



Numerical artifacts in the discontinuous Generalized Porous Medium Equation: How to avoid spurious temporal oscillations

Danielle C. Maddix^{a,*}, Luiz Sampaio^b, Margot Gerritsen^{a,b}

^a Institute of Computational and Mathematical Engineering, Stanford University, United States

^b Energy Resources Engineering, Stanford University, United States

ARTICLE INFO

Article history:

Received 19 January 2018

Received in revised form 17 March 2018

Accepted 16 April 2018

Available online 1 May 2018

Keywords:

Discontinuous Generalized Porous Medium Equation

Stefan problem

Nonlinear degenerate parabolic equations

Temporal oscillations

Numerical shock detection

Jump condition

ABSTRACT

Numerical discretizations of the Generalized Porous Medium Equation (GPME) with discontinuous coefficients are analyzed with respect to the formation of numerical artifacts. In addition to the degeneracy and self-sharpening of the GPME with continuous coefficients, detailed in [1], increased numerical challenges occur in the discontinuous coefficients case. These numerical challenges manifest themselves in spurious temporal oscillations in second order finite volume discretizations with both arithmetic and harmonic averaging. The integral average, developed in [2], leads to improved solutions with monotone and reduced amplitude temporal oscillations. In this paper, we propose a new method called the Shock-Based Averaging Method (SAM) that incorporates the shock position into the numerical scheme. The shock position is numerically calculated by discretizing the theoretical speed of the front from the GPME theory. The speed satisfies the jump condition for integral conservation laws. SAM results in a non-oscillatory temporal profile, producing physically valid numerical results. We use SAM to demonstrate that the choice of averaging alone is not the cause of the oscillations, and that the shock position must be a part of the numerical scheme to avoid the artifacts.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

The purpose of the paper is to identify the cause of the numerical artifacts reported in the literature [1–3] for second order finite volume discretizations of the Generalized Porous Medium Equation (GPME) with discontinuous coefficients, and to suggest a numerical approach that does not have these problems. The GPME, commonly known as the Filtration Equation, can be expressed in both conservative and integral forms as:

* Corresponding author.

E-mail address: dcmaddix@stanford.edu (D.C. Maddix).

$$\begin{aligned}
p_t &= \nabla \cdot (k(p) \nabla p) \\
&= \Delta \Phi(p), \quad \text{where} \\
\Phi(p) &= \int_0^p k(\tilde{p}) d\tilde{p}, \quad k(p) = \Phi'(p).
\end{aligned} \tag{1.1}$$

In the discontinuous coefficients case, $k(p)$ is given as:

$$k(p) = \begin{cases} k_{\max}, & p \geq p^* \\ k_{\min}, & p^* > p, \end{cases} \quad \text{where} \tag{1.2}$$

k_{\max} , k_{\min} and p^* are real positive constants.

In this paper, we are interested in a subclass of the GPME known as the Stefan problem [4–7], for which

$$\Phi(p) = \begin{cases} c_1(p - c_3)_+, & \text{if } p \geq 0, \\ c_2 p, & \text{otherwise,} \end{cases} \tag{1.3}$$

for arbitrary $c_1, c_2, c_3 \in \mathbb{R}$. We look at a particular Stefan problem, where $p \geq 0$, $c_1 = k_{\max}$ and $c_3 = p^*$. Then, $\Phi(p)$ in Eqn. (1.1) can be expressed as the positive part function

$$\Phi(p) = k_{\max}(p - p^*)_+ = \begin{cases} k_{\max}(p - p^*), & \text{if } p \geq p^*, \\ 0, & \text{otherwise,} \end{cases} \tag{1.4}$$

and $k(p)$ is given by Eqn. (1.2) with $k_{\min} = 0$.

The Stefan problem can also be formulated in its classical form [8,9] as

$$\begin{aligned}
\frac{\partial p}{\partial t} &= k_{\max} \Delta p, \quad p \geq p^*, \\
\frac{\partial p}{\partial t} &= k_{\min} \Delta p, \quad p < p^*.
\end{aligned}$$

In the classical Stefan problem, the above two parabolic equations are defined on domains that are separated by a moving interface $x^*(t)$. The Stefan condition is given on the interface as

$$(p_L - p_R) \frac{dx^*(t)}{dt} = -k_{\max} \frac{\partial p_L}{\partial x} + k_{\min} \frac{\partial p_R}{\partial x},$$

where $p_L \equiv \lim_{x \rightarrow x^*(t)^-} p(x, t)$ and $p_R \equiv \lim_{x \rightarrow x^*(t)^+} p(x, t)$. The Stefan condition can be derived by using the Rankine–Hugoniot jump condition for this conservation law [10].

The Stefan Problem has been key to both the numerical and theoretical developments of the GPME. The Stefan Problem is used in modeling phase transitions, and was developed to study the evolution of a medium of two phases, water and ice [7]. Brattkus and Meiron [11] use the Stefan problem to model crystal growth. Sethian and Strain [12] and Chen et al. [9] develop a modified Stefan problem to model crystal growth as well as dendritic solidification. Eqn. (1.2) is often used to illustrate the numerical challenges present in more complex porous media applications. For example, in van der Meer et al. [2] a foam model prototype is developed with $k_{\max} = 1$, $k_{\min} = \epsilon \rightarrow 0$ and $p^* = 0.5$.

The continuous GPME already poses numerical challenges, caused by self-sharpening and degeneracy for near-zero $k(p)$ [1]. Due to this self-sharpening and degeneracy in the continuous case, it is not just the discontinuity that poses the numerical challenges in the Stefan problem in Eqn. (1.4). The discontinuity in $k(p)$ does make the challenges more severe and because of this discontinuity, the Modified Equation Analysis approach in [1] for the continuous GPME is not applicable. Here, we look at the discontinuous GPME, and also refer to subclasses of the continuous GPME. The Porous Medium Equation (PME) subclass, where $k(p) = p^m$ and $m \geq 1$, is used to model gas flow through a porous medium [1,7,13]. Another application in thermodynamics is the superslow diffusion equation [1,7], where $k(p) = \exp(-1/p)$. Further applications of the GPME for continuous $k(p)$ are detailed in [1].

1.1. Understanding the behavior of the GPME

The GPME in Eqn. (1.1) at first appears like a heat equation. Contrary to solutions of the heat equation, where the propagation speed is infinite, solutions of the GPME are known to have a finite speed of propagation [7]. This results from the degeneracy of the GPME for compactly supported initial data, and is a property that distinguishes the GPME from classical parabolic theory. This degeneracy leads to self-sharpening and moving interface solutions, as illustrated in Fig. 1. In addition, for certain compactly supported smooth initial data, the waiting time phenomenon can occur. This is discussed in the literature [14,15] [16, Chapter 3] and is illustrated in Fig. 3. The term waiting time refers to the fact that an interface moves only after sufficient sharpening.

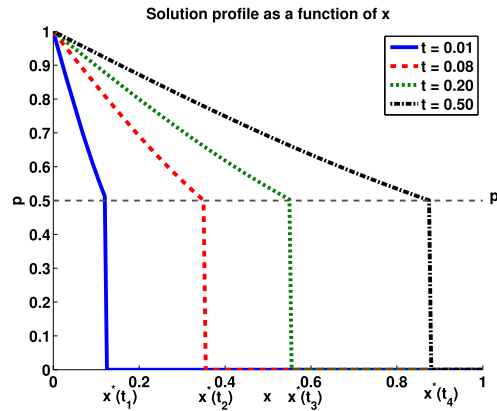


Fig. 1. The exact solution of Eqn. (1.1) with $k(p)$ given by Eqn. (1.2), where $k_{\max} = 1$ and $k_{\min} = 0$, evolves as a rightward moving shock over time. The moving shock position is given by $x^*(t)$ and the fixed p value at the shock is given by $p^* = 0.5$.

