

UNCONDITIONALLY ENERGY STABLE AND FIRST-ORDER ACCURATE NUMERICAL SCHEMES FOR THE HEAT EQUATION WITH UNCERTAIN TEMPERATURE-DEPENDENT CONDUCTIVITY

JOSEPH ANTHONY FIORDILINO* AND MATTHEW WINGER

Abstract. In this paper, we present first-order accurate numerical methods for solution of the heat equation with uncertain temperature-dependent thermal conductivity. Each algorithm yields a shared coefficient matrix for the ensemble set improving computational efficiency. Both mixed and Robin-type boundary conditions are treated. In contrast with alternative, related methodologies, stability and convergence are unconditional. In particular, we prove unconditional, energy stability and optimal-order error estimates. A battery of numerical tests are presented to illustrate both the theory and application of these algorithms.

Key words. Time-stepping, finite element method, heat equation, temperature-dependent thermal conductivity, uncertainty quantification.

1. Introduction

Demand for superior predictions of scientific and engineering problems is ever increasing. Improvement of available computational resources and both development and application of numerical methodologies work synergistically to meet the aforementioned demand. In particular, numerical schemes are devised to improve model accuracy (e.g., via inclusion of additional physics), replicate additional properties of the continuous problem (e.g., long-time stability), incorporate uncertainty quantification via statistical techniques, etc. The focus of this manuscript is on improving the efficiency of ensemble simulations, which facilitate uncertainty quantification, applied to heat conduction dynamics with increased model physics.

The crisis of predictability in numerical weather prediction, led to the discovery of chaos and the use of ensemble simulations to produce predictive results with uncertainty quantified. Some key figures include, Charney [5], Philips [41], Thompson [45], Lorenz [34, 35, 36]; see, e.g., [27, 32] and references therein for a historical perspective. Ensemble calculations typically involve J solves of a set of equations with slightly perturbed initial data. Calculations are performed as either J sequential, fine mesh runs or J parallel, coarse mesh runs of a given code. The ensemble average tends to perform better as a prediction than any of the individual realizations; see, e.g., Chapter 6 Section 5 of [27] or [2, 13, 28]. Evidently, increased computational resources are needed over a single realization run. Moreover, since both increased ensemble size J and mesh density h yield superior results, there is an inherent competition for available computational resources.

Received by the editors on September 30, 2021 and, accepted on September 18, 2023.
2000 *Mathematics Subject Classification.* 65M12, 65M15, 65M60, and 80A20.

*Corresponding author.

The last six years have seen increased focus in improving efficiency of ensemble calculations [21, 22, 23, 24, 40, 43, 14, 15, 16, 17, 10, 8, 12, 9, 33, 6, 26, 38, 39, 42] and references therein. The driver for much of this work is owed to a breakthrough work by Jiang and Layton [21], as applied to non-isothermal fluid flow. Therein, they recognized that a consistent modification of the convective term, utilizing the ensemble mean and fluctuation of the viscosity together with lagging of the fluctuation term, would yield a shared coefficient matrix independent of the ensemble member j . The result was a reduction in both storage requirements and solution turnover time.

Recent years have seen increased focus towards problems with uncertain parameters and considerations of alternative physics. Of particular interest here, first- and second-order ensemble algorithms for iso-thermal fluid flow with constant viscosity were developed in [14, 15]. Further, first-order methods were presented for the heat equation with constant thermal conductivity under mixed boundary conditions in [42] and both space and time dependent thermal conductivity under Dirichlet boundary conditions in [38]. Moreover, first and second-order methods were developed for spatially dependent thermal conductivities in [9]. Notably, stochastics were incorporated in [38, 39] via the Monte Carlo method and in [33] for the convection-diffusion equation via stochastic collocation.

In each of the above works, both stability and convergence were conditionally dependent on the ratio between the fluctuating and mean values of the relevant parameter. In contrast, the ensemble methods presented herein are unconditionally, nonlinearly, energy stable and first-order accurate, with $\Delta t = \mathcal{O}(h)$. Moreover, we consider the heat equation with uncertain temperature-dependent thermal conductivity due to uncertain initial conditions. Physically, this is more realistic as most materials' thermal conductivity exhibit non-trivial temperature-dependence. Mathematically, the resulting equation becomes nonlinear, in the diffusive term, presenting new challenges over the analogous linear problem.

Let $\Omega \subset \mathbb{R}^d$ be an open, bounded, Lipschitz domain. Given initial temperature $T^0(x) = T(x, 0)$, thermal conductivity κ and heat source f , find $T(x, t) : \Omega \times (0, t^*] \rightarrow R$ satisfying

$$(1) \quad \frac{\partial T}{\partial t} - \nabla \cdot (\kappa \nabla T) = f \text{ in } \Omega.$$

We consider two boundary configurations: mixed and Robin. Throughout, κ is the thermal conductivity of the solid medium which depends on the temperature profile; that is, $\kappa \equiv \kappa(T)$. For the mixed boundary condition, the boundary $\partial\Omega$ is partitioned such that $\partial\Omega = \overline{\Gamma_D} \cup \overline{\Gamma_N}$ with $\Gamma_D \cap \Gamma_N = \emptyset$ (Γ_D for Dirichlet condition and Γ_N for Neumann condition). Let n denote the outward normal, then

$$(2) \quad T = 0 \text{ on } \Gamma_D, \quad \nabla T \cdot n = 0 \text{ on } \Gamma_N.$$

Moreover, the Robin condition is prescribed via

$$(3) \quad \alpha T + \kappa \nabla T \cdot n = \beta \text{ on } \partial\Omega,$$

where $\alpha \in [0, 1]$ is the emissivity, and β a prescribed function on the boundary.

The paper is organized as follows. In Section 2, we introduce mathematical preliminaries required in the analysis, including semi-discrete numerical schemes and

finite element preliminaries. The fully discrete schemes are introduced in Section 3. Sections 4 and 5 are devoted to the stability and error analysis of the fully discrete algorithms. Results from a battery of numerical tests are provided in Section 6. These serve to illustrate the validity of the proven theory and value of the algorithms. Finally, conclusions are drawn in Section 7.

2. Mathematical Preliminaries

Herein, we introduce notation and preliminaries that are necessary for presentation and analysis. $H^s(\Omega)$ denotes the Hilbert space of $L^2(\Omega)$ functions with distributional derivatives of order $s \geq 0$ in $L^2(\Omega)$. The corresponding norms and seminorms are $\|\cdot\|_s$ and $|\cdot|_s$. In the special case $s = 0$, $H^0(\Omega) = L^2(\Omega)$ and the associated inner product and induced norm are (\cdot, \cdot) and $\|\cdot\|$. Moreover, $(\cdot, \cdot)_{\partial\Omega}$ and $\|\cdot\|_{\partial\Omega}$ denote the $L^2(\partial\Omega)$ inner product and induced norm on the boundary.

Define the Hilbert spaces,

$$X := H^1(\Omega), Y := \{S \in H^1(\Omega) : S = 0 \text{ on } \Gamma_D\},$$

with dual norm $\|\cdot\|_{-1}$ understood to correspond to either X or Y . The following Poincaré-like inequalities are critical in the analysis.

Lemma 1. *Let γ be a linear form on $H^1(\Omega)$ whose restriction to constant functions is nonzero. Then, $\exists C_P > 0$ such that $\forall S \in X$*

$$(4) \quad C_P \|S\|_1 \leq \|\nabla S\| + |\gamma(S)|.$$

Moreover, if $S \in Y$ then $\exists C_{PF} > 0$ satisfying

$$(5) \quad \|S\| \leq C_{PF} \|\nabla S\|.$$

Proof. See Lemma B.63 pp. 490, for the former, and Lemma B.66 pp. 491, for the latter, in [3]. □

The Poincaré-Friedrichs inequality, inequality (5), guarantees that $|\cdot|_1$ is an equivalent norm to $\|\cdot\|_1$ in Y . Recall, Young’s inequality is given by

$$(6) \quad ab \leq \frac{\epsilon}{q} a^q + \frac{\epsilon^{-r/q}}{r} b^r, \quad 1 < q, r < \infty, \frac{1}{q} + \frac{1}{r}, a, b \geq 0.$$

The special case $q = r = 2$ will be used throughout.

Let $\{T(x, t; \omega_j)\}_{j=1}^J$ denote the ensemble set of solution variables to equation (1), with corresponding boundary conditions; ω_j parametrizes each ensemble member $j \in [1, J]$. Then, the weak formulation of system (1) and (2) is: Find $T : [0, t^*] \rightarrow Y$ for a.e. $t \in (0, t^*]$ satisfying for $j = 1, 2, \dots, J$:

$$(7) \quad \left(\frac{\partial T}{\partial t}, S\right) + (\kappa \nabla T, \nabla S) = (f, S) \quad \forall S \in Y.$$

Furthermore, the weak formulation of system (1) and (3) is : Find $T : [0, t^*] \rightarrow X$ for a.e. $t \in (0, t^*]$ satisfying for $j = 1, 2, \dots, J$:

$$(8) \quad \left(\frac{\partial T}{\partial t}, S\right) + (\kappa \nabla T, \nabla S) + (\alpha T, S)_{\partial\Omega} = (f, S) + (\beta, S)_{\partial\Omega} \quad \forall S \in X.$$

Throughout, the thermal conductivity is assumed bounded and continuously differentiable such that:

$$(9) \quad |\kappa(T) - \kappa(S)| \leq C_\kappa |T - S| \quad \forall S, T,$$

$$(10) \quad 0 < \kappa_{min} \leq \kappa(S) \leq \kappa_{max} < \infty \quad \forall S.$$

Remark: Practically, κ_{max} can be estimated a priori knowing that equation (1) is elliptic satisfying a maximum principle [4].

A discrete Gronwall inequality will be critical in the subsequent stability and error analysis. Let N be a positive integer and set both $\Delta t = \frac{t^*}{N}$ and $t^n = n\Delta t$ for $0 \leq n \leq N$. Then, $[0, t^*] = \bigcup_{n=0}^{N-1} [t^n, t^{n+1}]$ is a partition of the time interval.

Lemma 2. (*Discrete Gronwall Lemma*). *Let $\Delta t, H, a_n, b_n, c_n,$ and d_n be finite nonnegative numbers for $n \geq 0$ such that for $N \geq 1$*

$$a_N + \Delta t \sum_0^N b_n \leq \Delta t \sum_0^{N-1} d_n a_n + \Delta t \sum_0^N c_n + H,$$

then for all $\Delta t > 0$ and $N \geq 1$

$$a_N + \Delta t \sum_0^N b_n \leq \exp(\Delta t \sum_0^{N-1} d_n) (\Delta t \sum_0^N c_n + H).$$

Proof. See Lemma 5.1 on pp. 369 of [20]. □

Lastly, the following norms are utilized in the error analysis: $\forall -1 \leq k < \infty,$

$$\|v\|_{\infty,k} := \max_{0 \leq n \leq N} \|v^n\|_k, \quad \|v\|_{p,k} := \left(\Delta t \sum_{n=0}^N \|v^n\|_k^p \right)^{1/p}.$$

We are now in a place to discuss the key idea of the numerical methods. Let $\kappa^n \equiv \kappa(T^n)$ and $\kappa'^n \equiv \kappa_{max} - \kappa^n$. Suppress the spatial discretization, apply an implicit-explicit time-discretization to the system (1) with (2):

Algorithm 1 (a):

$$(11) \quad \frac{T^{n+1} - T^n}{\Delta t} - \kappa_{max} \Delta T^{n+1} + \nabla \cdot (\kappa'^n \nabla T^n) = f^{n+1},$$

Remark: For the Robin boundary condition (3), the form of the above scheme is modified such that $(\alpha T^{n+1}, S)_{\partial\Omega}$ and $(\beta, S)_{\partial\Omega}$ appear on the left- and right-hand sides, respectively.

Applying a standard FEM discretization in space for the above system, we arrive at the following block linear system for each ensemble member j :

$$(12) \quad \left(\frac{1}{\Delta t} M + \kappa_{max} D \right) T^{n+1} = \left(f^{n+1} + \frac{1}{\Delta t} M + N_\kappa(T^n) \right) T^n,$$

where M is the mass matrix, D is the diffusion matrix, and $N_\kappa(T^n)$ is the matrix associated with conductivity fluctuations. The above linear system is equivalent to

the following: Let A be the resulting coefficient matrix (independent of j). Then, the following set of J linear systems must be solved at each timestep:

$$(13) \quad [A] [x_1|x_2|\dots|x_J] = [b_1|b_2|\dots|b_J].$$

The matrix A is symmetric positive definite (SPD) since both $\frac{1}{\Delta t}M$ and $\kappa_{max}D$ are SPD. The system (13) can be solved with efficient block solvers [7, 18]. Further, since only one coefficient matrix is required for computation per timestep, the storage requirement is thereby reduced.

