boldsymbol{g}}^k}\right) + \frac{\vec{\boldsymbol{s}}^k (M_{fp}^T \vec{\boldsymbol{s}}^k)^T}{\vec{\boldsymbol{s}}^{kT} M_{fp} \vec{\boldsymbol{g}}^k}$;
Update $\vec{\boldsymbol{U}}_{0,h}^{(k+1)} = \vec{\boldsymbol{U}}_{0,h}^{(k)} - \eta^k D_k \nabla \vec{J}_h(\boldsymbol{U}_{0,h}^{(k)})$;
Set $k = k + 1$ and error $= \|\nabla \vec{J}_h(\boldsymbol{U}_{0,h}^{(k)})\|_0^2$;
end while
Output: $\vec{\boldsymbol{U}}_{0,h}^{(k+1)}$;

The BFGS algorithm can be briefly described as follows:

- Initialize $\boldsymbol{U}_{0,h}^{(0)}$ and a bounded positive definite operator D_0.
- Update $\boldsymbol{U}_{0,h}^{(k+1)} = \boldsymbol{U}_{0,h}^{(k)} - \eta^k D_k \nabla J_h(\boldsymbol{U}_{0,h}^{(k)})$ with

$$D_k = \begin{cases} D_0, & k = 0, \\ \left(I - \theta^k(\boldsymbol{s}^k \otimes \boldsymbol{g}^k)\right) D_{k-1} \left(I - \theta^k(\boldsymbol{g}^k \otimes \boldsymbol{s}^k)\right) + \theta^k(\boldsymbol{s}^k \otimes \boldsymbol{s}^k), & k \geq 1. \end{cases}$$

- η^k can be simply chosen as $\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots\}$, or determined by using exact line search if possible, or by using inexact line search methods, such as the Armijo and Wolfe condition [50, 55].

Next, we show how to explicitly compute the matrix form of the operator D_k in coding implementation. Based on the definition in (5.23), for $\boldsymbol{a}, \boldsymbol{b} \in \boldsymbol{L}^2(\Omega)$, we deduce

$$(\boldsymbol{a} \otimes \boldsymbol{b})\boldsymbol{c} = (\boldsymbol{b}, \boldsymbol{c})\boldsymbol{a} = \vec{\boldsymbol{b}}^T M_{fp} \vec{\boldsymbol{c}} \boldsymbol{a} = (M_{fp}^T \vec{\boldsymbol{b}})^T \vec{\boldsymbol{c}} \boldsymbol{a} = \vec{\boldsymbol{a}} (M_{fp}^T \vec{\boldsymbol{b}})^T \boldsymbol{c} \ \forall \boldsymbol{c} \in \boldsymbol{L}^2(\Omega). \tag{5.24}$$

Therefore, the matrix representation of $\boldsymbol{a} \otimes \boldsymbol{b}$ acting on $\boldsymbol{L}^2(\Omega)$ is $\vec{\boldsymbol{a}}(M_{fp}^T \vec{\boldsymbol{b}})^T$. Meanwhile, the calculation of θ^k is straightforward:

$$\theta^k = \frac{1}{(\boldsymbol{s}^k, \boldsymbol{g}^k)} = \frac{1}{\vec{\boldsymbol{s}}^{kT} M_{fp} \vec{\boldsymbol{g}}^k}.$$

Now we can rewrite (5.22) or the matrix representation of D_k, still denoted as D_k, in the form

$$D_k = \left(I - \frac{\vec{\boldsymbol{s}}^k (M_{fp}^T \vec{\boldsymbol{g}}^k)^T}{\vec{\boldsymbol{s}}^{kT} M_{fp} \vec{\boldsymbol{g}}^k}\right) D_{k-1} \left(I - \frac{\vec{\boldsymbol{g}}^k (M_{fp}^T \vec{\boldsymbol{s}}^k)^T}{\vec{\boldsymbol{s}}^{kT} M_{fp} \vec{\boldsymbol{g}}^k}\right) + \frac{\vec{\boldsymbol{s}}^k (M_{fp}^T \vec{\boldsymbol{s}}^k)^T}{\vec{\boldsymbol{s}}^{kT} M_{fp} \vec{\boldsymbol{g}}^k}. \tag{5.25}$$

We summarize the BFGS iterative algorithm as follows.

5.3. The steepest descent method. The CG and BFGS methods can display a fast convergence rate and solve the discrete optimality system (4.11) effectively in most cases. However, their descent directions are sort of sensitive to the stability of

Algorithm 3 Steepest descent algorithm

Input: $\eta^k, \vec{\boldsymbol{U}}_{0,h}^{(0)}$, and ϵ;
Set error$=1$ and $k=1$;

while error $>\epsilon$ **do**
Compute $\nabla \vec{J}_h(\boldsymbol{U}_{0,h}^{(k)}) = \gamma \vec{\boldsymbol{U}}_{0,h}^{(k)} - \vec{\boldsymbol{U}}_h^{*0(k)}$ via sequentially solving (5.2)–(5.3) forward and backward;
Update $\vec{\boldsymbol{U}}_{0,h}^{(k+1)} = \vec{\boldsymbol{U}}_{0,h}^{(k)} - \eta^k \nabla \vec{J}_h(\boldsymbol{U}_{0,h}^{(k)})$;
Set $k=k+1$ and error $= \|\nabla \vec{J}_h(\boldsymbol{U}_{0,h}^{(k)})\|_0^2$;
end while
Output: $\vec{\boldsymbol{U}}_{0,h}^{(k+1)}$;

the data assimilation problem, which can hinder the convergence for a problem with extreme low stability that might be caused by a small regularization parameter γ in the cost functional (4.2). This shortcoming motivates us to propose an SD method [55] that gains more stability at the cost of a lower convergence rate.

Recalling (4.11) or (5.1),

$$\nabla J_h\left(\boldsymbol{U}_{0,h}^{(k)}\right) = \gamma \boldsymbol{U}_{0,h}^{(k)} - \boldsymbol{U}_h^{*0(k)} \tag{5.26}$$

is the gradient of J_h at $\boldsymbol{U}_{0,h}^{(k)}$ of the kth iteration.