The theoretical propagation speed for compactly supported initial data is known and is given by Darcy's Law [7] as

$$V = - \lim_{x \rightarrow x^*(t)^-} \nabla v,$$

where

$$v = \int_0^p \frac{\Phi'(\tilde{p})}{\tilde{p}} d\tilde{p}, \quad (1.5)$$

and $x^*(t)$ is the shock position. Combining these expressions gives

$$V = - \lim_{x \rightarrow x^*(t)^-} \frac{\Phi'(p) \nabla p}{p} = - \lim_{x \rightarrow x^*(t)^-} \frac{k(p) \nabla p}{p}. \quad (1.6)$$

Eqn. (1.6) holds for any $k(p)$. In the particular case of the Stefan problem in Eqn. (1.4), $k(p) = k_{\max} = 1.0$ to the left of the shock. Substituting the $k(p)$ limit into Eqn. (1.6) gives

$$V = - \lim_{x \rightarrow x^*(t)^-} \frac{\nabla p}{p} = - \lim_{x \rightarrow x^*(t)^-} \nabla \log(p). \quad (1.7)$$

The expression for the velocity V can also be expressed in terms of fluxes. The flux F for the integral conservation law in Eqn. (1.1) is given by

$$F(p) = -k(p) \nabla p = -\nabla \Phi(p). \quad (1.8)$$

Then, from Eqn. (1.6),

$$V = \frac{F(p_L)}{p_L}, \quad (1.9)$$

where $p_L \equiv \lim_{x \rightarrow x^*(t)^-} p(x, t)$. For a compactly supported initial condition and $k_{\min} = 0$, the flux and p values to the right of the shock are zero. In this case, the velocity in Eqn. (1.9) can be expressed in terms of the familiar jump condition for integral conservation laws as

$$V = \frac{F(p_L) - F(p_R)}{p_L - p_R}, \quad (1.10)$$

where $p_R \equiv \lim_{x \rightarrow x^*(t)^+} p(x, t)$ [7,17]. The velocity in Eqn. (1.10) holds in general for any $k_{\min}, p_R \geq 0$, by the Rankine-Hugoniot condition [18].

1.2. Numerical methods to approach this problem

The degeneracy, self-sharpening and nonlinearity of the GPME pose interesting numerical challenges. The numerical approaches used by practitioners vary based on the field. In the porous media communities, central flux-based finite volume methods are widely employed to solve the related variable coefficients problem $p_t = \nabla \cdot (k(x) \nabla p)$. The coefficient $k(x)$ is

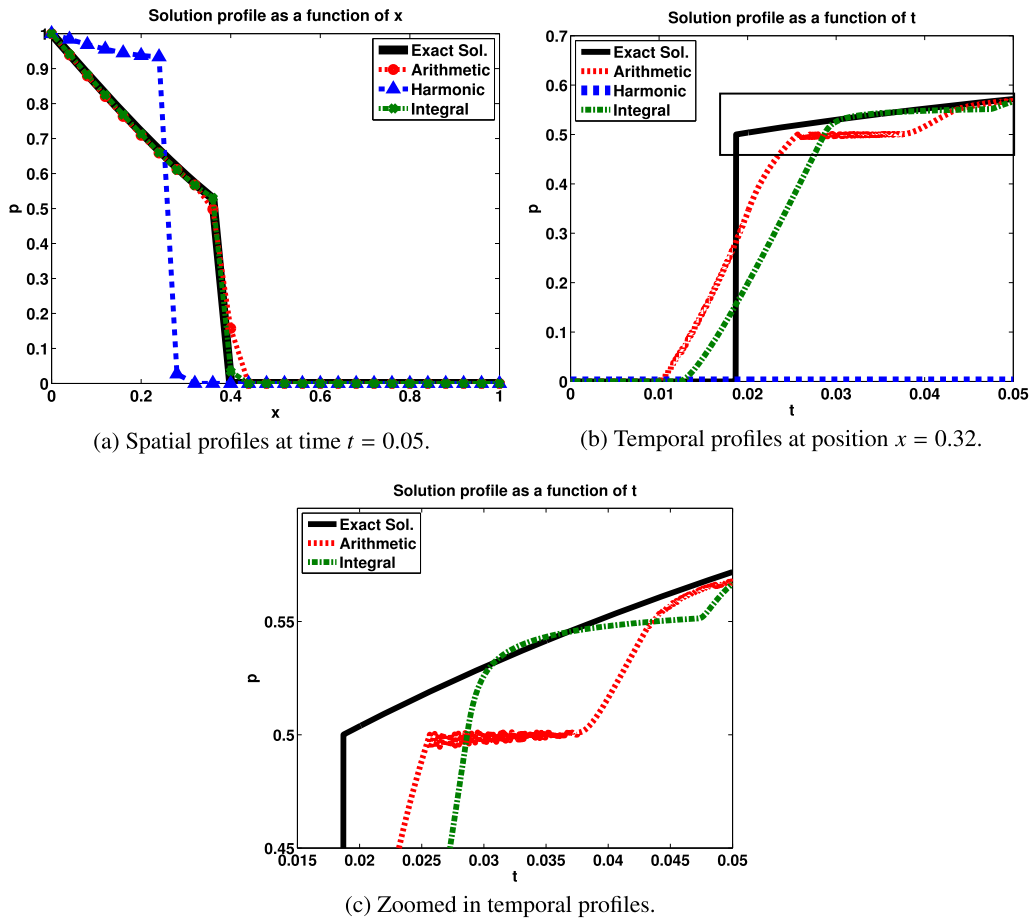


Fig. 2. Comparison of various averages, where $k_{\max} = 1$, $k_{\min} = 0$ and $p^* = 0.5$. The spatial step size $\Delta x = 0.04$ and the time step size $\Delta t = \Delta x^2/32$.

defined at the cell-centers, and harmonic and arithmetic averaging are commonly used to compute the coefficient, which may be discontinuous, at the cell interface. Harmonic averaging is often preferred in these problems because it leads to more physical solutions. This common finite volume averaged-based approach has also been extended to the nonlinear GPME in Eqn. (1.1). For our case, the selection of the $k(p)$ averaging is not as straightforward as in the variable coefficient case. After carefully comparing arithmetic to harmonic averaging, Lipnikov et al. [3] prefer arithmetic averaging over harmonic for the continuous PME with near-zero $k(p) = p^m$. An integral average is developed in van der Meer et al. [2] for the discontinuous GPME in Eqn. (1.2). With these average-based approaches, it is difficult to satisfy the jump condition in Eqn. (1.10). Violation of this condition manifests itself in locking and spurious temporal oscillations, as illustrated in Figs. 2 and 3 and further discussed in Section 5. In other fields, adaptive and moving mesh approaches have been applied to this problem also with increased refinement near the shock. In [13], for example, a finite element moving mesh method for the PME is developed. Adaptive Mesh Refinement (AMR) will be discussed further in Section 5.

Variational particle schemes have been developed for the PME in [19–21], where a variational principle is defined and steepest descent is applied to optimize the corresponding energy function. Variational principles have also been defined for the Stefan problem in [22] and a similar gradient flow method can be applied for the discontinuous GPME.

In the crystal application communities, numerical work has been done on solving the Stefan problem, as detailed in [23–26] and the references therein. In [27,28], an enthalpy scheme for Stefan problems is introduced. This method consists of a compact finite difference stencil with an implicit temporal scheme. The shock position is not incorporated into the scheme, and the error is observed to be concentrated at the moving interface. Brattkus and Meiron [11] show that finite difference and finite element methods that do not take special care at the interface result in a significant error of $\mathcal{O}(\sqrt{\Delta x})$. They propose an effective method based on an integral equation formulation that requires an additional integral evaluation.