2.1. Finite Element Preliminaries. Let $\{\mathcal{T}_h\}_{0 < h < 1}$ be a family of quasi-uniform meshes with maximum element length $h = \max_{K \in \mathcal{T}_h} h_K$. We define the geometric interpolation of Ω as $\Omega_h = \bigcup_{K \in \mathcal{T}_h} K$. Throughout, Ω is assumed to be a convex polytope so that $\Omega = \Omega_h$. Let $X_h \subset X$ and $Y_h \subset Y$ be conforming finite element spaces defined as

$$\begin{aligned} X_h &:= \{S_h \in C^0(\bar{\Omega}_h) : \forall K \in \mathcal{T}_h, S_h|_K \in \mathbb{P}_l(K)\} \cap X, \\ Y_h &:= \{S_h \in C^0(\bar{\Omega}_h) : \forall K \in \mathcal{T}_h, S_h|_K \in \mathbb{P}_l(K)\} \cap Y. \end{aligned}$$

The spaces above satisfy the following approximation properties: $\forall 1 \leq l \leq k$,

$$(14) \quad \inf_{S_h \in Z} \left\{ \|T - S_h\| + h \|\nabla(T - S_h)\| \right\} \leq Ch^{k+1} |T|_{k+1} \quad T \in Z \cap H^{k+1}(\Omega),$$

and $Z = X$ or Y . Throughout, C denotes a generic positive constant independent of h or Δt .

3. Numerical Scheme

Let T_h^n be the fully discrete approximate solution at time level t^n , $\kappa_h^n = \kappa(T_h^n)$, and $\kappa_h'^n = \kappa_{max} - \kappa_h^n$. Then, the fully discrete schemes are:

Algorithm 1:

(a) Given $T_h^n \in Y_h$, find $T_h^{n+1} \in Y_h$ satisfying

$$(15) \quad \left(\frac{T_h^{n+1} - T_h^n}{\Delta t}, S_h \right) + (\kappa_{max} \nabla T_h^{n+1}, \nabla S_h) - (\kappa_h'^n \nabla T_h^n, \nabla S_h) = (f^{n+1}, S_h) \quad \forall S_h \in Y_h.$$

(b) Given $T_h^n \in X_h$, find $T_h^{n+1} \in X_h$ satisfying the fully discrete scheme as follows:

$$(16) \quad \begin{aligned} &\left(\frac{T_h^{n+1} - T_h^n}{\Delta t}, S_h \right) + (\kappa_{max} \nabla T_h^{n+1}, \nabla S_h) - (\kappa_h'^n \nabla T_h^n, \nabla S_h) + (\alpha T_h^{n+1}, S_h)_{\partial\Omega} \\ &= (f^{n+1}, S_h) + (\beta, S_h)_{\partial\Omega} \quad \forall S_h \in X_h. \end{aligned}$$

Remark: If the thermal conductivity is provided with explicit dependence on space and time, e.g., $\kappa \equiv \kappa(x, t)$, then

$$(\kappa_{max} \nabla T_h^{n+1}, \nabla S_h) - (\kappa_h'^n \nabla T_h^n, \nabla S_h) \leftarrow (\kappa_{max}^n \nabla T_h^{n+1}, \nabla S_h) - (\kappa_h'^n \nabla T_h^n, \nabla S_h)$$

with $\kappa_{max}^n = \max_{1 \leq j \leq J} \sup_{x \in \Omega} \kappa(x, t^n)$ and $\kappa_h'^n = \kappa_{max}^n - \kappa(x, t^n)$, in the above, yield unconditionally stable and first-order accurate methods. The analysis is novel but analogous to that presented below. Advantageously, κ_{max} needs not be

estimated a priori. Moreover, the consistency error is tighter with more relaxed requirements on solution regularity.

Finally, the following regularity assumptions of the temperature T are needed for the subsequent analyses of the schemes (15) and (16) respectively:

$$(17) \quad T \in L^\infty(0, t^*; Y \cap H^{k+1}(\Omega)), \nabla T \in L^\infty(0, t^*; L^\infty(\Omega)), \\ \frac{\partial T}{\partial t} \in L^2(0, t^*; X), \frac{\partial^2 T}{\partial t^2} \in L^2(0, t^*; L^2(\Omega)),$$

$$(18) \quad T \in L^\infty(0, t^*; X \cap H^{k+1}(\Omega)), \nabla T \in L^\infty(0, t^*; L^\infty(\Omega)), \\ \frac{\partial T}{\partial t} \in L^2(0, t^*; X), \frac{\partial^2 T}{\partial t^2} \in L^2(0, t^*; H^{-1}(\Omega)).$$

4. Stability Analysis

In Theorem 3, stability of the temperature approximation is proven for Algorithm 1, both (15) and (16). Later, first-order convergence is proven in Theorem 5.

Theorem 3. *Consider Algorithm 1(a) and suppose $f \in L^2(0, t^*; H^{-1}(\Omega))$ and $\beta \in H^{-1}(\partial\Omega)$, then*

$$(19) \quad \|T_h^N\|^2 + \|\sqrt{\kappa_{max}} \nabla T_h^N\|^2 + \sum_{n=0}^{N-1} \left(\|T_h^{n+1} - T_h^n\|^2 + \Delta t \|\sqrt{\kappa_h^{t_n}} \nabla (T_h^{n+1} - T_h^n)\|^2 \right) \\ + \frac{\Delta t}{2} \sum_{n=0}^{N-1} \|\sqrt{\kappa_h^n} \nabla T_h^{n+1}\|^2 \leq \|T_h^0\|^2 + \|\sqrt{\kappa_{max}} \nabla T_h^0\|^2 + \frac{2\Delta t}{\kappa_{min}} \sum_{n=0}^{N-1} \|f^{n+1}\|_{-1}^2.$$

Moreover, for Algorithm 1(b), we have

$$(20) \quad \|T_h^N\|^2 + \|\sqrt{\kappa_{max}} \nabla T_h^N\|^2 + \sum_{n=0}^{N-1} \left(\|T_h^{n+1} - T_h^n\|^2 + \Delta t \|\sqrt{\kappa_h^{t_n}} \nabla (T_h^{n+1} - T_h^n)\|^2 \right) \\ + \frac{C_P^2 \Delta t}{8} \sum_{n=0}^{N-1} \|\sqrt{\kappa_h^n} \nabla T_h^{n+1}\|_1^2 \\ \leq \|T_h^0\|^2 + \|\sqrt{\kappa_{max}} \nabla T_h^0\|^2 + \frac{4\Delta t}{C_P^2 \kappa_{min}} \sum_{n=0}^{N-1} \left(\|f^{n+1}\|_{-1}^2 + 2\|\beta\|_{-1, \partial\Omega}^2 \right).$$

Proof. Setting $S_h = 2\Delta t T_h^{n+1}$ in (15) and using the polarization identity on the first term, we have

$$(21) \quad \|T_h^{n+1}\|^2 - \|T_h^n\|^2 + \|T_h^{n+1} - T_h^n\|^2 + 2\Delta t \|\sqrt{\kappa_{max}} \nabla T_h^{n+1}\|^2 \\ - 2\Delta t (\kappa_h^{t_n} \nabla T_h^n, \nabla T_h^{n+1}) = (f^{n+1}, 2\Delta t T_h^{n+1}).$$

Now,

$$(22) \quad \begin{aligned} & 2\Delta t \|\sqrt{\kappa_{max}} \nabla T_h^{n+1}\|^2 - 2\Delta t (\kappa_h'^n \nabla T_h^n, \nabla T_h^{n+1}) \\ & = 2\Delta t (\kappa_{max} \nabla T_h^{n+1}, \nabla (T_h^{n+1} - T_h^n)) + 2\Delta t (\kappa_h^n \nabla T_h^n, \nabla T_h^{n+1}). \end{aligned}$$

Using the polarization identity twice and rearranging terms in equation (21) yields,

$$(23) \quad \begin{aligned} & \|T_h^{n+1}\|^2 - \|T_h^n\|^2 + \|T_h^{n+1} - T_h^n\|^2 + \|\sqrt{\kappa_{max}} \nabla T_h^{n+1}\|^2 - \|\sqrt{\kappa_{max}} \nabla T_h^n\|^2 \\ & + \Delta t \|\sqrt{\kappa_h'^n} \nabla (T_h^{n+1} - T_h^n)\|^2 + \Delta t \|\sqrt{\kappa_h^n} \nabla T_h^{n+1}\|^2 + \Delta t \|\sqrt{\kappa_h^n} \nabla T_h^n\|^2 \\ & = 2\Delta t (f^{n+1}, T_h^{n+1}). \end{aligned}$$

Application of Cauchy-Schwarz and Young's inequalities on the forcing term leads to

$$(24) \quad 2\Delta t (f^{n+1}, T_h^{n+1}) \leq \frac{\Delta t}{\epsilon_1 \kappa_{min}} \|f^{n+1}\|_{-1}^2 + \epsilon_1 \Delta t \|\sqrt{\kappa_h^n} \nabla T_h^{n+1}\|^2.$$

Drop $\Delta t \|\sqrt{\kappa_h^n} \nabla T_h^n\|^2$, use estimate (24) with $\epsilon_1 = 1/2$, and rearrange terms. Then,

$$(25) \quad \begin{aligned} & \|T_h^{n+1}\|^2 - \|T_h^n\|^2 + \|T_h^{n+1} - T_h^n\|^2 + \|\sqrt{\kappa_{max}} \nabla T_h^{n+1}\|^2 - \|\sqrt{\kappa_{max}} \nabla T_h^n\|^2 \\ & + \Delta t \|\sqrt{\kappa_h'^n} \nabla (T_h^{n+1} - T_h^n)\|^2 + \frac{\Delta t}{2} \|\sqrt{\kappa_h^n} \nabla T_h^{n+1}\|^2 \leq \frac{2\Delta t}{\kappa_{min}} \|f^{n+1}\|_{-1}^2. \end{aligned}$$

Summing from $n = 0$ to $n = N - 1$, we arrive at

$$(26) \quad \begin{aligned} & \|T_h^N\|^2 + \|\sqrt{\kappa_{max}} \nabla T_h^N\|^2 + \sum_{n=0}^{N-1} \left(\|T_h^{n+1} - T_h^n\|^2 + \Delta t \|\sqrt{\kappa_h'^n} \nabla (T_h^{n+1} - T_h^n)\|^2 \right) \\ & + \frac{\Delta t}{2} \sum_{n=0}^{N-1} \|\sqrt{\kappa_h^n} \nabla T_h^{n+1}\|^2 \leq \|T_h^0\|^2 + \|\sqrt{\kappa_{max}} \nabla T_h^0\|^2 + \frac{2\Delta t}{\kappa_{min}} \sum_{n=0}^{N-1} \|f^{n+1}\|_{-1}^2. \end{aligned}$$

Similarly, setting $S_h = 2\Delta t T_h^{n+1}$ in equation (16), we have

$$(27) \quad \begin{aligned} & \|T_h^{n+1}\|^2 - \|T_h^n\|^2 + \|T_h^{n+1} - T_h^n\|^2 + 2\Delta t \|\sqrt{\kappa_{max}} \nabla T_h^{n+1}\|^2 \\ & + 2\Delta t \|\sqrt{\alpha} T_h^{n+1}\|_{\partial\Omega}^2 - 2\Delta t (\kappa_h'^n \nabla T_h^n, \nabla T_h^{n+1}) \\ & = (f^{n+1}, 2\Delta t T_h^{n+1}) + 2\Delta t (\beta, T_h^{n+1})_{\partial\Omega}. \end{aligned}$$