Based on (5.26) we can illustrate the SD method as follows:

- Initialize $\boldsymbol{U}_{0,h}^{(0)}$.
- Compute $\nabla J_h(\boldsymbol{U}_{0,h}^{(k)}) = \gamma \boldsymbol{U}_{0,h}^{(k)} - \boldsymbol{U}_h^{*0(k)}$.
- Update $\boldsymbol{U}_{0,h}^{(k+1)} = \boldsymbol{U}_{0,h}^{(k)} - \eta^k(\gamma \boldsymbol{U}_{0,h}^{(k)} - \boldsymbol{U}_h^{*0(k)})$.
- η^k can be simply chosen as $\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots\}$, or determined by using exact line search if possible, or determined with inexact line search methods, such as the Armijo and Wolfe condition [50, 55].

The SD method shares the simplest iteration update and the best stability behavior among almost all of the gradient-based methods, which makes it popular for a lot of optimization problems. Of course, such a benefit is accompanied by a slower convergence speed, especially for problems with a lower stability.

We summarize this SD iteration as follows.

Remark 5.2. The superlinear, linear, or sublinear convergence rates for the CG, BFGS, and SD methods are not universally guaranteed. The actual rate strongly relies on the conditioning number $\kappa(\mathcal{A})$, where $\mathcal{A}$ is defined in (5.15). Although the CG and BFGS methods have a lot of similarities, they are good at solving different problems. In general, they both converge faster than the SD method, which is more stable.

6. Numerical experiments. This section presents numerical results to demonstrate the optimal convergence established in section 4 and the performance of the state prediction using the algorithms developed in section 5. The Taylor–Hood finite elements are utilized for the spatial discretization of the Stokes equation and the quadratic finite elements are utilized for the hydraulic head of the Darcy equation.

6.1. Verification of the finite element convergence rate. In this example, we let $\mathbb{K}=\mathbb{I}$, $\alpha=1, g=1$, $\Omega_p=(0,\pi)\times(0,1)$, $\Omega_f=(0,\pi)\times(-1,0)$, $\Gamma: x=0$, and

$\boldsymbol{U}|_{\partial\Omega} = 0$. Based on the numerical example in [37], whose analytic solutions satisfy the Beavers–Joseph interface conditions, we choose the following initial functions and source term functions:

$$\begin{aligned}
\boldsymbol{W}_0 &= ((2-\pi\sin(\pi x))(-y+\cos(\pi(1-y))), x^2y^2+e^{-y}, (-2/3)xy^3+2-\pi\sin(\pi x))^T,\\
f_p &= \cos(2\pi t)(\pi^2(2\cos(\pi(1-y))-2\pi\sin(\pi x)\cos(\pi(1-y))+\pi y\sin(\pi x)))\\
&\quad -2\pi\sin(2\pi t)(2-\pi\sin(\pi x))(-y+\cos(\pi(1-y))),\\
f_1 &= \cos(2\pi t)(-2y^2-2x^2-e^{-y}+\pi^2\cos(\pi x)\cos(2\pi y))\\
&\quad -2\pi\sin(2\pi t)(x^2y^2+e^{-y})\sin(2\pi t)(-2\pi),\\
f_2 &= \cos(2\pi t)(4xy-\pi^3\sin(\pi x)+2\pi(2-\pi\sin(\pi x))\sin(2\pi y))\\
&\quad -2\pi\sin(2\pi t)\left(\frac{2}{3}xy^3+2-\pi\sin(\pi x)\right).
\end{aligned}$$

To have a set of smooth observation data satisfying both the interface conditions and homogeneous boundary conditions, we numerically solve the Stokes–Darcy model with $h = 1/64$, $\tau = 1/4000$, initial function $\boldsymbol{W}_0$, and source term $\boldsymbol{F} = (f_p, f_1, f_2)^T$ in the time interval $[0, 0.75]$. Then the numerical solution in the time interval $(0.25, 0.75]$ is considered as the observation data $\widehat{\boldsymbol{U}}$. The solution at $t = 0.25$ is the $\boldsymbol{U}_0$ we intend to reconstruct.

For the data assimilation problem, we use the mesh sizes of 1/8, 1/16, 1/32, 1/64 and time step sizes of 1/16, 1/128, 1/1024, 1/4000 to produce numerical solutions, based on the CG method with stopping criteria $\epsilon = 10^{-5}$. For each γ, the numerical solution with $h = 1/64, \tau = 1/4000$ is considered to replace the analytical solution when computing the numerical errors. Tables 6.1–6.3 illustrate the convergence performance. From Table 6.1 and Table 6.2, we can see that the $\boldsymbol{L}^2$ norm errors for ϕ and $\boldsymbol{u}$ appear to converge optimally. In addition, the relative errors in Table 6.3 become larger when γ decreases, which is consistent with the conclusion that the coefficient $C_{\gamma,\Omega,\Gamma,T}$ in Theorem 4.10 is proportional to $\frac{1}{\gamma}$.

6.2. Iterative data assimilation methods. We first investigate the three iterative methods developed in section 5. Let $\mathbb{K} = \begin{pmatrix} 0.8 & 0 \\ 0 & 1.2 \end{pmatrix}$, $\nu = 1.2$, $\alpha = 0.12$, $g = 9.8$, $\Omega_p = (0,\pi)\times(0,1)$, $\Omega_f = (0,\pi)\times(-1,0)$, $\Gamma : x = 0$, $\boldsymbol{U}|_{\partial\Omega} = 0$, and

TABLE 6.1
The finite element convergence rate of the recovered initial condition ϕ_0.

γ	$\|\|\phi_0 - \phi_{0,\frac{1}{8}}\|\|_0$	$\|\|\phi_0 - \phi_{0,\frac{1}{16}}\|\|_0$	Rate	$\|\|\phi_0 - \phi_{0,\frac{1}{32}}\|\|_0$	Rate
1	1.16×10^{-2}	1.30×10^{-3}	3.15	1.20×10^{-4}	3.44
$\frac{1}{5}$	4.19×10^{-2}	4.90×10^{-3}	3.09	4.70×10^{-4}	3.38
$\frac{1}{50}$	9.83×10^{-2}	1.22×10^{-2}	3.01	1.30×10^{-3}	3.23
$\frac{1}{200}$	1.14×10^{-1}	1.41×10^{-2}	3.02	1.50×10^{-3}	3.23

TABLE 6.2
The finite element convergence rate of the recovered initial condition $\boldsymbol{u}_0$.