In two-phase flow applications of the Stefan problem, level set methods [8] could also be useful in tracking the evolving interface. Thus far, they have not been directly applied to Eqn. (1.4). Related problems with Stefan boundary conditions are discussed in Sethian and Strain [12] and Chen et al. [9]. More recent approaches using the level set method in the solidification community are detailed in [29,30]. An advantage of the level set approach over traditional front-tracking

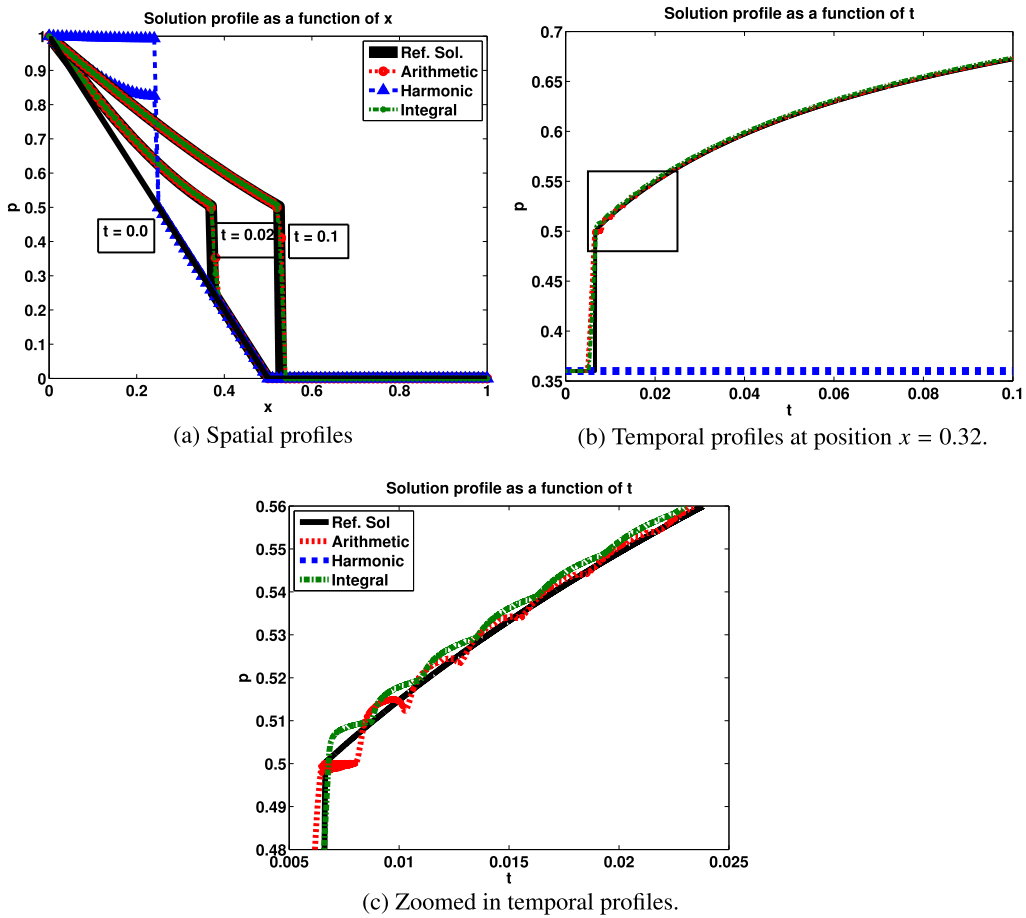


Fig. 3. Comparison of various averages, where $k_{\max} = 1$, $k_{\min} = 0$ and $p^* = 0.5$ for a waiting time phenomenon example. The piecewise linear initial condition self-sharpens until it is sharp enough for the support interface position to move rightward at later times past the initial interface position at $x = 0.5$. The spatial step size $\Delta x = 0.01$ and the time step size $\Delta t = \Delta x^2/32$.

approaches [31,32] is that the interface is treated implicitly. In doing so, the level set method can be more easily extended to higher dimensions, and is robust under complex topologies. Numerical methods combining the level set approach with the Extended Finite Element Method (X-FEM) have also been developed to solve these Stefan problems [33,34]. A modified Stefan problem is also solved in Zhao et al. [35] using a phase-field method. Phase-field methods converge as the interface thickness parameter tends to zero. The use of phase-field approaches in solidification and the effect of the interface thickness parameter have also been studied in [36,37]. While these papers in the solidification and phase change communities provide approaches to solve the Stefan problem, they do not give details on understanding why the numerical artifacts occur with finite volume average-based approaches. This leads us to the main goal of this paper.

1.3. Main goal of this paper

The main goal of this paper is to shed light on the origin of the numerical problems reported in the literature for finite volume averaged-based approaches of the discontinuous GPME. Before we discuss the cause of the artifacts, we first introduce an alternate numerical method for Eqn. (1.4) that does not exhibit the numerical errors observed in the literature. This newly developed Shock-Based Averaging Method (SAM) is discussed in Section 2. We discuss this first because it shows that incorporation of the shock position in the scheme is key, and also helps with the understanding of the artifacts in later sections. Section 3 discusses the derivation of SAM when the shock position is known. The shock position can also be approximated using the jump condition, which is detailed in Section 4. In Section 5, we cast SAM in the finite volume framework to help identify what has been lacking in the other approaches. The main issue is that these discretizations do not contain enough information about the shock.

2. Proposed numerical method: shock-based averaging method (SAM)

Due to the discontinuities in $k(p)$ and p , we adopt a finite volume approach to the integral form of the governing equation:

$$p_t = \nabla \cdot (k(p) \nabla p), \quad \text{where} \quad k(p) = \begin{cases} k_{\max}, & p \geq p^* \\ k_{\min}, & p^* > p, \end{cases} \quad (2.1)$$

where p^* is the left limiting value. We first define a finite volume grid with cell-centers x_j for $j = 1, \dots, N+1$. The boundaries of the domain are at x_1 and x_{N+1} , respectively and the corresponding unknowns p are fixed by Dirichlet boundary conditions. The remaining $N-1$ degrees of freedom p_j , and the corresponding coefficients k_j , are defined at the nodes x_j for $j = 2, \dots, N$. We define control volumes CV_j with width Δx_j . The cell faces $x_{j+1/2}$ of each CV_j are at a distance of $\Delta x_j/2$ from the cell-centers x_j . We assume that the solution is monotone and non-increasing, and that the shock is located between x_i and x_{i+1} , such that $p_i \geq p^* \geq p_{i+1}$.

The semi-discrete numerical discretization for CV_j with volume Δx_j is given by

$$\Delta x_j \frac{dp_j}{dt} = F_j^- - F_j^+, \quad (2.2)$$

where F_j^- represents the in-flux and F_j^+ represents the out-flux of CV_j . The numerical fluxes

$$F_j^+ = -k_{j+1/2} \frac{p_{j+1} - p_j}{\Delta x}, \quad j \neq i, \quad (2.3)$$

and

$$F_j^- = -k_{j-1/2} \frac{p_j - p_{j-1}}{\Delta x}, \quad j \neq i+1, \quad (2.4)$$

are defined at all faces away from the shock cell with the standard two-point flux approximation, where $\Delta x \equiv \Delta x_j$. The coefficient $k_{j+1/2}$ at the cell face represents a local average of its neighboring coefficients. Then, for any two-neighbor average away from the shock cell, the coefficient is constant and given by

$$k_{j+1/2} = \begin{cases} k_{\max}, & 1 \leq j \leq i-1, \\ k_{\min}, & i+1 \leq j \leq N. \end{cases} \quad (2.5)$$

Analogous definitions are used for $j-1/2$. Because of the discontinuity, Eqns. (2.3)–(2.4) cannot be used for the flux at the cell interface $x_{i+1/2}$.