Following the analysis above and rearranging leads to

$$(28) \quad \begin{aligned} & \|T_h^{n+1}\|^2 - \|T_h^n\|^2 + \|T_h^{n+1} - T_h^n\|^2 + \|\sqrt{\kappa_{max}} \nabla T_h^{n+1}\|^2 - \|\sqrt{\kappa_{max}} \nabla T_h^n\|^2 \\ & + \Delta t \|\sqrt{\kappa_h'^n} \nabla (T_h^{n+1} - T_h^n)\|^2 + \Delta t \|\sqrt{\kappa_h^n} \nabla T_h^{n+1}\|^2 + \Delta t \|\sqrt{\kappa_h^n} \nabla T_h^n\|^2 \\ & + 2\Delta t \|\sqrt{\alpha} T_h^{n+1}\|_{\partial\Omega}^2 = 2\Delta t (f^{n+1}, T_h^{n+1}) + 2\Delta t (\beta, T_h^{n+1})_{\partial\Omega}. \end{aligned}$$

Apply Cauchy-Schwarz and Young's inequalities on the two terms on the right hand side of (28). Then,

$$(29) \quad 2\Delta t (f^{n+1}, T_h^{n+1}) \leq \frac{\Delta t}{\epsilon_2 \kappa_{min}} \|f^{n+1}\|_{-1}^2 + \epsilon_2 \Delta t \|\sqrt{\kappa_h^n} \nabla T_h^{n+1}\|_1^2,$$

$$(30) \quad 2\Delta t(\beta, T_h^{n+1})_{\partial\Omega} \leq \frac{\Delta t}{\epsilon_3 \kappa_{min}} \|\beta\|_{-1, \partial\Omega}^2 + \epsilon_3 \Delta t \|\sqrt{\kappa_h^n} T_h^{n+1}\|_1^2.$$

From Lemma 1, we have $\|\sqrt{\kappa_h^n} \nabla T_h^{n+1}\|^2 + \|\sqrt{\alpha} T_h^{n+1}\|^2 \geq \frac{C_P^2}{2} \|\sqrt{\kappa_h^n} T_h^{n+1}\|_1^2$ and thus

$$(31) \quad \begin{aligned} \frac{C_P^2 \Delta t}{2} \|\sqrt{\kappa_h^n} T_h^{n+1}\|_1^2 &\leq \Delta t \|\sqrt{\kappa_h^n} \nabla T_h^{n+1}\|^2 + \Delta t \|\sqrt{\alpha} T_h^{n+1}\|^2 \\ &\leq \Delta t \|\sqrt{\kappa_h^n} \nabla T_h^{n+1}\|^2 + \left(\Delta t \|\sqrt{\kappa_h^n} \nabla T_h^n\|^2 + 2\Delta t \|\sqrt{\alpha} T_h^{n+1}\|^2 \right). \end{aligned}$$

Combine estimates (29)-(30) with $\epsilon_2 = 2\epsilon_3 = \frac{C_P^2}{4}$, the above estimate in equation (28), and rearrange. Then,

$$(32) \quad \begin{aligned} &\|T_h^{n+1}\|^2 - \|T_h^n\|^2 + \|T_h^{n+1} - T_h^n\|^2 + \|\sqrt{\kappa_{max}} \nabla T_h^{n+1}\|^2 - \|\sqrt{\kappa_{max}} \nabla T_h^n\|^2 \\ &\quad + \Delta t \|\sqrt{\kappa_h^n} \nabla (T_h^{n+1} - T_h^n)\|^2 + \frac{C_P^2 \Delta t}{8} \|\sqrt{\kappa_h^n} T_h^{n+1}\|_1^2 \\ &\leq \frac{4\Delta t}{C_P^2 \kappa_{min}} \|f^{n+1}\|_{-1}^2 + \frac{8\Delta t}{C_P^2 \kappa_{min}} \|\beta\|_{-1, \partial\Omega}^2. \end{aligned}$$

Summing from $n = 0$ to $n = N - 1$ leads to

$$(33) \quad \begin{aligned} &\|T_h^N\|^2 + \|\sqrt{\kappa_{max}} \nabla T_h^N\|^2 + \sum_{n=0}^{N-1} \left(\|T_h^{n+1} - T_h^n\|^2 + \Delta t \|\sqrt{\kappa_h^n} \nabla (T_h^{n+1} - T_h^n)\|^2 \right) \\ &\quad + \frac{C_P^2 \Delta t}{8} \sum_{n=0}^{N-1} \|\sqrt{\kappa_h^n} \nabla T_h^{n+1}\|_1^2 \\ &\leq \|T_h^0\|^2 + \|\sqrt{\kappa_{max}} \nabla T_h^0\|^2 + \frac{4\Delta t}{C_P^2 \kappa_{min}} \sum_{n=0}^{N-1} \left(\|f^{n+1}\|_{-1}^2 + 2\|\beta\|_{-1, \partial\Omega}^2 \right). \end{aligned}$$

□

The stability result above shows that we have control over the temperature approximation in both $L^\infty(0, t^*; L^2(\Omega))$ and $L^2(0, t^*; H^1(\Omega))$, unconditionally. Moreover, we see that the numerical dissipation is enhanced, compared to standard Backward Euler, with the additional term $\Delta t \sum_{n=0}^{N-1} \|\sqrt{\kappa_h^n} \nabla (T_h^{n+1} - T_h^n)\|^2$.

5. Convergence Analysis

In this section, we first analyze the consistency errors for each numerical scheme. Convergence is then proven at the anticipated, optimal rates. Recall, the true solution satisfies

$$(34) \quad \begin{aligned} &\left(\frac{T^{n+1} - T^n}{\Delta t}, S \right) + (\kappa^{n+1} \nabla T^{n+1}, \nabla S) \\ &= (f^{n+1}, S) + \left(\frac{T^{n+1} - T^n}{\Delta t} - \frac{\partial T^{n+1}}{\partial t}, S \right) \quad \forall S \in Y, \end{aligned}$$

under mixed boundary conditions. For the Robin boundary condition, the true solution satisfies

$$(35) \quad \begin{aligned} & \left(\frac{T^{n+1} - T^n}{\Delta t}, S \right) + (\kappa^{n+1} \nabla T^{n+1}, \nabla S) + (\alpha T^{n+1}, S)_{\partial\Omega} \\ & = (f^{n+1}, S) + \left(\frac{T^{n+1} - T^n}{\Delta t} - \frac{\partial T^{n+1}}{\partial t}, S \right) + (\beta, S)_{\partial\Omega} \quad \forall S \in X. \end{aligned}$$

Denote $e^n = (T^n - I_h T^n) - (T_h^n - I_h T^n) = \phi^n - \psi_h^n$, where $I_h T^n$ is an arbitrary interpolate of T^n and e^n is the error at the time $t = t_n$; e.g., the Lagrange interpolate [3] is common and applicable. Letting $S = S_h \in X_h$ or Y_h and subtracting (34) and (35) from (15) and (16), respectively, yields the error equations:

$$(36) \quad \left(\frac{e^{n+1} - e^n}{\Delta t}, S_h \right) + (\kappa_{max} \nabla e^{n+1}, \nabla S_h) - (\kappa_h^m \nabla e^n, \nabla S_h) = \xi_1(T^{n+1}, S_h) \quad \forall S_h \in Y_h,$$

and

$$(37) \quad \begin{aligned} & \left(\frac{e^{n+1} - e^n}{\Delta t}, S_h \right) + (\kappa_{max} \nabla e^{n+1}, \nabla S_h) - (\kappa_h^m \nabla e^n, \nabla S_h) + (\alpha e^{n+1}, S_h) \\ & = \xi_1(T^{n+1}, S_h) \quad \forall S_h \in X_h, \end{aligned}$$

where $\xi_1(T^{n+1}, S_h)$ is defined as

$$(38) \quad \begin{aligned} \xi_1(T^{n+1}, S_h) : &= \left(\frac{T^{n+1} - T^n}{\Delta t} - \frac{\partial T^{n+1}}{\partial t}, S_h \right) + ((\kappa_{max} - \kappa^{n+1}) \nabla T^{n+1}, \nabla S_h) \\ & - (\kappa_h^m \nabla T^n, \nabla S_h). \end{aligned}$$

Lemma 4. For T satisfying the system (1) and (2) and regularity assumptions (17), the consistency error satisfies

$$(39) \quad \begin{aligned} \|\xi_1(T^{n+1}, S_h)\|^2 &\leq \frac{C_P^2 \Delta t^2}{2\kappa_{min} \epsilon_4} \left\| \frac{\partial^2 T}{\partial t^2} \right\|_{L^2(t^n, t^{n+1}; L^2(\Omega))}^2 + \frac{C_\kappa^2}{2\kappa_{min} \epsilon_6} \|\nabla T^n\|_\infty^2 \|e^n\|^2 \\ &+ \frac{C_\kappa^2 \Delta t}{2\kappa_{min} \epsilon_5} \|\nabla T^{n+1}\|_\infty^2 \left\| \frac{\partial T}{\partial t} \right\|_{L^2(t^n, t^{n+1}; L^2(\Omega))}^2 \\ &+ (\epsilon_7^{-1} + \epsilon_8^{-1}) \Delta t \frac{\kappa_{max}^2}{2\kappa_{min}} \|\nabla T\|_{L^2(t^n, t^{n+1}; L^2(\Omega))}^2 \\ &+ \left(\sum_{i=4}^8 \epsilon_i \right) \|\sqrt{\kappa_h^n} \nabla S_h\|^2. \end{aligned}$$

Moreover, suppose T satisfies the system (1) and (3) and regularity assumption (18). Then,

$$\begin{aligned}
(40) \quad \|\xi_1(T^{n+1}, S_h)\|^2 &\leq \frac{\Delta t^2}{2\kappa_{min}\epsilon_9} \left\| \frac{\partial^2 T}{\partial t^2} \right\|_{L^2(t^n, t^{n+1}; H^{-1}(\Omega))}^2 \\
&+ \frac{C_\kappa^2}{2\kappa_{min}\epsilon_6} \|\nabla T^n\|_\infty^2 \|e^n\|^2 + \frac{C_\kappa^2 \Delta t}{2\kappa_{min}\epsilon_5} \|\nabla T^{n+1}\|_\infty^2 \left\| \frac{\partial T}{\partial t} \right\|_{L^2(t^n, t^{n+1}; L^2(\Omega))}^2 \\
&+ (\epsilon_7^{-1} + \epsilon_8^{-1}) \Delta t \frac{\kappa_{max}^2}{2\kappa_{min}} \left\| \nabla \frac{\partial T}{\partial t} \right\|_{L^2(t^n, t^{n+1}; L^2(\Omega))}^2 \\
&+ \left(\sum_{i=5}^9 \epsilon_i \right) \|\sqrt{\kappa_h^n} \nabla S_h\|_1^2.
\end{aligned}$$

Proof. Recall, $\xi_1(T^{n+1}, S_h)$ is defined by

$$\begin{aligned}
(41) \quad \xi_1(T^{n+1}, S_h) &:= \left(\frac{T^{n+1} - T^n}{\Delta t} - \frac{\partial T^{n+1}}{\partial t}, S_h \right) + ((\kappa_{max} - \kappa^{n+1}) \nabla T^{n+1}, \nabla S_h) \\
&- (\kappa_h^n \nabla T^n, \nabla S_h).
\end{aligned}$$

Applying Taylor's Theorem with integral remainder, Lemma 1, and both Cauchy-Schwarz and Young's inequalities, we have

$$(42) \quad \left(\frac{T^{n+1} - T^n}{\Delta t} - \frac{\partial T^{n+1}}{\partial t}, S_h \right) \leq \frac{C_{PF}^2 \Delta t}{2\kappa_{min}\epsilon_4} \left\| \frac{\partial^2 T}{\partial t^2} \right\|_{L^2(t^n, t^{n+1}; L^2(\Omega))}^2 + \frac{\epsilon_4}{2} \|\sqrt{\kappa_h^n} \nabla S_h\|^2.$$

The last two terms in (41) can be reorganized as,

$$\begin{aligned}
(43) \quad &(\kappa^{n+1} \nabla T^{n+1}, \nabla S_h) - (\kappa_h^n \nabla T^n, \nabla S_h) \\
&= (\kappa_{max} \nabla(T^{n+1} - T^n), \nabla S_h) + (\kappa_h^n \nabla T^n, \nabla S_h) - (\kappa^{n+1} \nabla T^{n+1}, \nabla S_h).
\end{aligned}$$