γ	$\|\|\boldsymbol{u}_0 - \boldsymbol{u}_{0,\frac{1}{8}}\|\|_0$	$\|\|\boldsymbol{u}_0 - \boldsymbol{u}_{0,\frac{1}{16}}\|\|_0$	Rate	$\|\|\boldsymbol{u}_0 - \boldsymbol{u}_{0,\frac{1}{32}}\|\|_0$	Rate
1	2.10×10^{-3}	2.42×10^{-4}	3.12	2.91×10^{-5}	3.06
$\frac{1}{5}$	1.03×10^{-2}	1.25×10^{-3}	3.04	1.41×10^{-4}	3.15
$\frac{1}{50}$	4.84×10^{-2}	6.00×10^{-3}	3.01	6.64×10^{-4}	3.17
$\frac{1}{200}$	6.69×10^{-2}	8.65×10^{-3}	2.95	9.81×10^{-4}	3.14

TABLE 6.3
Relative errors according to γ, $R_h^{\phi_0} = \frac{||\phi_0 - \phi_{0,h}||_0}{||\phi_0||_0}$, *and* $R_h^{\boldsymbol{u}_0} = \frac{||\boldsymbol{u}_0 - \boldsymbol{u}_{0,h}||_0}{||\boldsymbol{u}_0||_0}$.

γ	$R^{\phi_0}_{\frac{1}{8}}$	$R^{\phi_0}_{\frac{1}{16}}$	$R^{\phi_0}_{\frac{1}{32}}$	$R^{\boldsymbol{u}_0}_{\frac{1}{8}}$	$R^{\boldsymbol{u}_0}_{\frac{1}{16}}$	$R^{\boldsymbol{u}_0}_{\frac{1}{32}}$
1	0.2184	0.0247	0.0023	0.0972	0.0119	0.0013
$\frac{1}{5}$	0.2341	0.0278	0.0027	0.1167	0.0143	0.0016
$\frac{1}{50}$	0.2496	0.0314	0.0033	0.1649	0.0208	0.0023
$\frac{1}{200}$	0.2524	0.0313	0.0034	0.1811	0.0236	0.0027

$$\boldsymbol{F} = \Big(\sin(2\pi t)(\pi\sin(x) + 3 + 2x(\pi - x)),\, \sin(2\pi t)(x^2 + y + \cos(y) + 2),$$
$$\sin(2\pi t)(\sin(y) + 2x + y + 5/2)\Big)^T,$$
$$\boldsymbol{W}_0 = \big(x(\pi - x)y(1 - y),\, x(\pi - x)y(1 - y)\sin(\pi x),\, 4x(\pi - x)y(1 - y)\sin(\pi x)\big)^T.$$

To construct observation data, we numerically solve the Stokes–Darcy model with initial function $\boldsymbol{W}_0$ and source term $\boldsymbol{F} = (f_p, f_1, f_2)^T$ in the time interval $[0, 2]$, and then take 100 snapshots from the numerical solution in the time interval $(1, 2]$ uniformly as the analytical solution $\boldsymbol{U}$. The observation $\widehat{\boldsymbol{U}}$ is produced by adding noise with multivariate Gaussian distribution $\mathcal{N}(0, I/50)$ to $\boldsymbol{U}$. The solution at $t = 1$ is the initial condition we intend to recover and will be used for the state forecast. The $\widetilde{\boldsymbol{L^2}}(0,T)$ and $\widetilde{\boldsymbol{L^\infty}}(0,T)$ norms, which are defined as $\|\boldsymbol{U} - \boldsymbol{U}_h\|_{\widetilde{\boldsymbol{L^2}}} = \sqrt{\sum_{n=2}^{N+1} \tau \frac{\|\boldsymbol{U}^n - \boldsymbol{U}_h^n\|_0^2}{\|\boldsymbol{U}^n\|_0^2}}$ and $\|\boldsymbol{U} - \boldsymbol{U}_h\|_{\widetilde{\boldsymbol{L^\infty}}} = \sum_{n=2}^{N+1} \tau \frac{\|\boldsymbol{U}^n - \boldsymbol{U}_h^n\|_{\boldsymbol{L}^\infty(\Omega)}}{\|\boldsymbol{U}^n\|_{\boldsymbol{L}^\infty(\Omega)}}$, are used to estimate the data assimilation accuracy. In the rest of paper, without special comment, $\epsilon = 10^{-5}$ will be the stopping criteria for all iterative methods, and the descent step size $\eta^k = 1$ is considered for the BFGS and SD methods.

First, Table 6.4 shows that an accurate state forecast is achieved for various values of γ. As γ decreases, the stability of the data assimilation problem decreases or the conditioning number $\kappa(\mathcal{A})$ in (5.15) increases, based on Theorem 3.2. Hence, the convergence of all iterative methods will become slower and slower; this can be verified by the increasing number of iterations in Table 6.4. The number of iterations used for each method indicates that the CG and BFGS methods converge much faster than the SD method, especially for small γ. The convergence rate can also be visualized via Figure 6.1. The CG method achieves the superlinear or nearly linear convergence rates for different regularization γ. The BFGS method behaves similarly but with a slower convergence rate. And the SD method shows the slowest sublinear convergence rate especially when γ is small. However, the SD method is always monotonically

TABLE 6.4
Data assimilation result: the $\widetilde{\boldsymbol{L^2}}$- *and* $\widetilde{\boldsymbol{L^\infty}}$-*norm errors between* $\boldsymbol{U}$ *and the numerical solution* $\boldsymbol{U}_h$.

γ	The CG method		The BFGS method		The SD method	
	$\|\|\boldsymbol{U}_h - \boldsymbol{U}\|\|_{\widetilde{\boldsymbol{L^2}}}$	$\|\|\boldsymbol{U}_h - \boldsymbol{U}\|\|_{\widetilde{\boldsymbol{L^\infty}}}$	$\|\|\boldsymbol{U}_h - \boldsymbol{U}\|\|_{\widetilde{\boldsymbol{L^2}}}$	$\|\|\boldsymbol{U}_h - \boldsymbol{U}\|\|_{\widetilde{\boldsymbol{L^\infty}}}$	$\|\|\boldsymbol{U}_h - \boldsymbol{U}\|\|_{\widetilde{\boldsymbol{L^2}}}$	$\|\|\boldsymbol{U}_h - \boldsymbol{U}\|\|_{\widetilde{\boldsymbol{L^\infty}}}$
$\frac{1}{10}$	0.327670	0.2087	0.32767	0.2087	0.327670	0.2087
$\frac{1}{100}$	0.05384	0.03838	0.053832	0.03838	0.05384	0.03838
$\frac{1}{1000}$	0.00596	0.00466	0.005941	0.00465	0.005938	0.00465
$\frac{1}{5000}$	0.00113	0.0008795	0.001126	0.0008914	0.001305	0.0009358
$\frac{1}{20000}$	0.0004604	0.0003047	0.0004744	0.0003154	0.0004866	0.0003456