2.1. Formulation of the fluxes near the shock

To estimate the fluxes out of CV_i and into CV_{i+1} , we borrow ideas from hyperbolic systems [38] and place a control volume around the discontinuity. Fig. 4 shows the auxiliary finite volume grid with the additional control volume CV_* around the physical shock position $x^*(t)$, where $p(x^*(t), t) \equiv p^*$ is defined. Fig. 4 also shows $\Delta x^*(t)$ as the distance between x_i and $x^*(t)$, where $0 \leq \Delta x^*(t) \leq \Delta x$. We then remove the cell face at position $x_i + \Delta x/2$ and add the cell faces of CV_* . The new cell faces are chosen to be centrally located at a distance of

$$\frac{|x^*(t) - x_i|}{2} \equiv \frac{\Delta x^*}{2},$$

from x_i and $x^*(t)$, and a distance of

$$\frac{|x^*(t) - x_{i+1}|}{2} \equiv \frac{\Delta x - \Delta x^*}{2},$$

from $x^*(t)$ and x_{i+1} . We have effectively added an additional degree of freedom p^* at $x^*(t)$. In Eqn. (2.1), p^* is known, and it does not need to be computed. We present the approach in its general form for future extensions.

With the additional control volume, we can now define the necessary out-flux F_i^+ of CV_i and in-flux F_{i+1}^- of CV_{i+1} . The coefficient at each of the new cell faces is known. By the monotonicity of p , for any $p(x, t)$ to the left of the interface, $k(p(x, t)) = k_{\max}$. Similarly, for any $p(x, t)$ to the right of the interface, $k(p(x, t)) = k_{\min}$. For this piecewise constant coefficient, any two-neighbor average on this auxiliary grid gives the same coefficient value at the faces. Based on these known coefficient values at the faces and the face-centered grid, the second-order fluxes are given by

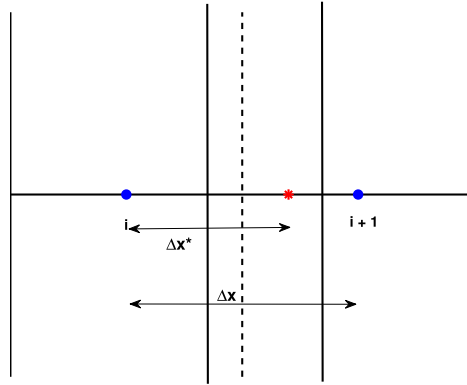


Fig. 4. Illustration of the additional control volume CV_* around the shock position $x^*(t)$. The grid points x_i and x_{i+1} are depicted by the blue dots, and $x^*(t)$ is depicted by the red star. The dashed line represents the deleted cell face between CV_i and CV_{i+1} . The new cell faces are located at a distance $\Delta x^*(t)/2$ from x_i and $[\Delta x - \Delta x^*(t)]/2$ from x_{i+1} . (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

$$F_i^+ = -k_{\max} \frac{p^* - p_i}{\Delta x^*}, \quad (2.6)$$

at the cell face between x_i and x^* and

$$F_{i+1}^- = -k_{\min} \frac{p_{i+1} - p^*}{\Delta x - \Delta x^*}, \quad (2.7)$$

at the cell face between x^* and x_{i+1} . The resulting scheme is conservative because F_i^+ is equal to the in-flux F_{*}^- of CV_* , by definition and similarly F_{i+1}^- is equal to the out-flux F_{*}^+ of CV_* . The fluxes are then substituted into the semi-discrete equation (2.2), where the cell volumes are now given by $\Delta x_i = (\Delta x + \Delta x^*)/2$ and $\Delta x_{i+1} = \Delta x - \Delta x^*/2$, respectively.

For $k_{\min} = 0$, we can relate the expression for F_i^+ in Eqn. (2.6) to the expression for $\Phi(p) = k_{\max}(p - p^*)_+$ in Eqn. (1.4) for the Stefan problem. The analytical flux $F(p_L)$ is given by $-\nabla \Phi(p_L)$ in Eqn. (1.8). The numerical flux F_i^+ can be interpreted as approximating this gradient with an upwind discretization as

$$-\frac{\Phi(p^*) - \Phi(p_i)}{\Delta x^*} = -k_{\max} \frac{p^* - p_i}{\Delta x^*} = F_i^+.$$

We see that the above formula is a first order approximation to the flux at $x^*(t)$. From the jump condition in Eqn. (1.9), the velocity can then be computed.

Round-off errors can arise when the shock location approaches the grid points x_i and x_{i+1} , and the denominators in Eqn. (2.6) and Eqn. (2.7) approach zero. To avoid problems, we specify a tolerance of ϵ that is proportional to Δx . If $\Delta x^* \leq \epsilon$, we neglect the leftmost portion near x_i , and set $k_{i+1/2} = k_{\min}$ for CV_i . Similarly, if $\Delta x - \Delta x^* \leq \epsilon$, we ignore the rightmost portion near x_{i+1} , and set $k_{i+1/2} = k_{\max}$ for CV_{i+1} . We then use these coefficients in the standard two-point flux approximation in Eqns. (2.3) and (2.4).

We will refer to the above approach as the Shock Based-Averaging Method (SAM).

3. SAM numerical results: exact shock location

In the numerical results presented throughout the paper, the test problem is given by Eqn. (2.1) with $k_{\max} = 1.0$, $k_{\min} = 0.0$ and $p^* = 0.5$. The choice of k_{\max} and p^* is arbitrary, while k_{\min} is set to zero to test the behavior of the numerical method in the degenerate case. The algorithm also works for arbitrarily small k_{\min} . The fixed Dirichlet boundary conditions are given by

$$\begin{aligned} p(0, t) &= 1.0, \quad \forall t \geq 0, \\ p(1, t) &= 0.0, \quad \forall t \geq 0. \end{aligned} \quad (3.1)$$

We first present results for the initial condition displayed in Fig. 5. This initial condition is formed as the exact solution to Eqn. (2.1) at a later time. The same parameters for k_{\max} and p^* , as given above for the test problem, are used to generate the initial condition, whereas $k_{\min} = 0.01$ for some smoothness at the bottom of the front. The coefficient value $k_{\min} = 0$ throughout the simulations.

For the representative test problem, an analytical solution exists, and is used to verify the implementation through detailed convergence studies. The derivation of the exact solution is generalized for arbitrary k_{\max} in Appendix A. In this

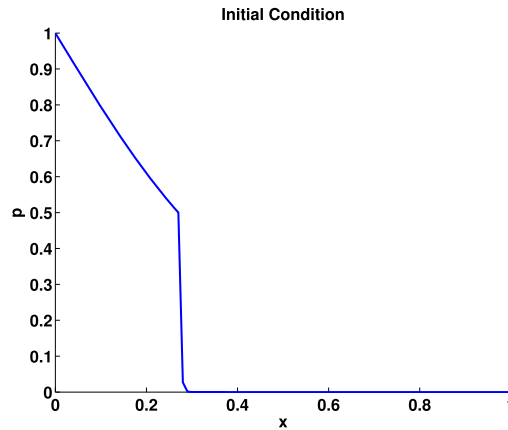


Fig. 5. Spatial profile at time $t = 0$.

section, we use the exact shock position to compute $\Delta x^*(t) \equiv |x^*(t) - x_i|$, where

$$x^*(t) = \alpha \sqrt{t}, \quad \alpha = 2\sqrt{k_{\max} z_1},$$

and z_1 is the solution to the nonlinear equation in Eqn. (A.4). In the following section, we will show how the shock location can be approximated, if it is not available.

After numerical stability and accuracy tests, the time step is selected to be $\Delta t = \Delta x^2/32$ for the explicit Forward Euler method. The discontinuity in the coefficient and the degeneracy make the time step criterion be more restrictive than it is for classical parabolic equations. If stability is the only interest, and not accuracy, the coefficient in the time step can be increased, but the time step must still be on the same order of $\mathcal{O}(\Delta x^2)$.