Adding and subtracting $(\kappa^n \nabla(T^{n+1} - T^n), \nabla S_h)$ to the right-hand side of the above and rearranging yields,

$$\begin{aligned}
(44) \quad &(\kappa_{max} \nabla(T^{n+1} - T^n), \nabla S_h) + (\kappa_h^n \nabla T^n, \nabla S_h) - (\kappa^{n+1} \nabla T^{n+1}, \nabla S_h) \\
&= (\kappa_{max} \nabla(T^{n+1} - T^n), \nabla S_h) - ((\kappa^{n+1} - \kappa^n) \nabla T^{n+1}, \nabla S_h) \\
&- (\kappa^n \nabla(T^{n+1} - T^n), \nabla S_h) - ((\kappa^n - \kappa_h^n) \nabla T^n, \nabla S_h).
\end{aligned}$$

Using properties (9)-(10), Taylor's Theorem with integral remainder in the first term, and both Cauchy-Schwarz and Young's inequalities, leads to

$$\begin{aligned}
(45) \quad &-((\kappa^{n+1} - \kappa^n) \nabla T^{n+1}, \nabla S_h) \leq C_\kappa \|T^{n+1} - T^n\| \|\nabla T^{n+1}\|_\infty \|\nabla S_h\| \\
&\leq \frac{C_\kappa^2 \Delta t}{2\kappa_{min}\epsilon_5} \|\nabla T^{n+1}\|_\infty^2 \left\| \frac{\partial T}{\partial t} \right\|_{L^2(t^n, t^{n+1}; L^2(\Omega))}^2 + \frac{\epsilon_5}{2} \|\sqrt{\kappa_h^n} \nabla S_h\|^2,
\end{aligned}$$

and

$$(46) \quad -((\kappa^n - \kappa_h^n) \nabla T^n, \nabla S_h) \leq \frac{C_\kappa^2}{2\kappa_{min}\epsilon_6} \|\nabla T^n\|_\infty^2 \|e^n\|^2 + \frac{\epsilon_6}{2} \|\sqrt{\kappa_h^n} \nabla S_h\|^2.$$

The other two estimates follow from Taylor's Theorem with integral remainder, Cauchy-Schwarz, and Young's inequality,

(47)

$$(\kappa_{max} \nabla(T^{n+1} - T^n), \nabla S_h) \leq \frac{\kappa_{max}^2 \Delta t}{2\kappa_{min} \epsilon_7} \left\| \nabla \frac{\partial T}{\partial t} \right\|_{L^2(t^n, t^{n+1}; L^2(\Omega))}^2 + \frac{\epsilon_7}{2} \|\sqrt{\kappa_h^n} \nabla S_h\|^2,$$

(48)

$$-(\kappa^n \nabla(T^{n+1} - T^n), \nabla S_h) \leq \frac{\kappa_{max}^2 \Delta t}{2\kappa_{min} \epsilon_8} \left\| \nabla \frac{\partial T}{\partial t} \right\|_{L^2(t^n, t^{n+1}; L^2(\Omega))}^2 + \frac{\epsilon_8}{2} \|\sqrt{\kappa_h^n} \nabla S_h\|^2.$$

Combining the above estimates (41)-(48) yields the result (39). For Robin boundary conditions, the first term of $\xi_1(T^{n+1}, S_h)$ is estimated as

(49)

$$\left(\frac{T^{n+1} - T^n}{\Delta t} - \frac{\partial T^{n+1}}{\partial t}, S_h \right) \leq \frac{\Delta t^2}{2\kappa_{min} \epsilon_9} \left\| \frac{\partial^2 T}{\partial t^2} \right\|_{L^2(t^n, t^{n+1}; H^{-1}(\Omega))}^2 + \frac{\epsilon_9}{2} \|\sqrt{\kappa_h^n} S_h\|_1^2.$$

Combining the above estimates (44)-(49) and using $\|\nabla S_h\| \leq \|S_h\|_1$ yields the result (40). \square

With the consistency error now analyzed, we can now prove the major convergence result.

Theorem 5. *Suppose T satisfies the assumptions of Lemma 4. Moreover, suppose $T_h^0 \in Y_h$ is an approximation of T^0 to within the accuracy of the interpolant. Then, $\exists C_\dagger$ such that scheme (15) satisfies*

(50)

$$\begin{aligned} \|e^N\|^2 + \Delta t \|\sqrt{\kappa_{max}} \nabla e^N\|^2 + \sum_{n=0}^{N-1} (\|e^{n+1} - e^n\|^2 + \Delta t \|\sqrt{\kappa_h^n} \nabla(e^{n+1} - e^n)\|^2) \\ + \frac{\Delta t}{2} \sum_{n=0}^{N-1} \|\sqrt{\kappa_h^n} \nabla e^{n+1}\|^2 \leq C \exp(C_\dagger) \left\{ (\kappa_{min}^{-1} + \kappa_{max}^2 \kappa_{min}^{-1} + 1 + \Delta t) h^{2k+2} \right. \\ \left. + (\kappa_{min}^{-1} + \kappa_{max}(1 + \Delta t + \Delta t^2)) h^{2k} + (1 + \kappa_{min}^{-1} + \kappa_{max}^2 \kappa_{min}^{-1} + \kappa_{max}) \Delta t^2 \right\}. \end{aligned}$$

Moreover, scheme (16) satisfies

(51)

$$\begin{aligned} \|e^N\|^2 + \|\sqrt{\kappa_{max}} \nabla e^N\|^2 + \sum_{n=0}^{N-1} (\|e^{n+1} - e^n\|^2 + \Delta t \|\sqrt{\kappa_h^n} \nabla(e^{n+1} - e^n)\|^2) \\ + \frac{C_P^2 \Delta t}{8} \sum_{n=0}^{N-1} \|\sqrt{\kappa_h^n} \nabla e^{n+1}\|_1^2 \leq C \exp(C_\dagger) \left\{ (\kappa_{min}^{-1} + \kappa_{max}^2 \kappa_{min}^{-1} + 1 + \Delta t) h^{2k+2} \right. \\ \left. + (\kappa_{min}^{-1} + \kappa_{max}(1 + \Delta t + \Delta t^2)) h^{2k} + (1 + \kappa_{min}^{-1} + \kappa_{max}^2 \kappa_{min}^{-1} + \kappa_{max}) \Delta t^2 \right\}. \end{aligned}$$

Proof. Using $e^n = \phi^n - \psi_h^n$, rearrange the error equation (36) with $S_h = 2\Delta t \psi_h^{n+1}$

(52)

$$\begin{aligned} & \|\psi_h^{n+1}\|^2 - \|\psi_h^n\|^2 + \|\psi_h^{n+1} - \psi_h^n\|^2 + 2\Delta t \|\sqrt{\kappa_{max}} \nabla \psi_h^{n+1}\|^2 - 2\Delta t (\kappa_h'^n \nabla \psi_h^n, \nabla \psi_h^{n+1}) \\ &= 2\Delta t \left(\frac{\phi^{n+1} - \phi^n}{\Delta t}, \psi_h^{n+1} \right) + 2\Delta t (\kappa_{max} \nabla \phi^{n+1}, \psi_h^{n+1}) - 2\Delta t (\kappa_h'^n \nabla \phi_h^n, \nabla \psi_h^{n+1}) \\ & \quad - \xi_1(T^{n+1}, 2\Delta t \psi_h^{n+1}). \end{aligned}$$

Recall equations (22) and (23) from Theorem 7, we proceed in similar fashion so that after applications of the polarization identity we arrive at

(53)

$$\begin{aligned} & \|\psi_h^{n+1}\|^2 - \|\psi_h^n\|^2 + \|\psi_h^{n+1} - \psi_h^n\|^2 + \|\sqrt{\kappa_{max}} \nabla \psi_h^{n+1}\|^2 - \|\sqrt{\kappa_{max}} \nabla \psi_h^n\|^2 \\ & \quad + \Delta t \|\sqrt{\kappa_h^n} \nabla (\psi_h^{n+1} - \psi_h^n)\|^2 + \Delta t \|\sqrt{\kappa_h^n} \nabla \psi_h^{n+1}\|^2 + \Delta t \|\sqrt{\kappa_h^n} \nabla \psi_h^n\|^2 \\ &= 2\Delta t \left(\frac{\phi^{n+1} - \phi^n}{\Delta t}, \psi_h^{n+1} \right) + 2\Delta t (\kappa_{max} \nabla \phi^{n+1}, \psi_h^{n+1}) - 2\Delta t (\kappa_h'^n \nabla \phi_h^n, \nabla \psi_h^{n+1}) \\ & \quad - 2\Delta t \xi_1(T^{n+1}, \psi_h^{n+1}). \end{aligned}$$

Now, application of Taylor's Theorem with integral remainder, the Cauchy-Schwarz inequality, Lemma 1, and Young's inequality on the first term on the right-hand-side of (52) yields

(54)

$$2\Delta t \left(\frac{\phi^{n+1} - \phi^n}{\Delta t}, \psi_h^{n+1} \right) \leq \frac{C_{PF}^2}{\epsilon_{10} \kappa_{min}} \|\phi_t\|_{L^2(t^n, t^{n+1}, L^2(\Omega))}^2 + \epsilon_{10} \Delta t \|\sqrt{\kappa_h^n} \nabla \psi_h^{n+1}\|^2.$$

Applying the Cauchy-Schwarz and Young inequalities to the second and third terms yield

$$(55) \quad 2\Delta t (\kappa_{max} \nabla \phi^{n+1}, \nabla \psi_h^{n+1}) \leq \frac{\kappa_{max}^2 \Delta t}{\epsilon_{11} \kappa_{min}} \|\nabla \phi^{n+1}\|^2 + \epsilon_{11} \Delta t \|\sqrt{\kappa_h^n} \nabla \psi_h^{n+1}\|^2,$$

$$(56) \quad 2\Delta t (\kappa_h'^n \nabla \phi_h^n, \nabla \psi_h^{n+1}) \leq \frac{\kappa_{max}^2 \Delta t}{\epsilon_{12} \kappa_{min}} \|\nabla \phi^n\|^2 + \epsilon_{12} \Delta t \|\sqrt{\kappa_h^n} \nabla \psi_h^{n+1}\|^2.$$

Using the above and Lemma 4 in equation (53), with $\epsilon_i = 1/20$ for $i = 4, 5, 6, 7, 8$ and $\epsilon_i = 1/12$ for $i = 10, 11, 12$, yields

$$\begin{aligned}
(57) \quad & \|\psi_h^{n+1}\|^2 - \|\psi_h^n\|^2 + \|\psi_h^{n+1} - \psi_h^n\|^2 + \|\sqrt{\kappa_{max}}\nabla\psi_h^{n+1}\|^2 - \|\sqrt{\kappa_{max}}\nabla\psi_h^n\|^2 \\
& + \Delta t\|\sqrt{\kappa_h^n}\nabla(\psi_h^{n+1} - \psi_h^n)\|^2 + \frac{\Delta t}{2}\|\sqrt{\kappa_h^n}\nabla\psi_h^{n+1}\|^2 + \Delta t\|\sqrt{\kappa_h^n}\nabla\psi_h^n\|^2 \\
& \leq \frac{12C_{PF}^2}{\kappa_{min}}\|\phi_t\|_{L^2(t^n, t^{n+1}; L^2(\Omega))}^2 + \frac{12\kappa_{max}^2\Delta t}{\kappa_{min}}(\|\nabla\phi^{n+1}\|^2 + \|\nabla\phi^n\|^2) \\
& + \frac{10C_{PF}^2\Delta t^2}{\kappa_{min}}\left\|\frac{\partial^2 T}{\partial t^2}\right\|_{L^2(t^n, t^{n+1}; L^2(\Omega))}^2 + \frac{10C_\kappa^2\Delta t}{\kappa_{min}}\|\nabla T^n\|_\infty^2\|e^n\|^2 \\
& + \frac{10C_\kappa^2\Delta t^2}{\kappa_{min}}\|\nabla T^{n+1}\|_\infty^2\left\|\frac{\partial T}{\partial t}\right\|_{L^2(t^n, t^{n+1}; L^2(\Omega))}^2 \\
& + \frac{5\kappa_{max}^2\Delta t^2}{\kappa_{min}}\left\|\nabla\frac{\partial T}{\partial t}\right\|_{L^2(t^n, t^{n+1}; L^2(\Omega))}^2.
\end{aligned}$$