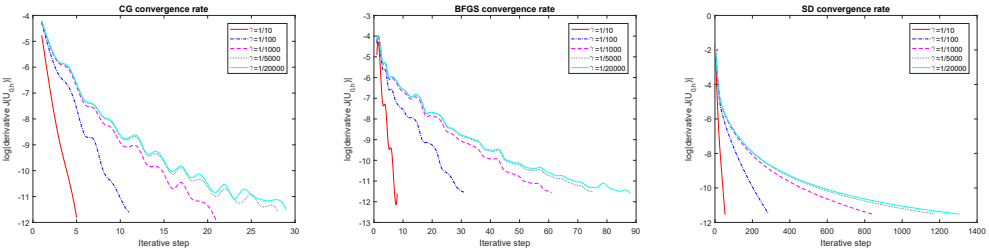

FIG. 6.1. *Convergence rate comparison for CG, BFGS, and SD methods with different* γ.

TABLE 6.5

Comparison of the convergence rate and computational cost for the CG, BFGS, and SD methods; # iteration = number of iteration, # PDE solving = total number of PDE solvings.

	The CG method		The BFGS method		The SD method	
γ	# iteration	# PDE solving	# iteration	# PDE solving	# iteration	# PDE solving
$\frac{1}{10}$	7	16	11	24	56	112
$\frac{1}{100}$	13	28	34	70	289	578
$\frac{1}{1000}$	23	48	64	130	841	1682
$\frac{1}{5000}$	30	62	78	158	1178	2356
$\frac{1}{20000}$	31	64	91	184	1306	2612

decreasing, while the CG and BFGS methods both have oscillations when approaching the minima for small γ. This verifies the stability advantage of the SD method.

The convergence rate provides information about the computation cost of each method, but very implicitly here. This is because the iteration cost of each method is not the same due to different update procedures. To clearly compare the computation efficiency, we notice that the PDE solving at each iteration is the main contribution to the cost; therefore, we use the total number of PDE solvings to approximately evaluate the computation cost. In Table 6.5, the CG method requires the small number of PDE solvings, the BFGS method asks for a moderate number of PDE solvings, and the SD method needs a large number of PDE solvings.

Another important convergence property of all iterative methods is that their convergence rate is not sensitive to the mesh size h in our case, i.e., when the mesh size is refined or the number of unknowns increases, the number of iterations does not increase. We prove this statement by choosing a different mesh size in Figure 6.2, where the convergence rate does not change much when refining h from $\frac{1}{16}$ to $\frac{1}{48}$.

Last but not least, besides the convergence rate, the choice of stopping criteria ϵ also plays a critical role in balancing the computational cost. As shown in Figure 6.1, the convergence speed slows down quickly as the derivative norm approaches a small number, in which the computation cost is consumed dramatically. This situation is getting even worse for the sublinear convergence method. In practice, a very small ϵ is unnecessary, since it may not help improve the data assimilation accuracy but only increase the computation cost greatly.

In our case, based on the observation availability and quality, we empirically assume that $\gamma = 1/20000$ is the regularization parameter to obtain the best assimilation result. In Table 6.6, we observe that the stopping criteria $\epsilon = 10^{-2}$ and 10^{-4} are not small enough to be used. On the other hand, compared to $\epsilon = 10^{-5}$, the simulation results based on $\epsilon = 10^{-6}$ and 10^{-7} are not improving noticeably anymore but are

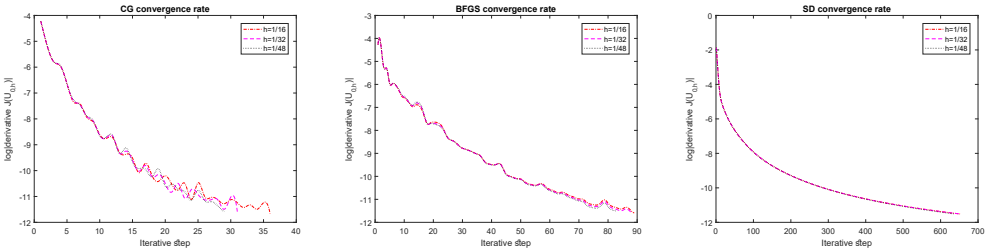

FIG. 6.2. *Convergence rate comparison of different mesh size h for CG, BFGS, and SD methods.*

TABLE 6.6

Comparison of the computational cost and data assimilation results of different stop criteria for the CG, BFGS, and SD methods, # Ite/PDE:= the number of iterations and the number of PDE solvings.

	The CG method		The BFGS method		The SD method	
ϵ	$\|\|\boldsymbol{U}_h - \boldsymbol{U}\|\|_{\widetilde{\boldsymbol{L^2}}}$	# Ite/PDE	$\|\|\boldsymbol{U}_h - \boldsymbol{U}\|\|_{\widetilde{\boldsymbol{L^2}}}$	# Ite/PDE	$\|\|\boldsymbol{U}_h - \boldsymbol{U}\|\|_{\widetilde{\boldsymbol{L^2}}}$	# Ite/PDE
10^{-2}	0.01493	4/10	0.02073	6/14	0.04003	13/26
10^{-4}	0.0003918	15/32	0.0006852	40/82	0.0007870	195/390
10^{-5}	0.0004899	33/68	0.0004898	91/184	0.0004866	654/1308
10^{-6}	0.0004885	83/168	0.0004875	203/408	0.0004872	3600/7200
10^{-7}	0.0004877	144/290	0.0004875	332/666	0.0004872	15343/30686

consuming significantly more iterations, especially for the SD method. This observation suggests using $\epsilon = 10^{-5}$ as the stopping criteria in our problem.