In Fig. 6, the numerical solution is plotted at the arbitrary position $x = 0.32$ over time. By self-similarity of the solution [1,7,13], the temporal plots have the same profile for any x -coordinate that the front has passed through. Fig. 6 illustrates that the numerical solution with SAM has an accurate and non-oscillatory temporal profile, even on a coarse grid. Fig. 7 reveals that the shock location is accurate in space for this moving front problem. The self-sharpening nature of the GPME [1,7,13] is depicted, by the smooth lower corner in the initial condition in Fig. 5 evolving into a sharp corner in Fig. 7. Fig. 7 also illustrates the sharp capture of the shock.

4. SAM numerical results: approximate shock location

The method in the prior section can be extended to more general GPME problems, where the exact shock position is not known. The finite speed of propagation property and theoretical speed of the front for the GPME, discussed in Section 1.1, can be utilized to numerically approximate the shock location. Again the fluxes in Eqns. (2.6) and (2.7) are used with the difference that $\Delta x^*(t)$ is numerically calculated. In this section, we show that this approximation does not introduce numerical artifacts, such as temporal oscillations. The convergence results are provided in Appendix B.

We discretize V in Eqn. (1.7) with upwinding for the derivative as

$$\hat{V} = -\frac{p_i - p_{i-1}}{\Delta x p_i}, \quad (4.1)$$

where i is the index, such that $p_i \geq p^* \geq p_{i+1}$. Eqn. (4.1) can be interpreted as an approximation of the jump condition in Eqn. (1.9), where p_L is approximated by p_i and $p_R = 0$. For problems where p_R is not initially zero, such as the waiting time problem in Section 1.2, we can also use a discrete approximation to Eqn. (1.10). Care must be taken numerically when approximating p_R in Eqn. (1.10), since p_{i+1} is not guaranteed to be zero: all that is guaranteed is that $0 \leq p_{i+1} \leq p^*$. In this case, we use p_{i+2} to approximate p_R .

To obtain the numerical shock position, a simple time integration is implemented. We let ξ^n represent the approximate $x^*(t)$ at time step $n \geq 0$ and substitute

$$\Delta x^*(t) \approx \xi^n - x_i, \quad (4.2)$$

into the expressions for the fluxes in Eqns. (2.6) and (2.7). We integrate the approximate shock speed in Eqn. (4.1) in time, using the following update

$$\xi^{n+1} = \xi^n + \Delta t \hat{V}. \quad (4.3)$$

We assume that the initial position ξ^0 is known from the problem definition or can also be approximated.

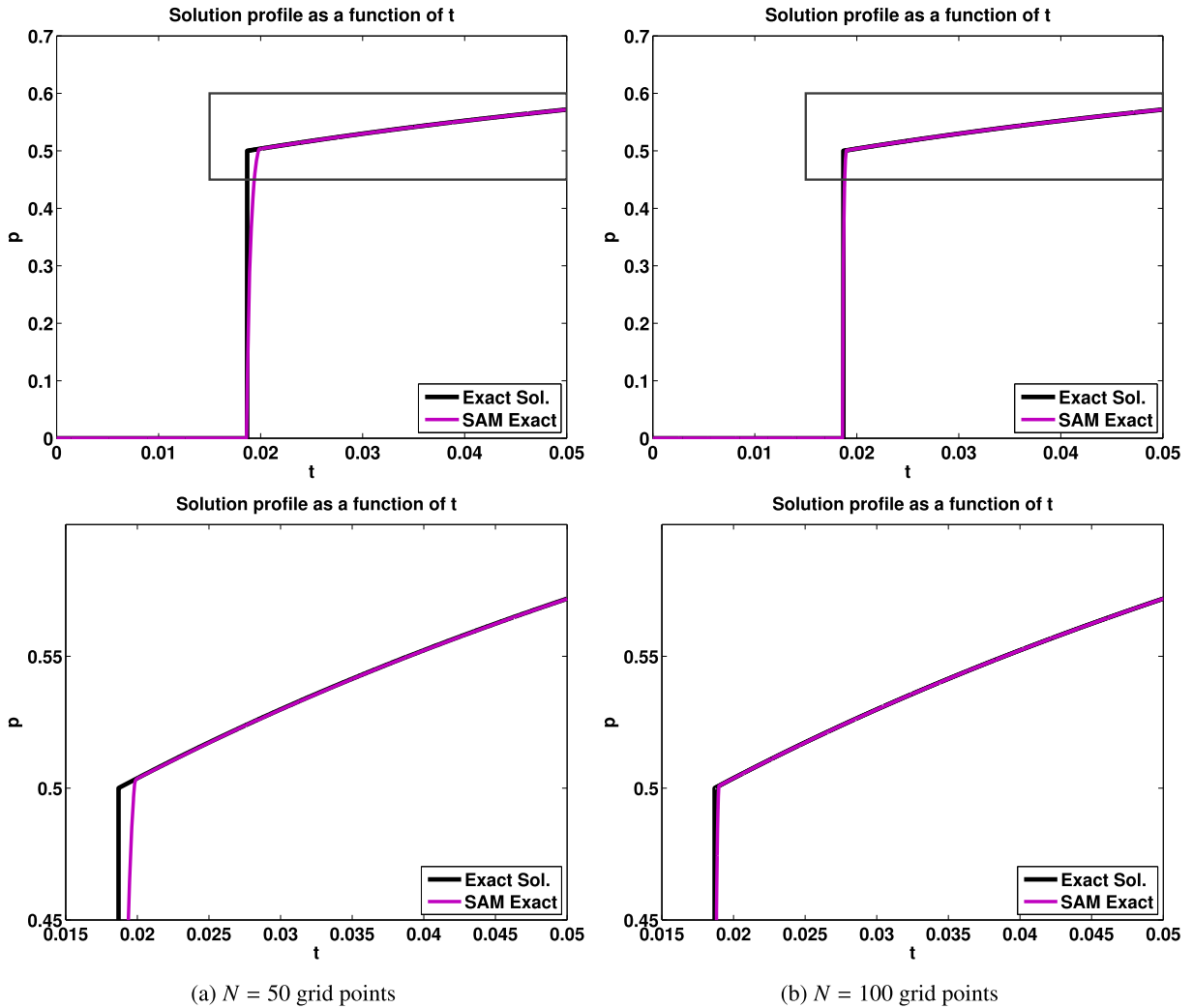


Fig. 6. Temporal profiles at position $x = 0.32$ with $\Delta t = \Delta x^2/32$.

We can also approximate $\Delta x^*(t)$ using a level set method [8]. In the level set method, the interface $x^*(t)$ is represented by the zero level set $\{x \mid \phi(x, t) = 0\}$ of a signed distance function $\phi(x, t)$. Using the velocity V of the interface in Eqn. (1.7), the evolution equation can be written in terms of $\phi(x, t)$ as

$$\phi_t + V \cdot \nabla \phi = 0. \quad (4.4)$$

The velocity can be extended to the entire domain or in a neighborhood of the interface by using velocity extension methods based on the Fast Marching Method [39]. The level set equation (4.4) can be solved numerically using Forward Euler in time and upwinding for the gradient, where V is discretized as \hat{V} in Eqn. (4.1). Since $\Delta x^*(t)$ is defined as the distance from the shock or interface $x^*(t)$ to the grid point x_i , $\Delta x^*(t)$ can be calculated by $|\phi(x_i)|$. Solving Eqn. (4.4) for ϕ_i^n and setting $\Delta x^*(t) \approx |\phi_i^n|$ gives the same results as solving Eqns. (4.2)–(4.3).

Figs. 8a and 8b illustrate the long-time behavior of the numerical shock position evolution. Fig. 8a displays that the numerical shock position aligns with the exact shock position as a function of time, and that the shock position evolution is accurately captured without oscillations. Both plots verify the numerical implementation, and show that there is no significant accuracy loss in estimating the shock position using Eqn. (4.3).

The results in Figs. 9 and 10 on the same test case from the prior section show that there is no significant change in the behavior of the SAM solution with the approximate shock speed than with the exact shock speed in the prior section.