Now, note that $\|e^n\|^2 \leq 2\|\psi^n\|^2 + 2\|\psi_h^n\|^2$, drop $\Delta t\|\sqrt{\kappa_h^n}\nabla\psi_h^n\|^2$, sum from $n = 0$ to $n = N - 1$, and rearrange terms to arrive at

$$\begin{aligned}
(58) \quad & \|\psi_h^N\|^2 + \Delta t\|\sqrt{\kappa_{max}}\nabla\psi_h^N\|^2 + \sum_{n=0}^{N-1}(\|\psi_h^{n+1} - \psi_h^n\|^2 + \Delta t\|\sqrt{\kappa_h^n}\nabla(\psi_h^{n+1} - \psi_h^n)\|^2) \\
& + \frac{\Delta t}{2}\sum_{n=0}^{N-1}\|\sqrt{\kappa_h^n}\nabla\psi_h^{n+1}\|^2 \\
& \leq \frac{20C_\kappa^2}{\kappa_{min}}\sum_{n=0}^{N-1}\|\nabla T^n\|_\infty^2\|\psi_h^n\|^2 + \frac{12C_{PF}^2}{\kappa_{min}}\|\phi_t\|_{L^2(0, t^*, L^2(\Omega))}^2 \\
& + \frac{24\kappa_{max}^2}{\kappa_{min}}\|\|\nabla\phi\|_{2,0}\|^2 + \frac{10C_{PF}^2\Delta t^2}{\kappa_{min}}\left\|\frac{\partial^2 T}{\partial t^2}\right\|_{L^2(0, t^*; L^2(\Omega))}^2 \\
& + \frac{10C_\kappa^2}{\kappa_{min}}\max_{0 \leq n \leq N}\|\nabla T^n\|_\infty^2\left(2\|\phi\|_{2,0}^2 + \Delta t^2\left\|\frac{\partial T}{\partial t}\right\|_{L^2(0, t^*; L^2(\Omega))}^2\right) \\
& + \frac{5\kappa_{max}^2\Delta t^2}{\kappa_{min}}\left\|\nabla\frac{\partial T}{\partial t}\right\|_{L^2(0, t^*; L^2(\Omega))}^2 + \|\psi_h^0\|^2 + \Delta t\|\sqrt{\kappa_{max}}\nabla\psi_h^0\|^2.
\end{aligned}$$

Collecting constants, application of Lemma 2, and a rearrangement yield

$$\begin{aligned}
(59) \quad & \|\psi_h^N\|^2 + \Delta t \|\sqrt{\kappa_{max}} \nabla \psi_h^N\|^2 + \sum_{n=0}^{N-1} (\|\psi_h^{n+1} - \psi_h^n\|^2 + \Delta t \|\sqrt{\kappa_h^n} \nabla (\psi_h^{n+1} - \psi_h^n)\|^2) \\
& + \frac{\Delta t}{2} \sum_{n=0}^{N-1} \|\sqrt{\kappa_h^n} \nabla \psi_h^{n+1}\|^2 \\
& \leq C \exp\left(\frac{20C_\kappa^2}{\kappa_{min}} \Delta t \sum_{n=0}^{N-1} \|\nabla T^n\|_\infty^2\right) \left\{ \kappa_{min}^{-1} (\|\phi_t\|_{L^2(0,t^*,L^2(\Omega))}^2) \right. \\
& + \kappa_{max}^2 \|\nabla \phi\|_{2,0}^2 + \max_{0 \leq n \leq N} \|\nabla T^n\|_\infty^2 \|\phi\|_{2,0}^2 \left. + \Delta t^2 \left(\kappa_{min}^{-1} \left\| \frac{\partial^2 T}{\partial t^2} \right\|_{L^2(0,t^*;L^2(\Omega))}^2 \right. \right. \\
& + \kappa_{min}^{-1} \max_{0 \leq n \leq N} \|\nabla T^n\|_\infty^2 \left\| \frac{\partial T}{\partial t} \right\|_{L^2(0,t^*;L^2(\Omega))}^2 + \kappa_{max}^2 \kappa_{min}^{-1} \left\| \nabla \frac{\partial T}{\partial t} \right\|_{L^2(0,t^*;L^2(\Omega))}^2 \left. \right) \\
& \left. + \|\psi_h^0\|^2 + \Delta t \|\sqrt{\kappa_{max}} \nabla \psi_h^0\|^2 \right\}.
\end{aligned}$$

Denote $C_\dagger = \frac{20C_\kappa^2}{\kappa_{min}} \Delta t \sum_{n=0}^{N-1} \|\nabla T^n\|_\infty^2$, take an infimum over Y_h , apply approximation property (14), and collect constants. Then,

$$\begin{aligned}
(60) \quad & \|\psi_h^N\|^2 + \Delta t \|\sqrt{\kappa_{max}} \nabla \psi_h^N\|^2 + \sum_{n=0}^{N-1} (\|\psi_h^{n+1} - \psi_h^n\|^2 + \Delta t \|\sqrt{\kappa_h^n} \nabla (\psi_h^{n+1} - \psi_h^n)\|^2) \\
& + \frac{\Delta t}{2} \sum_{n=0}^{N-1} \|\sqrt{\kappa_h^n} \nabla \psi_h^{n+1}\|^2 \leq C \exp(C_\dagger) \left\{ \kappa_{min}^{-1} (1 + \kappa_{max}^2 + \kappa_{min}) h^{2k+2} \right. \\
& \left. + \kappa_{min}^{-1} h^{2k} + (1 + \kappa_{min}^{-1} + \kappa_{max}^2 \kappa_{min}^{-1} + \kappa_{max}) \Delta t^2 \right\}.
\end{aligned}$$

Application of the triangle inequality yields the result, estimate (50). For the latter result (51) pertaining to Robin boundary conditions, we have the following estimate for the boundary term:

$$(61) \quad 2\Delta t (\alpha \phi^{n+1}, \psi_h^{n+1})_{\partial\Omega} \leq \frac{\|\alpha\|_{L^\infty(\partial\Omega)}^2 \Delta t}{2\kappa_{min} \epsilon_0} \|\phi_h^{n+1}\|_{-1,\partial\Omega}^2 + \frac{\epsilon_0 \Delta t}{2} \|\sqrt{\kappa_h^n} \psi_h^{n+1}\|_1^2.$$

Using the above estimate, noting that $\|\nabla S\| \leq \|S\|_1$ and selecting $\epsilon_i = C_P^2/40$ for $i = 5, 6, 7, 8, 9$ and $\epsilon_i = C_P^2/16$ for $i = 0, 10, 11, 12$, leads to the result. \square

The above Theorem allows for Lagrange elements of arbitrary polynomial order to be used. However, if $P1$ is used, the optimal, first-order accuracy is achieved with $h = \mathcal{O}(\Delta t)$.

TABLE 1. Errors and rates for algorithm (15)

m	$\ \nabla \langle e \rangle\ _{2,0}$	Rate	$\ \langle e \rangle\ _{\infty,0}$	Rate
4	2.54E-01	-	1.72E-02	-
8	1.27E-01	1	4.23E-03	2.03
16	6.04E-02	1.08	1.04E-03	2.03
32	3.04E-02	0.99	2.80E-04	1.9
64	1.55E-02	0.98	7.29E-05	1.95

Corollary 6. *Suppose the finite element space $Z = Y_h$ or X_h is given by P1. Then, under the assumptions of Theorem 5, the error for Algorithms 1(a) and (b) satisfy (62)*

$$\begin{aligned} & \|e^N\|^2 + \Delta t \|\sqrt{\kappa_{max}} \nabla e^N\|^2 + \sum_{n=0}^{N-1} (\|e^{n+1} - e^n\|^2 + \Delta t \|\sqrt{\kappa_h^n} \nabla (e^{n+1} - e^n)\|^2) \\ & + \frac{\Delta t}{2} \sum_{n=0}^{N-1} \|\sqrt{\kappa_h^n} \nabla e^{n+1}\|^2 \leq C(h^2 + \Delta t^2) \end{aligned}$$

and
(63)

$$\begin{aligned} & \|e^N\|^2 + \Delta t \|\sqrt{\kappa_{max}} \nabla e^N\|^2 + \sum_{n=0}^{N-1} (\|e^{n+1} - e^n\|^2 + \Delta t \|\sqrt{\kappa_h^n} \nabla (e^{n+1} - e^n)\|^2) \\ & + \frac{C_P^2 \Delta t}{8} \sum_{n=0}^{N-1} \|\sqrt{\kappa_h^n} \nabla e^{n+1}\|^2 \leq C(h^2 + \Delta t^2). \end{aligned}$$

6. Numerical Experiments

In this section, we illustrate the stability and convergence of the numerical schemes (15) and (16) using P1 Lagrange elements to approximate temperature distributions. The first numerical experiment tests for convergence and considers the effects of ensemble and perturbation sizes, where an analytical solution is constructed via the method of manufactured solutions. From this test it is shown that the numerical methods (15) and (16) are first-order accurate in the appropriate norms. Moreover, the ensemble and perturbation sizes have little effect on the accuracy in this setting.

The next numerical experiment is a 3D printing application in the spirit of the work by Vora and Dahotre [46] using a temperature-dependent thermal conductivity that resembles data from [48]. A simple application to uncertainty quantification, calculation of error envelopes, is illustrated using temperature-dependent thermal conductivity. The final validation experiment is the steady-state solution of a non-linear heat transfer problem from [47] which is compared to a given analytical solution. The software used for all tests is FREEFEM++ [19].

6.1. Numerical Convergence Study. For the first numerical experiment we will illustrate the convergence rates for the proposed algorithms (15) and (16). Let the

TABLE 2. Errors and rates for algorithm (16)

m	$\ \nabla \langle e \rangle\ _{2,0}$	Rate	$\ \langle e \rangle\ _{\infty,0}$	Rate
4	2.49E-01	-	1.81E-02	-
8	1.29E-01	0.95	4.03E-03	2.17
16	6.07E-02	1.1	1.06E-03	1.93
32	3.05E-02	1	3.69E-04	1.53
64	1.55E-02	0.98	1.57E-04	1.24

domain Ω be the unit square $(0, 1)^2$ and final time $t^* = 1$; see Figure 1 for the domain and boundary conditions. Let $c = 0.1$, with $J = 4$ and $T(x, y, t, \omega_j) = (1 + \epsilon_j)T(x, y, t)$, where $\epsilon_j = \mathcal{O}(10^{-1})$ for $1 \leq j \leq 4$. The manufactured solution and thermal conductivity are

$$\begin{aligned} T(x, y, t) &= 20 \cos(t) (\cos(x(x-1)) \sin(y(y-1)) - y(y-1)) \\ \kappa(T) &= 10 + cT, \end{aligned}$$

where both the heat source and boundary terms are adjusted appropriately. For algorithm (16), the Robin boundary condition is used with $\alpha = 1/2$ and appropriate β .

Remark: The perturbations are randomly generated. For the first test, they are 0.9578666373, 0.9721124752, 0.03562315298, and 0.4332194024.

The finite element mesh Ω_h is a Delaunay triangulation generated from m points on each side of Ω . P1 Lagrange elements are used. We calculate the error in the approximations of the average temperature with the $L^\infty(0, t^*; L^2(\Omega))$ and $L^2(0, t^*; H^1(\Omega))$ norms. Rates are calculated from the errors at two $\Delta t_{1,2}$ via

$$\frac{\log_2(e(\Delta t_1)/e(\Delta t_2))}{\log_2(\Delta t_1/\Delta t_2)}.$$

We set $\Delta t = 0.5/m$ and vary m between 4, 8, 16, 32, and 64. Results are presented in Tables 1 and 2. We see first-order convergence in the $L^\infty(0, t^*; L^2(\Omega))$ norm and first-order convergence in the $L^2(0, t^*; H^1(\Omega))$ norm for each algorithm. These results are as expected based on the convergence analysis, Theorem 5.

For the second test, we study the effect of the size of the perturbation on convergence. We repeat the above convergence test, changing only the perturbation size. For $1 \leq j \leq 4$, let $\epsilon_j = \mathcal{O}(10^{-l})$ for $l = -2, -1, 0, 1, 2, 3$, and 4. The errors of the average solution in $L^\infty(0, t^*; L^2(\Omega))$ are presented in Tables 3 and 4 for methods (15) and (16), respectively. We see that as the perturbation size is reduced, the results increasingly agree with one another. Notably, the algorithm remains stable irrespective of the perturbation size, consistent with Theorem 3.