6.3. Data assimilation performance. In this section, we focus on verifying the proposed data assimilation methods by testing against a more practical case. We consider the situation in which observation is available only in limited windows. We will use the following four observation windows for experimental tests: observation window 1 (O1): $[0,\pi]\times[-0.75,0]$, $[0,\pi]\times[0,0.75]$; O2: $[0,\pi]\times[-1,-0.25]$, $[0,\pi]\times[0.25,1]$; O3: $[0,\pi]\times[-0.5,0]$, $[0,\pi]\times[0,0.5]$; O4: $[0,\pi]\times[-1,-0.5]$, $[0,\pi]\times[0.5,1]$. Set $\mathbb{K}=\begin{pmatrix}0.06 & 0\\ 0 & 0.08\end{pmatrix}$, $\nu=0.2$, $\alpha=0.28$. Other parameters in the Stokes–Darcy model, such as g,Ω_p,Ω_f, Γ, and $\boldsymbol{F}$, are the same as in subsection 6.2. In each scenario, the observation is also provided the same way as in subsection 6.2 with Gaussian noise $\mathcal{N}(0,I/50)$.

Using the given model parameters ν, $\mathbb{K}$, and α, we alter the observation windows to compute numerical results with the proposed data assimilation methods. Assume $\gamma=1/20000$ is the regularization parameter to obtain the best assimilation result for all cases. In Table 6.7, errors between data assimilation results $\boldsymbol{U}_h$ and the analytical solution $\boldsymbol{U}$ for all scenarios show that the methods proposed in this paper are applicable to assimilate the state solution in general. Moreover, even the observational data is only partially provided; the proposed methods can still provide useful predictions on the entire domain. We also notice that the observations from different windows have different impact on the assimilation performance.

In Table 6.8, the total number of iterations and PDE solvings used for each method confirms again the convergence speed of CG, BFGS, and SD methods, respectively. Overall, we believe that the CG method should take priority for most of the data assimilation scenarios; the CG method is best for most applied to some special cases such as nonlinear models, and the SD method is a backup for dealing with extreme ill-conditioning problems.

Table 6.7
Data assimilation result for fixed model parameters with different observation windows.

Obw	The CG method		The BFGS method		The SD method	
	$\|\|\boldsymbol{U}_h - \boldsymbol{U}\|\|_{\widetilde{L^2}}$	$\|\|\boldsymbol{U}_h - \boldsymbol{U}\|\|_{\widetilde{L^\infty}}$	$\|\|\boldsymbol{U}_h - \boldsymbol{U}\|\|_{\widetilde{L^2}}$	$\|\|\boldsymbol{U}_h - \boldsymbol{U}\|\|_{\widetilde{L^\infty}}$	$\|\|\boldsymbol{U}_h - \boldsymbol{U}\|\|_{\widetilde{L^2}}$	$\|\|\boldsymbol{U}_h - \boldsymbol{U}\|\|_{\widetilde{L^\infty}}$
O1	0.001439	0.002177	0.001440	0.002164	0.002409	0.003065
O2	0.002461	0.002969	0.002641	0.003219	0.003041	0.003839
O3	0.01155	0.01508	0.011385	0.014879	0.01606	0.01960
O4	0.007655	0.01045	0.0079556	0.010392	0.009523	0.014243

Table 6.8
The number of iterations and PDE solvings used for the CG, BFGS, and SD method. # iteration=number of iteration, # PDE solving= total number of PDE solving.

Obw	The CG method		The BFGS method		The SD method	
	# Iteration	# PDE solving	# Iteration	# PDE solving	# Iteration	# PDE solving
O1	103	208	193	388	1038	2076
O2	107	216	202	406	1824	3648
O3	124	250	226	454	2220	4440
O4	90	182	194	390	2673	5346

7. Conclusion. In this paper, we proposed a variational method to solve a data assimilation problem of the Stokes–Darcy model by using iterative algorithms. The well-posedness of this problem was rigorously analyzed. For numerical computations, we proposed a discretization using FEM and the backward Euler scheme, and analyzed its convergence properties. The rescaling and auxiliary techniques played key roles in proving the optimal convergence rate of the proposed numerical scheme. Three decoupled iterative numerical algorithms, the CG method, the BFGS method, and the SD method, were developed to reduce the computational cost. Various examples were used to validate the proposed methods.

REFERENCES

[1] H. D. I. Abarbanel, P. J. Rozdeba, and S. Shirman, *Machine learning: Deepest learning as statistical data assimilation problems*, Neural Comput., 30 (2018), pp. 2025–2055, https://doi.org/10.1162/neco_a_01094.

[2] V. I. Agoshkov, E. I. Parmuzin, and V. P. Shutyaev, *A numerical algorithm of variational data assimilation for reconstruction of salinity fluxes on the ocean surface*, Russian J. Numer. Anal. Math. Model., 23 (2008), pp. 135–161.

[3] A. Apte, C. K. R. T. Jones, A. M. Stuart and J. Voss, *Data assimilation: Mathematical and statistical perspectives*, Internat. J. Numer. Methods Fluids, 56 (2008), pp. 1033–1046.

[4] T. Arbogast and D. S. Brunson, *A computational method for approximating a Darcy-Stokes system governing a vuggy porous medium*, Comput. Geosci., 11 (2007), pp. 207–218.

[5] T. Arbogast, M. A. Hesse, and A. L. Taicher, *Mixed methods for two-phase Darcy-Stokes mixtures of partially melted materials with regions of zero porosity*, SIAM J. Sci. Comput., 39 (2017), pp. B375–B402.

[6] M. G. Armentano and M. L. Stockdale, *Approximations by mini mixed finite element for the Stokes–Darcy coupled problem on curved domains*, Int. J. Numer. Anal. Model., 18 (2021), pp. 203–234.

[7] D. Auroux and M. Nodet, *The back and forth nudging algorithm for data assimilation problems: Theoretical results on transport equations*, ESAIM Control Optim. Calc. Var., 18 (2012), pp. 318–342, https://doi.org/10.1051/cocv/2011004.

[8] A. Azouani, E. Olson, and E. S. Titi, *Continuous data assimilation using general interpolant observables*, J. Nonlinear Sci., 24 (2014), pp. 277–304, https://doi.org/10.1007/s00332-013-9189-y.