Figs. 11a–11c show results for SAM for the waiting phenomenon problem discussed in Section 1.1. For this initial condition, the exact solution is not known. The figures show that the SAM solution is accurate and does not possess the numerical artifacts shown in Fig. 3.

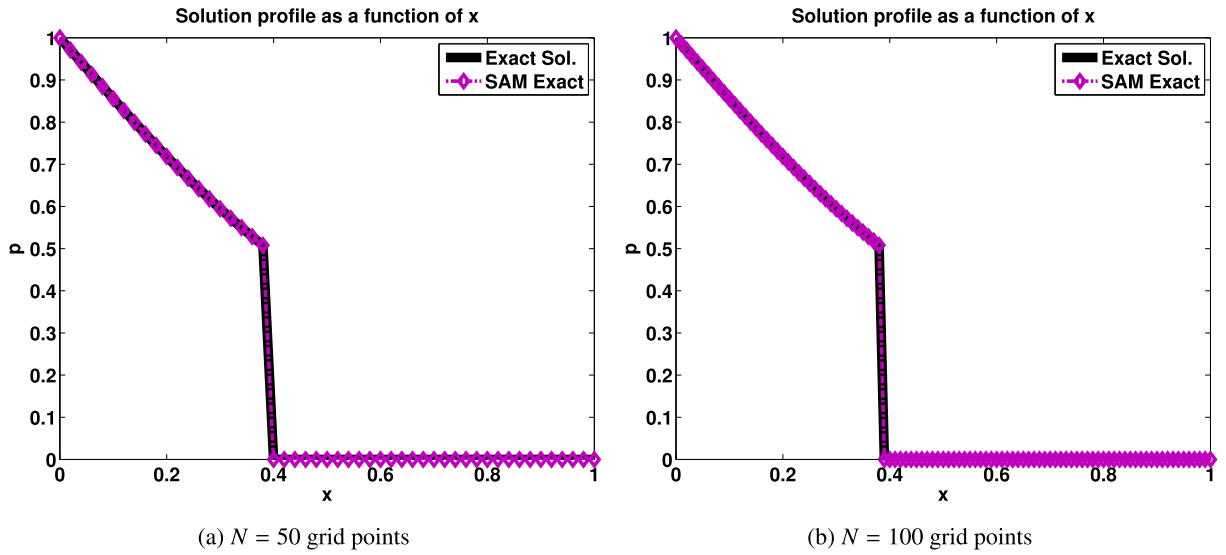


Fig. 7. Spatial profiles at time $t = 0.05$ with $\Delta t = \Delta x^2/32$.

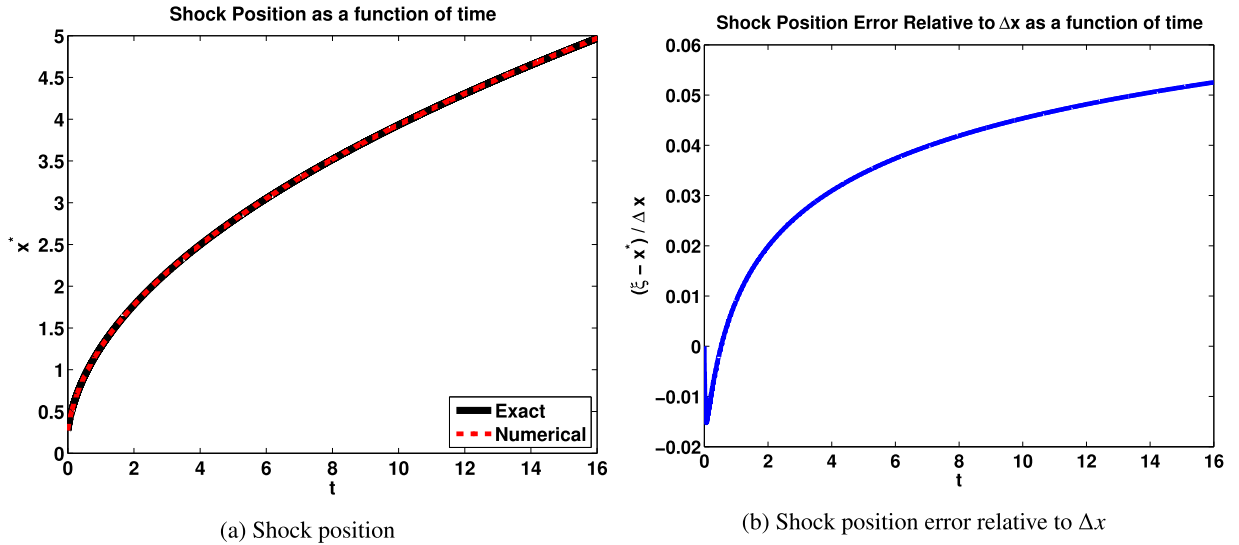


Fig. 8. Comparison of the temporal profiles of the shock position and relative error for $\Delta x = 0.04$ and $\Delta t = \Delta x^2/32$.

5. Average-based approaches

In the prior sections, we showed that we developed an accurate method, SAM, for the Stefan problem. The main goal of the paper is to understand why the artifacts are occurring with finite volume averaged-based methods from the literature. We will now recast SAM in this framework to shed light on what is happening in these methods that are used frequently in the porous media community.

We implement the finite volume average-based methods using the Forward in Time, Central in Space (FTCS) discretization on a uniform Cartesian grid, as is commonly done in the porous media literature. The numerical fluxes are given by

$$F_j^+ = -k_{j+1/2} \frac{p_{j+1} - p_j}{\Delta x}, \quad (5.1)$$

and

$$F_j^- = -k_{j-1/2} \frac{p_j - p_{j-1}}{\Delta x}, \quad (5.2)$$

for an arbitrary average $k_{j+1/2}$ of the neighboring coefficients at the cell face $x_{j+1/2}$ for all $j = 1, \dots, N$. For verification, we successfully repeated the test case presented for the arithmetic and integral average in van der Meer et al. [2].

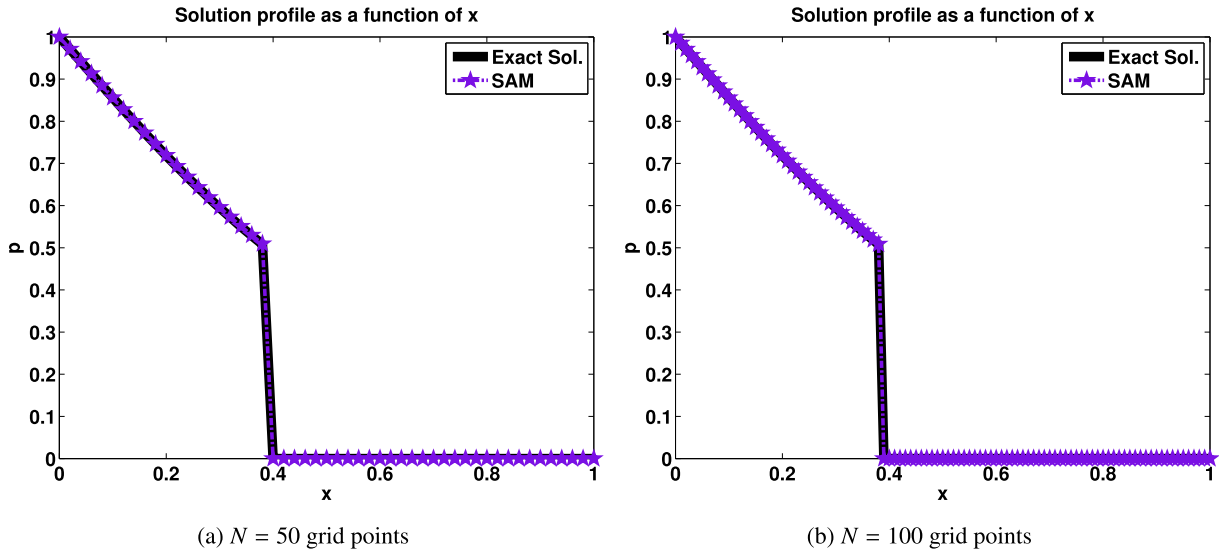


Fig. 9. Spatial profiles at time $t = 0.05$ with $\Delta t = \Delta x^2/32$.