Finally, we investigate the effect of the ensemble size J . We fix $m = 16$, $\Delta t = 0.5/m$, $\epsilon_j = \mathcal{O}(10^{-1})$, and then let J vary from 1, 2, 4, 8, 16, 32, and 64. The associated average errors are calculated and plotted, Figure 2. Once again, results are consistent with the theory.

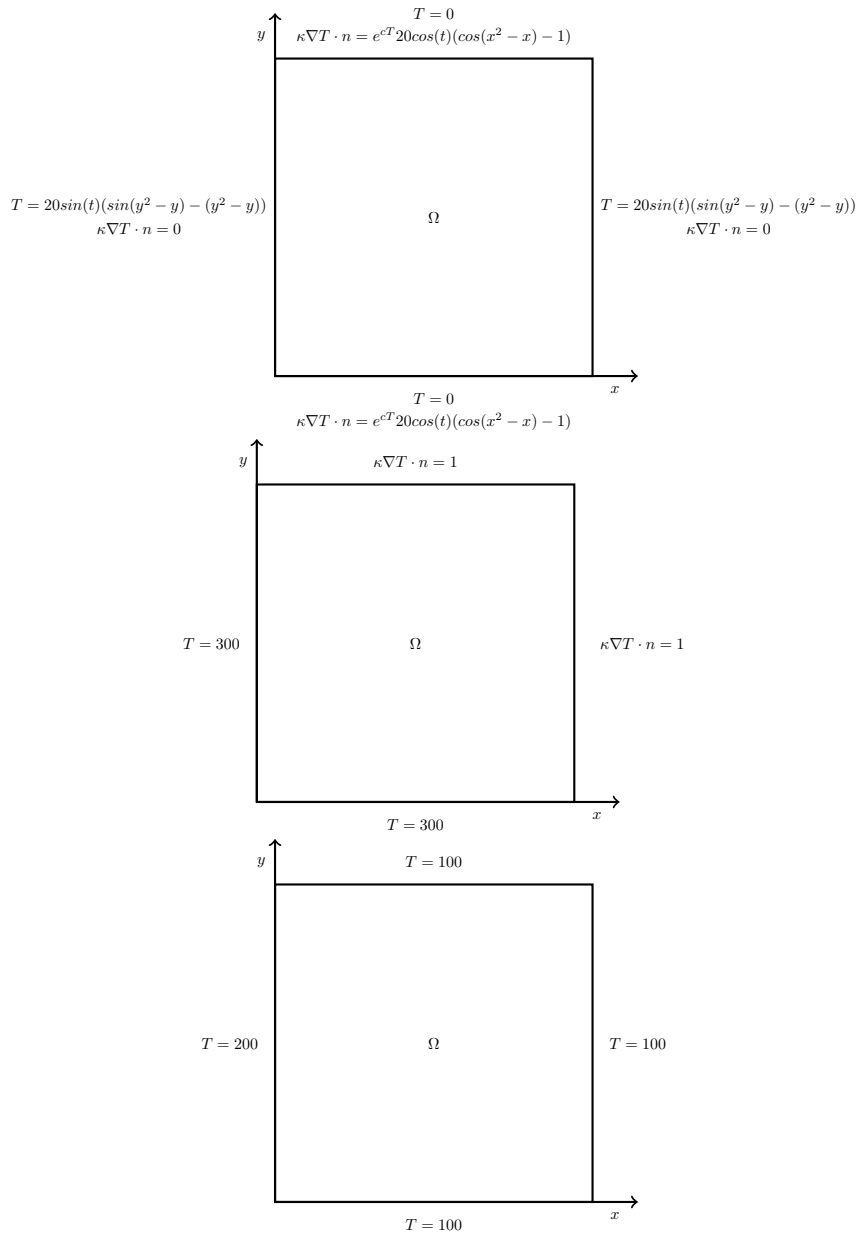


FIGURE 1. Domain and boundary conditions for (top) convergence test manufactured solution, (middle) 3D printing problem, and (bottom) steady-state solution.

TABLE 3. Comparison of $\| \langle e \rangle \|_{\infty,0}$ with algorithm (15), varying perturbation size, ϵ .

Mesh m	Perturbation Size					
	$\mathcal{O}(10^{-2})$	$\mathcal{O}(10^{-1})$	$\mathcal{O}(10^0)$	$\mathcal{O}(10^1)$	$\mathcal{O}(10^2)$	$\mathcal{O}(10^3)$
4	1.09E-02	1.15E-02	1.72E-02	8.17E-02	1.42E+00	1.62E+01
8	2.65E-03	2.79E-03	4.23E-03	5.22E-02	1.67E+00	2.10E+01
16	6.52E-04	6.86E-04	1.04E-03	3.05E-02	1.37E+00	2.06E+01
32	1.77E-04	2.07E-04	2.80E-04	1.63E-02	9.48E-01	1.88E+01
64	1.01E-04	1.17E-04	7.29E-05	8.40E-03	5.63E-01	1.58E+01

TABLE 4. Comparison of $\| \langle e \rangle \|_{\infty,0}$ with algorithm (16), varying perturbation size, ϵ .

Mesh m	Perturbation Size					
	$\mathcal{O}(10^{-2})$	$\mathcal{O}(10^{-1})$	$\mathcal{O}(10^0)$	$\mathcal{O}(10^1)$	$\mathcal{O}(10^2)$	$\mathcal{O}(10^3)$
4	1.15E-02	1.21E-02	1.81E-02	8.55E-02	1.01E+00	1.16E+01
8	2.54E-03	2.67E-03	4.03E-03	3.86E-02	1.27E+00	1.59E+01
16	6.36E-04	6.66E-04	1.06E-03	2.34E-02	1.07E+00	1.60E+01
32	2.07E-04	2.15E-04	3.69E-04	1.27E-02	7.57E-01	1.49E+01
64	8.98E-05	9.44E-05	1.57E-04	6.59E-03	4.58E-01	1.27E+01

TABLE 5. Comparison of steady-state solutions with exact solution.

Position		$m = 8$		$m = 16$		Analytical
x	y	T	% Error	T	% Error	T
0.25	0.50	161.939	3.5	161.919	1.5	161.904
0.375	0.630	143.281	2.8	143.259	0.6	143.253
0.50	0.50	132.309	2.1	132.293	0.5	132.288
0.50	0.75	124.361	1.9	124.342	0	124.342
0.625	0.625	120.343	1.5	120.332	0.4	120.328
0.75	0.50	113.423	1	113.415	0.2	113.413
0.75	0.75	109.731	0.8	109.725	0.2	109.723
0.25	0.75	151.584	6.5	151.541	2.2	151.519

6.2. 3D Printing Application. We now consider an application problem in the spirit of [46] to illustrate the use of ensembles. The problem is the two-dimensional heat transfer of a solid medium subject to laser heating from above by a single pulse. We implement the thermal conductivity for Aluminum Oxide found in [48] which is dependent on the temperature of the material. Moreover, we set $J = 12$ whereby for each ensemble member we generate two random perturbations $\omega_j, \xi_j \in \mathcal{O}(10^{-1})$ for the initial condition $T(x, y, 0; \omega_j, \xi_j) = T_{0,j} = 300(1 + \omega_j/10) + 100(\xi_j - 0.5)$. The lower left corner walls are maintained at temperature $T(1, y, t; \omega_j) = T(x, 0, t; \omega_j) = 300$ and upper right corner walls allow for heat flow out of the element via

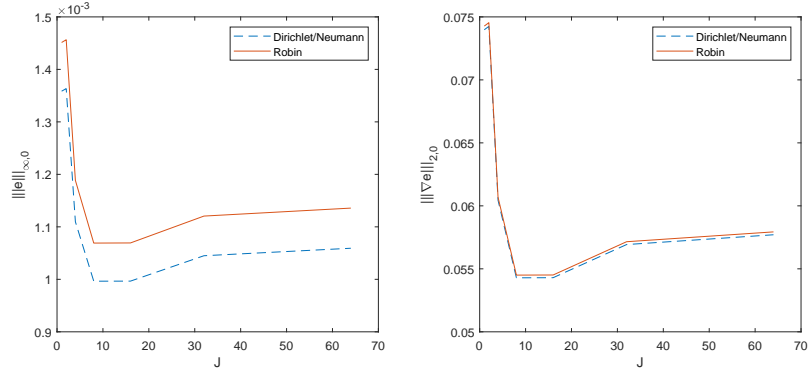


FIGURE 2. Errors versus ensemble size J .

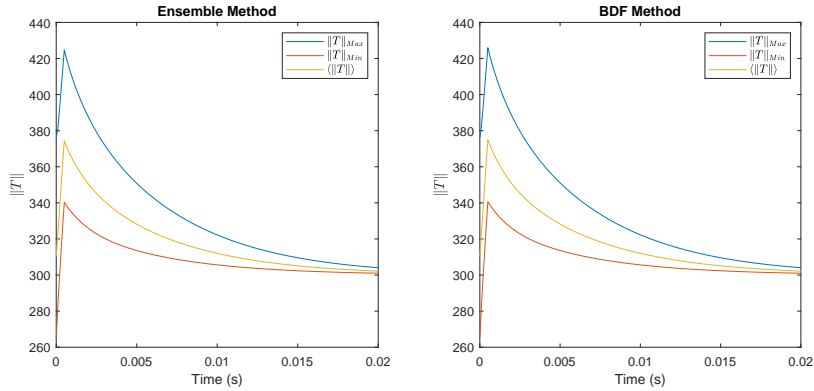


FIGURE 3. Comparison of the approximated maximum, minimum and average temperatures obtained using the ensemble method (left) and typical backwards differentiation formula (right).

$\kappa \nabla T \cdot n = 1$; see Figure 1. Moreover, the heat source, $f(x, y, t)$, is given by

$$f(x, y, t; \omega_j) = \begin{cases} (400000) \exp(-8((x - 0.5)^2 + (y - 0.5)^2)) & 0 \leq t \leq 0.0005, \\ 0 & 0.0005 < t, \end{cases}$$

representing a pulse laser with Gaussian beam profile.

The finite element mesh is a division of $(0, 1)^2$ into 64^2 squares with diagonals connected with a line within each square in the same direction. First, we use algorithm (15) with timestep $\Delta t = 0.0001$ and final time $t^* = 0.02$. Next we solve the ensemble members individually using a typical backwards differentiation formula, providing a comparison of the solutions using each method. The values for the maximum and minimum approximate temperature distributions and mean distribution in the $L^2(\Omega)$ norm are computed and presented in Figure 3, we can see the results from the ensemble method agree with with a more traditional method. We

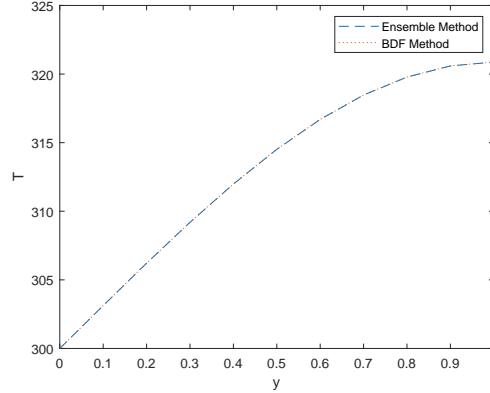


FIGURE 4. Mean temperature approximation of all ensemble members at $t = 0.01$, measured along $x = 0.5$.

see that the temperature approximations generated by perturbed thermal conductivities envelop the mean, evidently useful in quantifying uncertainty. Additionally we compare the mean temperature distribution along the center of the domain at $x = 0.5$ at the time $t = 0.01$ in Figure 4.

6.3. Steady State Experiment. The final numerical experiment is the solution of a two-dimensional steady-state nonlinear heat transfer problem with temperature dependent thermal conductivity as performed in [47]. We use a single ensemble with $J = 1$ and initial condition $T(x, y, 0) = 100$. The left wall is set at $T(0, y, t) = 200$ and the remaining boundaries are $T(1, y, t) = T(x, 0, t) = T(x, 1, t) = 100$. We use the heat source $f(x, y, t) = 0$, and set $\kappa(T) = \frac{\kappa_0}{c\rho}T$ where $\kappa_0 = 400$ is the reference thermal conductivity, and the specific heat and density are set to be $c = 400$ and $\rho = 9000$, respectively. These boundary conditions can be seen in Figure 1.