[9] L. Badea, M. Discacciati, and A. Quarteroni, *Numerical analysis of the Navier-Stokes/Darcy coupling*, Numer. Math., 115 (2010), pp. 195–227.
[10] G. Beavers and D. Joseph, *Boundary conditions at a naturally permeable wall*, J. Fluid Mech., 30 (1967), pp. 197–207.
[11] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Data assimilation in reduced modeling*, SIAM/ASA J. Uncertain. Quantif., 5 (2017), pp. 1–29, https://doi.org/10.1137/15M1025384.
[12] J. Blum, F.-X. L. Dime, and I. M. Navon, *Data assimilation for geophysical fluids*, in Handbook of Numerical Analysis. Special Volume: Computational Methods for the Atmosphere and the Oceans, Handb. Numer. Anal. 14, Elsevier/North-Holland, Amsterdam, 2009, pp. 385–441.
[13] Y. Boubendir and S. Tlupova, *Domain decomposition methods for solving Stokes–Darcy problems with boundary integrals*, SIAM J. Sci. Comput., 35 (2013), pp. B82–B106.
[14] A. Brandt and L. Y. Zaslavsky, *Multiscale algorithm for atmospheric data assimilation*, SIAM J. Sci. Comput., 18 (1997), pp. 949–956.
[15] C. H. Bruneau, P. Fabrie, and F. Veersé, *Optimal control data assimilation with an atmospheric model*, Numer. Funct. Anal. Optim., 18 (1997), pp. 691–722.
[16] M. Cai, M. Mu, and J. Xu, *Numerical solution to a mixed Navier-Stokes/Darcy model by the two-grid approach*, SIAM J. Numer. Anal., 47 (2009), pp. 3325–3338.
[17] Y. Cao, M. Gunzburger, X.-M. He, and X. Wang, *Robin-Robin domain decomposition methods for the steady Stokes–Darcy model with Beaver-Joseph interface condition*, Numer. Math., 117 (2011), pp. 601–629.
[18] Y. Cao, M. Gunzburger, X.-M. He, and X. Wang, *Parallel, non-iterative, multi-physics domain decomposition methods for time-dependent Stokes–Darcy systems*, Math. Comp., 83 (2014), pp. 1617–1644.
[19] Y. Cao, M. Gunzburger, X. Hu, F. Hua, X. Wang, and W. Zhao, *Finite element approximation for Stokes–Darcy flow with Beavers-Joseph interface conditions*, SIAM J. Numer. Anal., 47 (2010), pp. 4239–4256.
[20] Y. Cao, M. Gunzburger, F. Hua, and X. Wang, *Coupled Stokes–Darcy model with Beavers-Joseph interface boundary condition*, Commun. Math. Sci., 8 (2010), pp. 1–25.
[21] A. Çeşmelioğlu and B. Rivière, *Primal discontinuous Galerkin methods for time-dependent coupled surface and subsurface flow*, J. Sci. Comput., 40 (2009), pp. 115–140.
[22] J. Chen, S. Sun, and X. Wang, *A numerical method for a model of two-phase flow in a coupled free flow and porous media system*, J. Comput. Phys., 268 (2014), pp. 1–16.
[23] W. Chen, M. Gunzburger, F. Hua, and X. Wang, *A parallel Robin-Robin domain decomposition method for the Stokes–Darcy system*, SIAM J. Numer. Anal., 49 (2011), pp. 1064–1084.
[24] W. Chen, D. Han, and X. Wang, *Uniquely solvable and energy stable decoupled numerical schemes for the Cahn-Hilliard-Stokes–Darcy system for two-phase flows in karstic geometry*, Numer. Math., 137 (2017), pp. 229–255.
[25] P. Chidyagwai and B. Rivière, *On the solution of the coupled Navier-Stokes and Darcy equations*, Comput. Methods Appl. Mech. Engrg., 198 (2009), pp. 3806–3820.
[26] R. Ştefănescu, A. Sandu, and I. M. Navon, *POD/DEIM reduced-order strategies for efficient four dimensional variational data assimilation*, J. Comput. Phys., 295 (2015), pp. 569–595, https://doi.org/10.1016/j.jcp.2015.04.030.
[27] M. Discacciati, *Domain Decomposition Methods for the Coupling of Surface and Groundwater Flows*, Ph.D. thesis, Ecole Polytechnique Fédérale de Lausanne, Switzerland, 2004.
[28] M. Discacciati and L. Gerardo-Giorda, *Optimized Schwarz methods for the Stokes–Darcy coupling*, IMA J. Numer. Anal., 38 (2018), pp. 1959–1983.
[29] M. Discacciati, P. Gervasio, A. Giacomini, and A. Quarteroni, *The interface control domain decomposition method for Stokes–Darcy coupling*, SIAM J. Numer. Anal., 54 (2016), pp. 1039–1068.
[30] M. Discacciati, E. Miglio, and A. Quarteroni, *Mathematical and numerical models for coupling surface and groundwater flows*, Appl. Numer. Math., 43 (2002), pp. 57–74.
[31] M. Discacciati and R. Oyarzúa, *A conforming mixed finite element method for the Navier-Stokes/Darcy coupled problem*, Numer. Math., 135 (2017), pp. 571–606.
[32] M. Discacciati, A. Quarteroni, and A. Valli, *Robin-Robin domain decomposition methods for the Stokes–Darcy coupling*, SIAM J. Numer. Anal., 45 (2007), pp. 1246–1268.
[33] C. C. Douglas, X. Hu, B. Bai, X.-M. He, M. Wei, and J. Hou, *A data assimilation enabled model for coupling dual porosity flow with free flow*, in Proceedings of the 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science, Wuxi, China, 2018, https://doi.org/10.1109/DCABES.2018.00085.