5.1. Arithmetic and harmonic averages

In the FTCS finite volume scheme with arithmetic

$$k_{j+1/2}^A = \frac{k_j + k_{j+1}}{2}, \quad (5.3)$$

and harmonic

$$k_{j+1/2}^H = \frac{2k_j k_{j+1}}{k_j + k_{j+1}}, \quad (5.4)$$

averaging, the fluxes in and out of the control volumes surrounding the shock are discretized using Eqns. (5.1)–(5.2). The corresponding arithmetic fluxes are given by

$$F_i^{+A} = F_{i+1}^{-A} = -k_{i+1/2}^A \frac{p_{i+1} - p_i}{\Delta x}. \quad (5.5)$$

The analogous expression holds for the harmonic fluxes F_i^{+H} and F_{i+1}^{-H} , where the arithmetic average $k_{i+1/2}^A$ in Eqn. (5.5) is replaced with the harmonic average $k_{i+1/2}^H$.

The temporal plots at $x = 0.32$ in Fig. 12 reveal that the numerical interface with harmonic averaging is locked, that is, does not move at all, and so does not advance to this position. Fig. 13 shows the spatial evolution and we again see the locking with harmonic averaging. This numerical solution in Fig. 13 evolves to a step function with values at 1 and 0. Eqn. (5.4) explains this locking numerical artifact, since $k_{j+1/2}^H = 0$ when either coefficient is zero. Although here locking is seen as a drawback, there are situations, where it can be desirable, such as in variable coefficient problems, $p_t = \nabla \cdot (k(x) \nabla p)$, when the interface separates a permeable and impermeable material. The arithmetic average is known to cause leakage across such interfaces. We could relax $k_{\min} = 0$ to $k_{\min} = \epsilon$ for some small $\epsilon > 0$. Fig. 14a shows that for $\epsilon = 0.01$, the numerical solution with harmonic averaging no longer completely locks, but is still lagging behind the true front location. From Eqn. (5.4), we see that $k_{i+1/2}^H$ favors the smaller coefficient and is approximately equal to 2ϵ , as the shock moves through the interval $[x_i, x_{i+1}]$. The constant and small averaged k value at the cell face results in a solution whose numerical speed is too slow. Grid dependent oscillations with much larger amplitudes than those in the solution with arithmetic averaging are present in its temporal profile in Figs. 14b–14c. In [2], the harmonic average is not considered, since it is known in the literature to behave poorly for the GPME with near-zero coefficients. In the rest of this subsection, we focus on the behavior of the solution with arithmetic averaging.

The numerical solutions depicted in Figs. 12 and 13 show results for the arithmetic average. We see that the numerical solution with arithmetic averaging does not sharply capture the shock. Fig. 12 shows spurious oscillations of low and high frequency. The low frequency oscillations are grid dependent. The high frequency oscillations are apparent in the zoomed in region around the solution shock value of $p^* = 0.5$. Fig. 13 shows that the solution with arithmetic averaging does not have spatial oscillations. In the temporal view, oscillations are present.

There are two main differences in the artifacts for the continuous and discontinuous GPME. Here, for the discontinuous case, there are temporal oscillations of low and high frequencies for the arithmetic average. For the continuous GPME,

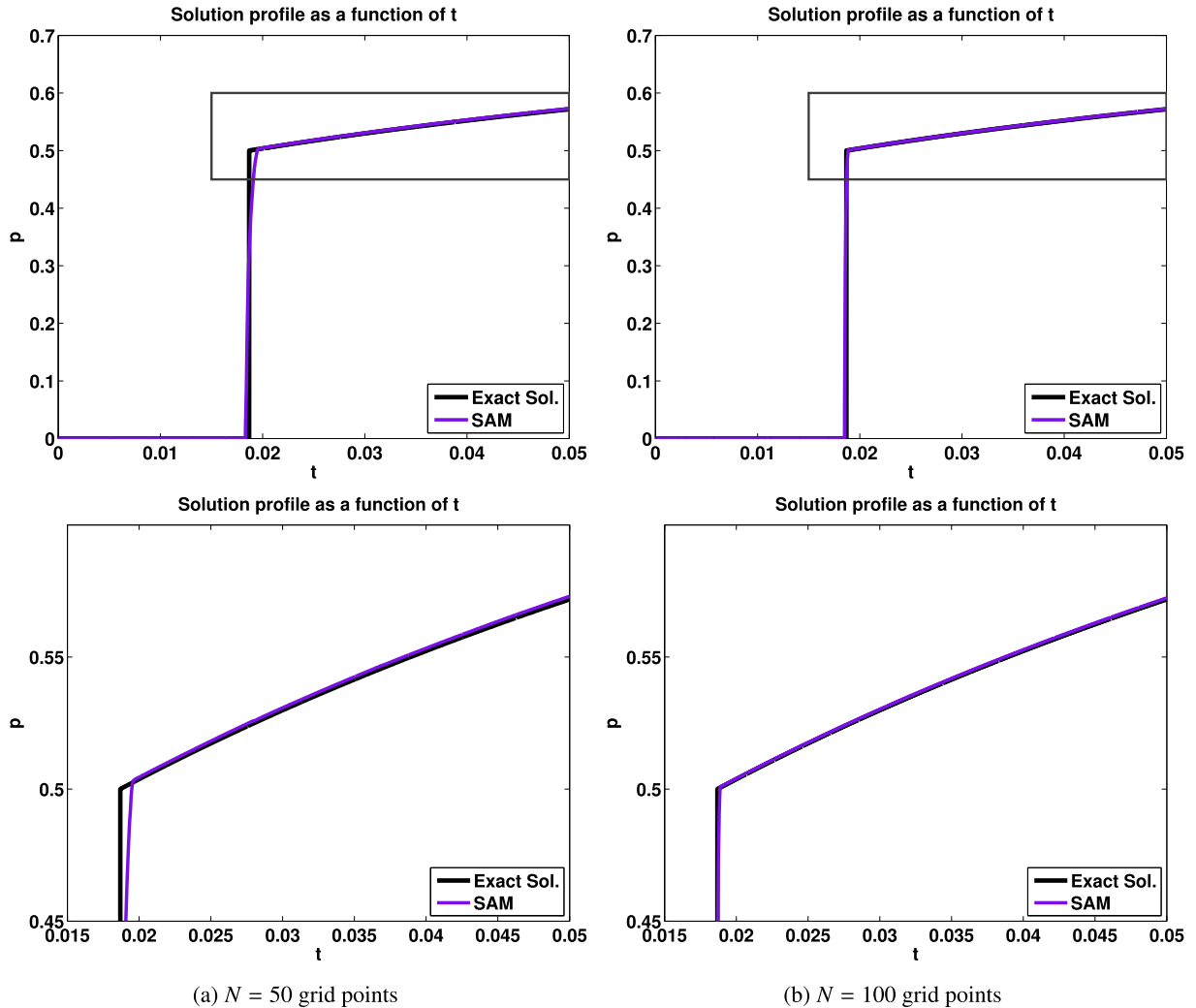


Fig. 10. Temporal profiles at position $x = 0.32$ with $\Delta t = \Delta x^2/32$.

temporal oscillations were only observed in solutions with harmonic averaging [1]. The lower frequency grid dependent oscillations were analyzed in [1] for discretizations with harmonic averaging. The high frequency oscillations are an additional numerical artifact with arithmetic averaging that were not observed for the harmonic average in [1].