The finite element mesh is a uniform division of the domain $(0, 1)^2$ into 8^2 and 16^2 squares whose diagonals are connected with a line in the same direction for each square. Values of the steady-state solution at each mesh size are approximated and presented with a comparison to the analytical solution given from [47] in Table 5; from this we can see Algorithm (15) reproduces the steady-state solution with high accuracy.

7. Conclusion

We presented two algorithms for calculating an ensemble of solutions to heat conduction problems with uncertain temperature-dependent thermal conductivity. In particular, these algorithms required the solution of a linear system, involving a shared coefficient matrix, for multiple right-hand sides at each timestep. Unconditional stability and convergence of the algorithms were proven. Moreover, numerical experiments were performed to illustrate the use of ensembles and the proven properties. Important next steps include allowing for phase changes in the

solid material (e.g., liquid phase) and incorporating more physics in the boundary conditions (e.g., surface-to-ambient radiation).

Acknowledgements

J. A. Fiordilino and M. Winger are supported by NISE/Section 219.

Public Release

Distribution Statement A: Approved for public release; distribution is unlimited. (NSWC Corona Public Release Control Number 21-012).

References

- [1] M. Anitescu, F. Pahlevani, and W. J. Layton, Implicit for local effects and explicit for nonlocal effects is unconditionally stable, *Electronic Transactions on Numerical Analysis* 18 (2004) 174-187.
- [2] F. Atger, The skill of ensemble prediction systems, *Monthly Weather Review* 127 (1999) 1941-1953.
- [3] A. Ern and J.-L. Guermond, *Theory and Practice of Finite Elements*, Springer-Verlag, New York, 2004.
- [4] L. C. Evans, *Partial Differential Equations*, American Mathematical Society, Rhode Island, 2010.
- [5] J. Charney, The Use of the Primitive Equations of Motion in Numerical Prediction, *Tellus* 7 (1955) 22-26.
- [6] G. Chen, L. Pi, L. Xu, and Y. Zhang, A Superconvergent Ensemble HDG Method for Parameterized Convection Diffusion Equations, *SIAM Journal on Numerical Analysis* 57 (2019) 2551-2578.
- [7] Y. T. Feng, D.R.J. Owen, and D. Perić, A block conjugate gradient method applied to linear systems with multiple right-hand sides, *Computer Methods in Applied Mechanics and Engineering* 127 (1995) 203-215.
- [8] J. A. Fiordilino, A second order ensemble timestepping algorithm for natural convection, *SIAM Journal on Numerical Analysis* 56 (2018) 816-837.
- [9] J. A. Fiordilino, Ensemble timestepping algorithms for the heat equation with uncertain conductivity, *Numerical Methods for Partial Differential Equations* 34 (2018) 1901-1916.
- [10] J. A. Fiordilino and S. Khankan, Ensemble Timestepping Algorithms for Natural Convection, *International Journal of Numerical Analysis and Modeling* 15 (2018) 524-551.
- [11] J. A. Fiordilino and Y. Li, New, efficient, and unconditional methods for calculating ensembles to the heat equation with temperature dependent thermal conductivity, *arXiv preprint* (2021).
- [12] J. A. Fiordilino and M. McLaughlin, An artificial compressibility ensemble timestepping algorithm for flow problems, *arXiv preprint* (2017) arXiv:1712.06271.
- [13] J. M. Fritsch, J. Hilliker, J. Ross, and R. L. Vislocky, Model consensus, *Weather and Forecasting* 15 (2000) 571-582.
- [14] M. Gunzburger, N. Jiang, and Z. Wang, An efficient algorithm for simulating ensembles of parameterized flow problems, *IMA Journal of Numerical Analysis* 39 (2019) 1180-1205.
- [15] M. Gunzburger, N. Jiang, and Z. Wang, A Second-Order Time-Stepping Scheme for Simulating Ensembles of Parameterized Flow Problems, *Computer Methods in Applied Mechanics and Engineering* 19 (2017).
- [16] M. D. Gunzburger, C. G. Webster, and G. Zhang, Stochastic finite element methods for partial differential equations with random input data, *Acta Numerica* 23 (2014) 521-650.
- [17] M. Gunzburger, T. Iliescu, M. Schneier, A Leray regularized ensemble-proper orthogonal decomposition method for parameterized convection-dominated flows, *IMA Journal of Numerical Analysis* 40 (2020) 886-913.

- [18] M. H. Gutknecht, Block Krylov Space Methods for Linear Systems with Multiple Right-Hand Sides: An Introduction, In *Modern Mathematical Models, Methods and Algorithms for Real World Systems*, Anamaya Publishers, New Delhi, India, 2006.
- [19] F. Hecht, New development in freefem++, *Journal of Numerical Mathematics* 20 (2012) 251-265.
- [20] J. G. Heywood and R. Rannacher, Finite-Element Approximation of the Nonstationary Navier-Stokes Problem Part IV: Error Analysis for Second-Order Time Discretization, *SIAM Journal on Numerical Analysis* 27 (1990) 353-384.
- [21] N. Jiang, A Higher Order Ensemble Simulation Algorithm for Fluid Flows, *Journal of Scientific Computing* 64 (2015) 264-288.
- [22] N. Jiang, S. Kaya, and W. Layton, Analysis of model variance for ensemble based turbulence modeling, *Computational Methods in Applied Mathematics* 15 (2015) 173-188.
- [23] N. Jiang and W. Layton, An Algorithm for Fast Calculation of Flow Ensembles, *International Journal for Uncertainty Quantification* 4 (2014) 273-301.
- [24] N. Jiang and C. Qiu, An efficient ensemble algorithm for numerical approximation of stochastic Stokes-Darcy equations, *Computer Methods in Applied Mechanics and Engineering* 343 (2019) 249-275.
- [25] V. John, *Finite Element Methods for Incompressible Flow Problems*, Springer Nature, Cham, Switzerland, 2017.
- [26] L. Ju, W. Leng, Z. Wang, and S. Yuan, Numerical investigation of ensemble methods with block iterative solvers for evolution problems, *Discrete and Continuous Dynamical Systems, Series B* (2019).
- [27] E. Kalnay, *Atmospheric modeling, data assimilation and predictability*, Cambridge University Press, New York, 2003.
- [28] E. Kalnay and M. Ham, Forecasting forecast skill in the Southern Hemisphere, *Preprints of the 3rd International Conference on Southern Hemisphere Meteorology and Oceanography*, Buenos Aires, 1989, 13-17.
- [29] G. E. Karniadakis and J. Glimm, Uncertainty quantification in simulation science, *Journal of Computational Physics* 217 (2006) 1-4.
- [30] W. J. Layton and C. Trenchea, Stability of the IMEX methods, CNLF and BDF2-AB2, for uncoupling systems of evolution equations, *Applied Numerical Mathematics* 62 (2012) 112-120.
- [31] P. Lermusiaux, Uncertainty estimation and prediction for interdisciplinary ocean dynamics, *Journal of Computational Physics* 217 (2006) 176-199.
- [32] J. M. Lewis, Roots of Ensemble Forecasting, *Monthly Weather Review* 133 (2005) 1865-1885.
- [33] N. Li, J. Fiordilino, and X. Feng, Ensemble time-stepping algorithm for the convection-diffusion equation with random diffusivity, *Journal of Scientific Computing* 79 (2) 1271-1293.
- [34] E. N. Lorenz, The predictability of hydrodynamic flow, *Transactions of the New York Academy of Sciences, Series II* 25 (1963) 409-432.
- [35] E. N. Lorenz, Deterministic non-periodic flow, *Journal of Atmospheric Sciences*, 20 (1963) 130-141.
- [36] E. N. Lorenz, A study of predictability of a 28-variable atmospheric model, *Tellus* 17 (1965) 321-333.
- [37] E. N. Lorenz The predictability of a flow which possesses many scales of motion, *Tellus* 21 (1968) 289-307.
- [38] Y. Luo and Z. Wang, An Ensemble Algorithm for Numerical Solutions to Deterministic and Random Parabolic PDEs, *SIAM Journal on Scientific Computing* 56 (2018) 859-876.
- [39] Y. Luo and Z. Wang, A Multilevel Monte Carlo Ensemble Scheme for Random Parabolic PDEs, *SIAM Journal on Scientific Computing* 41 (2019) A622-A642.
- [40] M. Mohebujjaman and L. Rebholz, An efficient algorithm for computation of MHD flow ensembles, *Computational Methods in Applied Mathematics* 17 (2017) 121-137.
- [41] N. A. Phillips, An Example of Non-Linear Computational Instability, *The Atmosphere and the Sea in Motion* 1959 501-504.
- [42] N. Sakthivel, J. A. Fiordilino, D. Banh, S. Sanyal, and H. Vora, Development of an Integrated Laser-aided Metal Additive Manufacturing System with Real-time Process, *Dimensions*,

- and Property Monitoring, Measurements and Control, TMS 2017 146th Annual Meeting & Exhibition, San Diego, CA, 2017.
- [43] A. Takhirov, M. Neda, and J. Waters, Time Relaxation Algorithm for Flow Ensembles, *Numerical Methods for Partial Differential Equations* 32 (2016) 757-777.
- [44] V. Thomée, Galerkin finite element methods for parabolic problems, Springer-Verlag, Berlin, 1984.
- [45] P. D. Thompson, Uncertainty of Initial State as a Factor in the Predictability of Large Scale Atmospheric Flow Patterns, *Tellus* 9 (1957) 275-295.
- [46] H. D. Vora and N. B. Dahotre, Laser Machining of Structural Alumina: Influence of Moving Laser Beam on the Evolution of Surface Topography, *International Journal of Applied Ceramic Technology* 12 (2015) 665-678.
- [47] Singh, Akhilendra and Singh, Indra Vir and Prakash, Ravi, "Numerical solution of temperature-dependent thermal conductivity problems using a meshless method," *Numerical Heat Transfer, Part A: Applications*, Vol. 50, No. 2, 125-145, (2006). Taylor & Francis.
- [48] Kolev, Nikolay Ivanov, Multiphase Flow Dynamics 5: Nuclear Thermal Hydraulics, Springer (2015) 764-765.

Appendix A. Analysis with Random Initial Condition

Motivated by the analysis performed in [33], we wish to provide the stability and convergence of (11) with mixed and Robin boundary conditions, where the initial condition is a real-valued random field on the problem domain. First we must redefine the problem (1) to include this randomness.

Note: The symbol for the domain Ω used in the previous analysis has been changed to D , as Ω will be used to represent the event space of the complete probability space.

A.1. Problem Setting. Let $D \subset \mathbb{R}^d$ be an open, bounded, Lipschitz domain and $(\Omega, \mathcal{F}, \mathcal{P})$ be a complete probability space with $\omega \in \Omega$ denoting a sampling point, where Ω is the event space, $\mathcal{F} \subset 2^\Omega$ is the σ -algebra, and $\mathcal{P} : \mathcal{F} \rightarrow [0, 1]$ is a probability measure. Given a random initial temperature $T(\omega, x, 0) = T^0(\omega, x)$, temperature dependent thermal conductivity $\kappa(T)$ and heat source f , find a random field $T : \Omega \times D \times [0, T] \rightarrow \mathbb{R}$, such that \mathcal{P} -almost satisfies:

$$(A.1) \quad \frac{\partial T}{\partial t} - \nabla \cdot [\kappa(\omega, x, t) \nabla T] = f(\omega, x, t) \quad \text{in } \Omega \times D \times [0, T],$$

where $T = T(\omega, x, t)$. For the mixed boundary condition, the boundary ∂D is partitioned such that $\partial D = \overline{\Gamma_D} \cup \overline{\Gamma_N}$ with $\Gamma_D \cap \Gamma_N = 0$ (Γ_D for Dirichlet condition and Γ_N for Neumann condition). Let n denote the outward normal, then

$$(A.2) \quad T = 0 \text{ on } \Gamma_D, \quad \nabla T \cdot n = 0 \text{ on } \Gamma_N.$$

Moreover, the Robin condition is prescribed via

$$(A.3) \quad \alpha T + \kappa(\omega, x, t) \nabla T \cdot n = \beta(\omega, x) \text{ on } \partial D,$$

where $\alpha \in [0, 1]$ is the emissivity, and β a prescribed function on the boundary. The thermal conductivity, $\kappa(\omega, x, t)$, external force term $f(\omega, x, t)$, $\beta(\omega, x)$ and initial condition $T^0(\omega, x)$ are real-valued random fields or processes on the physical domain D ; that is

$$\kappa(\omega, x, t), T^0(\omega, x) : \Omega \times \bar{D} \rightarrow \mathbb{R}, \quad f(\omega, x, t) : \Omega \times \bar{D} \times [0, T] \rightarrow \mathbb{R}.$$