[34] V. J. Ervin, E. W. Jenkins, and H. Lee, *Approximation of the Stokes–Darcy system by optimization*, J. Sci. Comput., 59 (2014), pp. 775–794.
[35] V. J. Ervin, E. W. Jenkins, and S. Sun, *Coupled generalized nonlinear Stokes flow with flow through a porous medium*, SIAM J. Numer. Anal., 47 (2009), pp. 929–952.
[36] G. Evensen, *The ensemble Kalman filter for combined state and parameter estimation: Monte Carlo techniques for data assimilation in large systems*, IEEE Control Syst. Mag., 29 (2009), pp. 83–104.
[37] W. Feng, X.-M. He, Z. Wang, and X. Zhang, *Non-iterative domain decomposition methods for a non-stationary Stokes–Darcy model with Beavers-Joseph interface condition*, Appl. Math. Comput., 219 (2012), pp. 453–463.
[38] S. W. Funke, M. Nordaas, O. Evju, M. S. Alnæ s, and K. A. Mardal, *Variational data assimilation for transient blood flow simulations: Cerebral aneurysms as an illustrative example*, Int. J. Numer. Methods Biomed. Eng., 35 (2019), e3152–27.
[39] Y. Gao, D. Han, X.-M. He, and U. Rüde, *Unconditionally stable numerical methods for Cahn-Hilliard-Navier-Stokes–Darcy system with different densities and viscosities*, J. Comput. Phys., 454 (2022), 110968.
[40] Y. Gao, X.-M. He, L. Mei, and X. Yang, *Decoupled, linear, and energy stable finite element method for the Cahn-Hilliard-Navier-Stokes–Darcy phase field model*, SIAM J. Sci. Comput., 40 (2018), pp. B110–B137.
[41] B. García-Archilla, J. Novo, and E. S. Titi, *Uniform in time error estimates for a finite element method applied to a downscaling data assimilation algorithm for the Navier-Stokes equations*, SIAM J. Numer. Anal., 58 (2020), pp. 410–429.
[42] M. Gardner, A. Larios, L. G. Rebholz, D. Vargun, and C. Zerfas, *Continuous data assimilation applied to a velocity-vorticity formulation of the 2D Navier-Stokes equations*, Electron. Res. Arch., 29 (2021), pp. 2223–2247, https://doi.org/10.3934/era.2020113.
[43] A. Gronskis, D. Heitz, and E. Mémin, *Inflow and initial conditions for direct numerical simulation based on adjoint data assimilation*, J. Comput. Phys., 242 (2013), pp. 480–497, https://doi.org/10.1016/j.jcp.2013.01.051.
[44] M. Gunzburger, X.-M. He, and B. Li, *On Ritz projection and multi-step backward differentiation schemes in decoupling the Stokes–Darcy model*, SIAM J. Numer. Anal., 56 (2018), pp. 397–427.
[45] D. Han, X.-M. He, Q. Wang, and Y. Wu, *Existence and weak-strong uniqueness of solutions to the Cahn-Hilliard-Navier-Stokes–Darcy system in superposed free flow and porous media*, Nonlinear Anal., 211 (2021), 112411.
[46] D. Han, D. Sun, and X. Wang, *Two-phase flows in karstic geometry*, Math. Methods Appl. Sci., 37 (2014), pp. 3048–3063.
[47] X.-M. He, N. Jiang, and C. Qiu, *An artificial compressibility ensemble algorithm for a stochastic Stokes–Darcy model with random hydraulic conductivity and interface conditions*, Internat. J. Numer. Methods Engrg., 121 (2020), pp. 712–739.
[48] X.-M. He, J. Li, Y. Lin, and J. Ming, *A domain decomposition method for the steady-state Navier–Stokes–Darcy model with Beavers-Joseph interface condition*, SIAM J. Sci. Comput., 37 (2015), pp. S264–S290.
[49] M. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Natl. Bur. Stand., 49 (1952), pp. 409–436, https://doi.org/10.6028/jres.049.044.
[50] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Constraints, Mathematical Modelling: Theory and Applications*, Springer, New York, 2009.
[51] J. Hou, D. Hu, X.-M. He, and C. Qiu, *Modeling and a Robin-type decoupled finite element method for dual-porosity-Navier-Stokes system with application to flows around multistage fractured horizontal wellbore*, Comput. Methods Appl. Mech. Engrg., 388 (2022), 114248.
[52] J. Hou, M. Qiu, X.-M. He, C. Guo, M. Wei, and B. Bai, *A dual-porosity-Stokes model and finite element method for coupling dual-porosity flow and free flow*, SIAM J. Sci. Comput., 38 (2016), pp. B710–B739.
[53] N. Jiang and C. Qiu, *An efficient ensemble algorithm for numerical approximation of stochastic Stokes–Darcy equations*, Comput. Methods Appl. Mech. Engrg., 343 (2019), pp. 249–275.
[54] G. Kanschat and B. Riviére, *A strongly conservative finite element method for the coupling of Stokes and Darcy flow*, J. Comput. Phys, 229 (2010), pp. 5933–5943.
[55] J. C. De los Reyes, *Numerical PDE-Constrained Optimization,* Springer Briefs Optim., Springer, Cham, 2015.
[56] W. Layton, H. Tran, and C. Trenchea, *Analysis of long time stability and errors of two partitioned methods for uncoupling evolutionary groundwater-surface water flows*, SIAM J. Numer. Anal., 51 (2013), pp. 248–272.