Additionally we have the following assumptions of the data:

(A1) The coefficient $\kappa(\omega, x, t)$ is uniformly bounded and coercive:
 $\exists \kappa_{min}, \kappa_{max} \in (0, \infty)$ such that $\mathcal{P}\{\omega \in \Omega : \kappa(\omega, x, t) \in [\kappa_{min}, \kappa_{max}], \forall x \in D\} = 1$;

(A2) The function $f(\omega, x, t)$ belongs to the space $L^2(0, T; L^2(D))$ and is square integrable with respect to \mathcal{P} in the sense of:

$$\int_{\Omega} \|f(\omega, x, t)\|_{L^2(0, T; L^2(D))}^2 d\mathcal{P}(\omega) < \infty;$$

(A2) The function $\beta(\omega, x)$ belongs to the space $L^2(\partial D)$ and is square integrable with respect to \mathcal{P} in the sense of:

$$\int_{\Omega} \|\beta(\omega, x)\|_{L^2(\partial D)}^2 d\mathcal{P}(\omega) < \infty.$$

A.2. Weak Formulation. Finally, in addition to the Hilbert spaces and associated norms defined in the previous analysis, we introduce the following Hilbert space

$$V = L^2(0, T; H_0^1(D)) \otimes L^2_{\mathcal{P}}(\Omega) \text{ with norm } \|T\|_V^2 = \int_0^T \int_D \mathbb{E}[|\nabla T|^2] dx dt,$$

where \mathbb{E} stands for the expectation. Now we introduce the weak form of the problem (A.1), defined as follows for the problem with mixed boundary conditions: a function $T \in V$ is a weak solution of (A.1) if it satisfies the initial condition $T(\omega, x, 0) = T^0(\omega, x)$ and for $T > 0$:

$$(A.4) \quad \int_D \mathbb{E}\left[\frac{\partial T}{\partial t} S\right] + \int_D \mathbb{E}[\kappa \nabla T \cdot \nabla S] = \int_D \mathbb{E}[f S], \quad \forall S \in H_0^1(D) \otimes L^2_{\mathcal{P}}(\Omega).$$

Where for Robin boundary conditions the terms $\int_{\partial D} \mathbb{E}[\alpha T^{n+1} S]$ and $\int_{\partial D} [\beta S]$ appear on the left- and right-hand sides, respectively.

A.3. Parameterization of Stochastic Equation. By applying the KL expansion as used in [33] we can reduce the infinite probability space to a K -dimensional space. With this we assume that $\{\xi_i(\omega)\}_{i=1}^K$ are random variables with probability density functions $\rho_i : \Gamma_i \rightarrow \mathbb{R}^+$ and their images $\Gamma_i = \xi_i(\Omega)$ are bounded intervals in \mathbb{R} for $i = 1, \dots, K$. Then, the joint probability density of $\xi = (\xi_1, \dots, \xi_K)$ is $\rho(\xi) = \prod_{i=1}^K \rho_i(\xi_i), \forall \xi \in \Gamma$, with the support $\Gamma = \prod_{i=1}^K \Gamma_i \subset \mathbb{R}^K$.

Now we can rewrite the original problem as the following finite dimensional problem:

$$(A.5) \quad \frac{\partial T}{\partial t} - \nabla \cdot [\kappa(\xi, x, t) \nabla T] = f(\xi, x, t) \quad \text{in } \Gamma \times D \times [0, T],$$

with initial condition $T(\xi, x, 0) = T^0(\xi, x)$ and where $T = T(\xi, x, t)$. The mixed boundary condition is given as

$$(A.6) \quad T = 0 \text{ on } \Gamma_D, \quad \nabla T \cdot n = 0 \text{ on } \Gamma_N.$$

Moreover, the Robin condition is prescribed via

$$(A.7) \quad \alpha T + \kappa(\xi, x, t) \nabla T \cdot n = \beta(\xi, x) \text{ on } \partial D,$$

Therefore, the weak form (A.4) has an equivalent form: a function $T \in V_\rho$ is a weak solution of (A.5) if it satisfies $T(\xi, x, 0) = T^0(\xi, x)$,

$$(A.8) \quad \int_\Gamma \left(\frac{\partial T}{\partial t}, S \right) \rho d\xi + \int_\Gamma (\kappa \nabla T, \nabla S) \rho d\xi = \int_\Gamma (f, S) \rho d\xi.$$

A.4. Numerical Scheme. Applying the method proposed in [33] to the weak form (A.8), we will need to solve the following fully-discrete methods for mixed and Robin boundary conditions, which are similar to those introduced in Section 3. Use the uniform time partition on $[0, T]$ with the time step $\delta T = T/N$. Define $T_j^n = T(\xi^j, x, t_n)$, $f_j^{n+1} = f(\xi^j, x, t_{n+1})$ and $\kappa' = \kappa_{max} - \kappa_j$. The fully discrete schemes for mixed and Robin boundary conditions are given by:

Algorithm 2: (a) Given $T_h^n \in Y_h$, find $T_h^{n+1} \in Y_h$ satisfying

$$(A.9) \quad \left(\frac{T_{j,h}^{n+1} - T_{j,h}^n}{\Delta t}, S_h \right) + (\kappa_{max} \nabla T_{j,h}^{n+1}, \nabla S_h) - (\kappa_{j,h}^n \nabla T_{j,h}^n, \nabla S_h) \\ = (f_j^{n+1}, S_h) \quad \forall S_h \in Y_h(D).$$

(b) Given $T_h^n \in X_h$, find $T_h^{n+1} \in X_h$ satisfying the fully discrete scheme as follows:

$$(A.10) \quad \left(\frac{T_{j,h}^{n+1} - T_{j,h}^n}{\Delta t}, S_h \right) + (\kappa_{max} \nabla T_{j,h}^{n+1}, \nabla S_h) - (\kappa_{j,h}^n \nabla T_{j,h}^n, \nabla S_h) + (\alpha T_{j,h}^{n+1}, S_h)_{\partial\Omega} \\ = (f_j^{n+1}, S_h) + (\beta, S_h)_{\partial\Omega} \quad \forall S_h \in X_h(D).$$

Additionally, we presume the regularity assumptions (17)-(18) hold for the mixed and Robin boundary conditions, respectively.

A.5. Stability. We can follow an identical analysis to Section 4 for Algorithms 2a and 2b to obtain the following Theorem:

Theorem 7. Consider Algorithm 2(a) and suppose $f_j \in L^2(0, t^*; H^{-1}(D))$ and $\beta_j \in H^{-1}(\partial D)$, then

$$(A.11) \quad \|T_{j,h}^N\|^2 + \|\sqrt{\kappa_{max}} \nabla T_{j,h}^N\|^2 + \sum_{n=0}^{N-1} \left(\|T_{j,h}^{n+1} - T_{j,h}^n\|^2 + \Delta t \|\sqrt{\kappa_{j,h}^n} \nabla (T_{j,h}^{n+1} - T_{j,h}^n)\|^2 \right) \\ + \frac{\Delta t}{2} \sum_{n=0}^{N-1} \|\sqrt{\kappa_{j,h}^n} \nabla T_{j,h}^{n+1}\|^2 \leq \|T_{j,h}^0\|^2 + \|\sqrt{\kappa_{max}} \nabla T_{j,h}^0\|^2 + \frac{2\Delta t}{\kappa_{min}} \sum_{n=0}^{N-1} \|f_j^{n+1}\|_{-1}^2.$$

Moreover, for Algorithm 2(b), we have

$$\begin{aligned}
 (A.12) \quad & \|T_{j,h}^N\|^2 + \|\sqrt{\kappa_{max}}\nabla T_{j,h}^N\|^2 + \sum_{n=0}^{N-1} \left(\|T_{j,h}^{n+1} - T_{j,h}^n\|^2 + \Delta t \|\sqrt{\kappa_{j,h}^n} \nabla (T_{j,h}^{n+1} - T_{j,h}^n)\|^2 \right) \\
 & + \frac{C_P^2 \Delta t}{8} \sum_{n=0}^{N-1} \|\sqrt{\kappa_{j,h}^n} \nabla T_{j,h}^{n+1}\|_1^2 \\
 & \leq \|T_{j,h}^0\|^2 + \|\sqrt{\kappa_{max}}\nabla T_{j,h}^0\|^2 + \frac{4\Delta t}{C_P^2 \kappa_{min}} \sum_{n=0}^{N-1} \left(\|f_j^{n+1}\|_{-1}^2 + 2\|\beta_j\|_{-1,\partial D}^2 \right).
 \end{aligned}$$

A.6. Convergence. Again, following the method used in Section 5 we have the following error bounds for mixed and Robin boundary conditions, respectively,

$$\begin{aligned}
 (A.13) \quad & \|e_j^N\|^2 + \Delta t \|\sqrt{\kappa_{max}}\nabla e_j^N\|^2 + \sum_{n=0}^{N-1} \left(\|e_j^{n+1} - e_j^n\|^2 + \Delta t \|\sqrt{\kappa_{j,h}^n} \nabla (e_j^{n+1} - e_j^n)\|^2 \right) \\
 & + \frac{\Delta t}{2} \sum_{n=0}^{N-1} \|\sqrt{\kappa_{j,h}^n} \nabla e_j^{n+1}\|^2 \leq C(h^2 + \Delta t^2),
 \end{aligned}$$

and

$$\begin{aligned}
 (A.14) \quad & \|e_j^N\|^2 + \Delta t \|\sqrt{\kappa_{max}}\nabla e_j^N\|^2 + \sum_{n=0}^{N-1} \left(\|e_j^{n+1} - e_j^n\|^2 + \Delta t \|\sqrt{\kappa_{j,h}^n} \nabla (e_j^{n+1} - e_j^n)\|^2 \right) \\
 & + \frac{C_P^2 \Delta t}{8} \sum_{n=0}^{N-1} \|\sqrt{\kappa_{j,h}^n} \nabla e_j^{n+1}\|^2 \leq C(h^2 + \Delta t^2).
 \end{aligned}$$

If we apply the expectation to the above, we obtain the following Theorem, similar to the one proven in [33]:

Theorem 8. *Suppose T satisfies the equation (A.5) with boundary conditions (A.6) and (A.7). Moreover, suppose $T_h^0 \in Y_h$ is an approximation of T^0 to within the accuracy of the interpolant. Then, $\exists C$ such that Algorithm 2(a) satisfies*

$$\begin{aligned}
 (A.15) \quad & \mathbb{E} [\|e_j^N\|^2] + \Delta t \mathbb{E} [\|\sqrt{\kappa_{max}}\nabla e_j^N\|^2] \\
 & + \sum_{n=0}^{N-1} \mathbb{E} \left[(\|e_j^{n+1} - e_j^n\|^2 + \Delta t \|\sqrt{\kappa_{j,h}^n} \nabla (e_j^{n+1} - e_j^n)\|^2) \right] \\
 & + \frac{\Delta t}{2} \sum_{n=0}^{N-1} \mathbb{E} \left[\|\sqrt{\kappa_{j,h}^n} \nabla e_j^{n+1}\|^2 \right] \leq C(h^2 + \Delta t^2)
 \end{aligned}$$

and Algorithm 2(b) satisfies

$$\begin{aligned}
 \text{(A.16)} \quad & \mathbb{E} [\|e^N\|^2] + \Delta t \mathbb{E} [\|\sqrt{\kappa_{max}} \nabla e^N\|^2] \\
 & + \sum_{n=0}^{N-1} \mathbb{E} \left[(\|e^{n+1} - e^n\|^2 + \Delta t \|\sqrt{\kappa_h^n} \nabla (e^{n+1} - e^n)\|^2) \right] \\
 & + \frac{C_P^2 \Delta t}{8} \sum_{n=0}^{N-1} \mathbb{E} [\|\sqrt{\kappa_h^n} \nabla e^{n+1}\|^2] \leq C(h^2 + \Delta t^2).
 \end{aligned}$$

¹Naval Surface Warfare Center Corona Division, Measurement Science and Engineering Department, Corona, CA 92860.

E-mail: joseph.a.fiordilino1@navy.mil (J. A. Fiordilino) and matthew.winger@navy.mil (M. Winger)