[57] W. J. Layton, F. Schieweck, and I. Yotov, *Coupling fluid flow with porous media flow*, SIAM J. Numer. Anal., 40 (2002), pp. 2195–2218.
[58] F.-X. Le Dimet, I. Souopgui, and H. E. Ngodock, *Sensitivity analysis applied to a variational data assimilation of a simulated pollution transport problem*, Internat. J. Numer. Methods Fluids, 83 (2017), pp. 465–482, https://doi.org/10.1002/fld.4274.
[59] J. Li, M. Yamamoto, and J. Zou, *Conditional stability and numerical reconstruction of initial temperature*, Commun. Pure Appl. Anal., 8 (2009), pp. 361–382, https://doi.org/10.3934/cpaa.2009.8.361.
[60] R. Li, J. Li, X.-M. He, and Z. Chen, *A stabilized finite volume element method for a coupled Stokes–Darcy problem*, Appl. Numer. Math., 133 (2018), pp. 2–24.
[61] J. L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer-Verlag, New York, 1971.
[62] Y. Liu, Y. Boubendir, X.-M. He, and Y. He, *New optimized Robin-Robin domain decomposition methods using Krylov solvers for the Stokes–Darcy system*, SIAM J. Sci. Comput., 44 (2022), pp. B1068–B1095.
[63] Y. Liu, Y. He, X. Li, and X.-M. He, *A novel convergence analysis of Robin-Robin domain decomposition method for Stokes–Darcy system with Beavers-Joseph interface condition*, Appl. Math. Lett., 119 (2021), 107181.
[64] M. A. A. Mahbub, X.-M. He, N. J. Nasu, C. Qiu, and H. Zheng, *Coupled and decoupled stabilized mixed finite element methods for non-stationary dual-porosity-Stokes fluid flow model*, Internat. J. Numer. Methods Engrg., 120 (2019), pp. 803–833.
[65] J. Mandel, L. S. Bennethum, J. D. Beezley, J. L. Coen, C. C. C. Douglas, M. Kim, and A. Vodacek, *A wildland fire model with data assimilation*, Math. Comput. Simul., 79 (2008), pp. 584–606.
[66] G. I. Marchuk and V. P. Shutyaev, *Solvability and numerical algorithms for a class of variational data assimilation problems. A tribute to J. L. Lions*, ESAIM Control Optim. Calc. Var., 8 (2002), pp. 873–883.
[67] G. I. Marchuk and V. P. Shutyaev, *Iterative algorithms for data assimilation problems*, in Frontiers in Mathematical Analysis and Numerical Methods, World Scientific, River Edge, NJ, 2004, pp. 197–206.
[68] P. A. Markowich, E. S. Titi, and S. Trabelsi, *Continuous data assimilation for the three-dimensional Brinkman-Forchheimer-extended Darcy model*, Nonlinearity, 29 (2016), pp. 1292–1328, https://doi.org/10.1088/0951-7715/29/4/1292.
[69] M. Meldi and A. Poux, *A reduced order model based on Kalman filtering for sequential data assimilation of turbulent flows*, J. Comput. Phys., 347 (2017), pp. 207–234, https://doi.org/10.1016/j.jcp.2017.06.042.
[70] V. Mons, J.-C. Chassaing, T. Gomez, and P. Sagaut, *Reconstruction of unsteady viscous flows using data assimilation schemes*, J. Comput. Phys., 316 (2016), pp. 255–280, https://doi.org/10.1016/j.jcp.2016.04.022.
[71] M. Mu and J. Xu, *A two-grid method of a mixed Stokes–Darcy model for coupling fluid flow with porous media flow*, SIAM J. Numer. Anal., 45 (2007), pp. 1801–1813.
[72] E. Olson and E. S. Titi, *Determining modes for continuous data assimilation in 2D turbulence*, J. Statist. Phys., 113 (2003), pp. 799–840, https://doi.org/10.1023/A:1027312703252. Progress in statistical hydrodynamics (Santa Fe, NM, 2002).
[73] C. Qiu, *Decoupling Methods for the Time-Dependent Navier–Stokes–Darcy Interface Model*, Ph.D. dissertation, Missouri University of Science and Technology, 2019.
[74] C. Qiu, X.-M. He, J. Li, and Y. Lin, *A domain decomposition method for the time-dependent Navier–Stokes–Darcy model with Beavers-Joseph interface condition and defective boundary condition*, J. Comput. Phys., 411 (2020), 109400.
[75] L. G. Rebholz and C. Zerfas, *Simple and efficient continuous data assimilation of evolution equations via algebraic nudging*, Numer. Methods Partial Differential Equations, 37 (2021), pp. 2588–2612, https://doi.org/10.1002/num.22751.
[76] S. Reich, *Data assimilation: The Schrödinger perspective*, Acta Numer., 28 (2019), pp. 635–711, https://doi.org/10.1017/s0962492919000011.
[77] F. A. Rihan, C. G. Collier, and I. Roulstone, *Four-dimensional variational data assimilation for Doppler radar wind data*, J. Comput. Appl. Math., 176 (2005), pp. 15–34.
[78] B. Rivière and I. Yotov, *Locally conservative coupling of Stokes and Darcy flows*, SIAM J. Numer. Anal., 42 (2005), pp. 1959–1977.
[79] D. Rozier, E. Birol, E. Cosme, P. Brasseur, J. M. Brankart, and J. Verron, *A reduced-order Kalman filter for data assimilation in physical oceanography*, SIAM Rev., 49 (2007), pp. 449–465.

[80] L. Shan and H. Zheng, *Partitioned time stepping method for fully evolutionary Stokes–Darcy flow with Beavers-Joseph interface conditions*, SIAM J. Numer. Anal., 51 (2013), pp. 813–839.

[81] M. Tang, Y. Liu, and L. J. Durlofsky, *A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems*, J. Comput. Phys., 413 (2020), 28, https://doi.org/10.1016/j.jcp.2020.109456.

[82] G. Triantafyllou, I. Hoteit, G. Korres, and G. Petihakis, *Ecosystem modeling and data assimilation of physical-biogeochemical processes in shelf and regional areas of the Mediterranean Sea*, Appl. Numer. Anal. Comput. Math., 2 (2005), pp. 262–280.

[83] D. Vassilev and I. Yotov, *Coupling Stokes–Darcy flow with transport*, SIAM J. Sci. Comput., 31 (2009), pp. 3661–3684.

[84] F. Veersé, D. Auroux and M. Fisher, *Limited-memory BFGS diagonal preconditioners for a data assimilation problem in meteorology*, Optim. Eng., 1 (2000), pp. 323–339.

[85] H. X. Vo and L. J. Durlofsky, *Data assimilation and uncertainty assessment for complex geological models using a new PCA-based parameterization*, Comput. Geosci., 19 (2015), pp. 747–767.

[86] R. G. Vuchkov, C. G. Petra, and N. Petra, *On the derivation of quasi-Newton formulas for optimization in function spaces*, Numer. Funct. Anal. Optim., 41 (2020), pp. 1564–1587, https://doi.org/10.1080/01630563.2020.1785496.

[87] M. Yamamoto and J. Zou, *Simultaneous reconstruction of the initial temperature and heat radiative coefficient*, Inverse Problems, 17 (2001), pp. 1181–1202.

[88] Z. Yang, X. Li, X.-M. He, and J. Ming, *A stochastic collocation method based on sparse grids for a stochastic Stokes–Darcy model*, Discrete Contin. Dyn. Syst. Ser. S, 15 (2022), pp. 893–912.

[89] Z. Yang, J. Ming, C. Qiu, X.-M. He, and M. Li, *A multigrid multilevel Monte Carlo method for Stokes–Darcy model with random hydraulic conductivity and Beavers-Joseph condition*, J. Sci. Comput., 90 (2022), 68.

[90] C. Zerfas, L. G. Rebholz, M. Schneier, and T. Iliescu, *Continuous data assimilation reduced order models of fluid flow*, Comput. Methods Appl. Mech. Engrg., 357 (2019), 112596, https://doi.org/10.1016/j.cma.2019.112596.

[91] B. Zhao, M. Zhang, and C. Liang, *Global well-posedness for Navier-Stokes–Darcy equations with the free interface*, Int. J. Numer. Anal. Model., 18 (2021), pp. 569–619.

[92] X. Zou, I. M. Navon, and F. X. Ledimet, *An optimal nudging data assimilation scheme using parameter estimation*, Quart. J. R. Meteorol. Soc., 118 (1992), pp. 1163–1186, https://doi.org/10.1002/qj.49711850808.