The low frequency temporal oscillations have been studied in the literature [1,2,40,41]. It has been shown in these works that the temporal oscillations occur as the front crosses a grid cell. The spurious oscillations then do not vanish with grid refinement. The frequency increases and the amplitude decreases, as the number of grid points increases. The differences in frequency and amplitude are illustrated by the solutions for $N = 50$ in Fig. 12a and $N = 100$ in Fig. 12b. Another observed characteristic of these oscillations in [2,40,41] is that the amplitude decreases as shock moves further away from the solution probe point. This is because the possible error in the shock position relative to the distance from the probe point to the shock decreases.

The high frequency oscillations were also observed in [2]. Figs. 15a–15b illustrate the temporal profile on a shorter time interval, where the high frequency oscillations occur for p values near $p^* = 0.5$. We see that p_{i+1} is slowly increasing, until it crosses the threshold at $p^* = 0.5$. The corresponding k_{i+1} then jumps from $k_{\min} = 0$ to $k_{\max} = 1$, according to the model for $k(p)$ in Eqn. (1.2). The arithmetic average $k_{i+3/2}^A$ in Eqn. (5.3) jumps from 0 to 0.5, as illustrated in Figs. 15a–15b. The increase in $k_{i+3/2}^A$ causes p to drop below $p^* = 0.5$ and $k_{i+3/2}^A$ to jump back down to 0 at the next time step. The cycle then repeats itself.

We use the FTCS discretization to explain the effect of the jump in $k_{i+3/2}^A$ on the p value at the next time step. We discretize the semi-discrete equation (2.2) with Forward Euler in time and substitute in the arithmetic fluxes to obtain

$$p_{i+1}^{n+1} = p_{i+1}^n + \frac{\Delta t}{\Delta x} \left(-k_{i+1/2}^A \left[\frac{p_{i+1}^n - p_i^n}{\Delta x} \right] + k_{i+3/2}^A \left[\frac{p_{i+2}^n - p_{i+1}^n}{\Delta x} \right] \right), \quad (5.6)$$

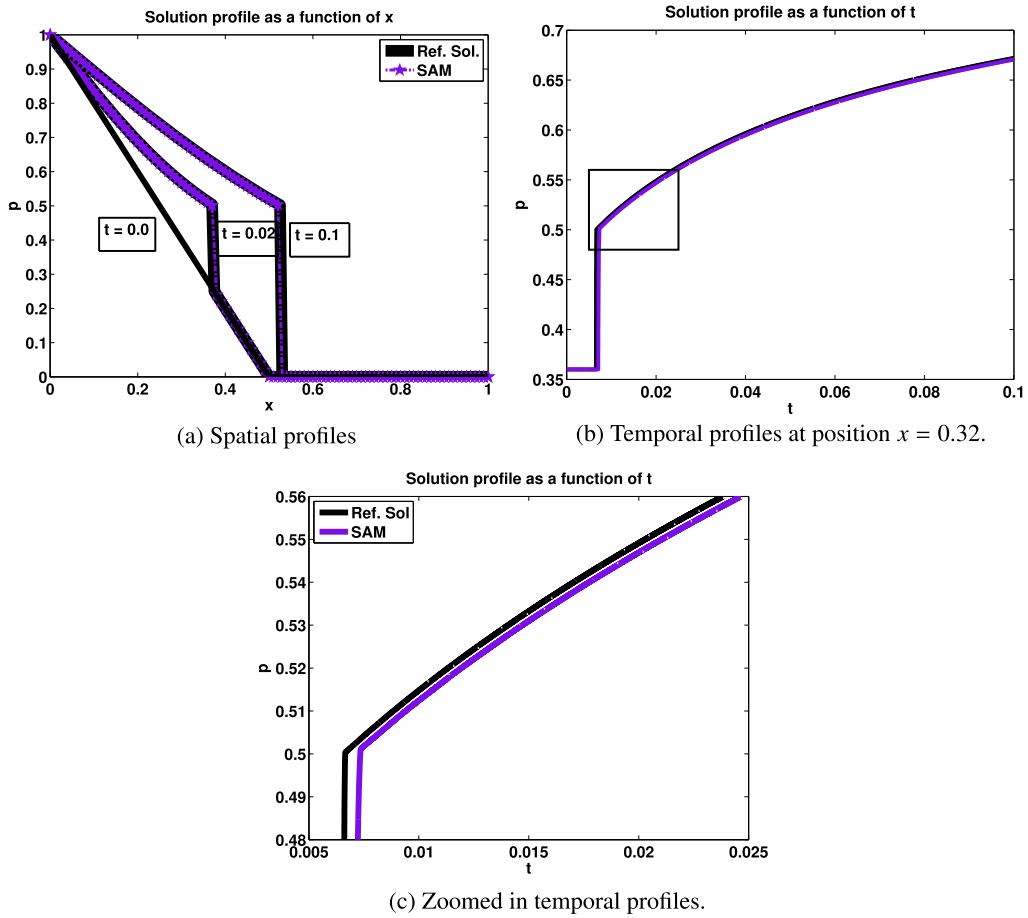


Fig. 11. SAM numerical solution with $N = 100$ grid points and $\Delta t = \Delta x^2/32$ for the waiting time phenomenon example from Fig. 3.

where n is a time step before the solution drop and $p_{i+1}^n \geq 0.5$. The first term in Eqn. (5.6) is small in magnitude, since $p_i^n \approx 0.5$. The second term in Eqn. (5.6) has a larger gradient of p , since $p_{i+2}^n \approx 0$. It is also negative, and so the increase in $k_{i+3/2}^A$ results in the drop in p_{i+1}^{n+1} . At the following time steps, $k_{i+3/2}^A = 0$, until the solution crosses the threshold again.

As p_{i+1} is increasing and is below the threshold, $k_{i+1/2}^A$ is fixed at 0.5, regardless of the shock position in the cell. From Eqn. (5.6), we see that the positive quantity $\Delta t/(2\Delta x^2)[p_i^n - p_{i+1}^n]$ is added to the current p value at each time step. The constant arithmetic average at the interface $x_{i+1/2}$ allows the solution to artificially increase above $p^* = 0.5$, resulting in the high frequency oscillations.

Figs. 15a–15b also show that the high frequency oscillations are dependent on the time step size. As Δt is decreased by half, the number of oscillations in Fig. 15b doubles from those in Fig. 15a. The amplitude of the high frequency oscillations also decreases, as Δt decreases. The amplitude decrease is expected from Eqn. (5.6), since the additional term is proportional to Δt . In [2], it is observed that as the time step is decreased, the arithmetic average solution in the high frequency region reaches a constant state at $p^* = 0.5$, rather than converging to the true solution.

Figs. 15a–15b display the profile of p_{i+1} over a time window when the shock is in the interval $[x_i, x_{i+1}]$. In Figs. 16a–16b, we now look at the profile of p_{i+2} over the same time window to see the effect of the jump in $k_{i+3/2}^A$ on the neighboring cell. Figs. 16a–16b show that the profile of p_{i+2} is an increasing piecewise constant. Since p_{i+2} and p_{i+3} are both less than p^* , $k_{i+5/2}^A = 0$ and the solution update for p_{i+2}^{n+1} reduces to

$$p_{i+2}^{n+1} = p_{i+2}^n + k_{i+3/2}^A \left(p_{i+1}^n - p_{i+2}^n \right) \frac{\Delta t}{\Delta x^2}. \quad (5.7)$$

Eqn. (5.7) shows that at the times when $p_{i+1}^n < p^* = 0.5$, $k_{i+3/2}^A = 0$, and so $p_{i+2}^{n+1} = p_{i+2}^n$. Otherwise, at the times when $k_{i+3/2}^A$ jumps to 0.5, the solution increases proportional to Δt . As expected, the increase in out-flux causes the solution to decrease in the left cell (Figs. 15a–15b) and the same increase in in-flux causes the solution in the right cell to increase (Figs. 16a–16b).

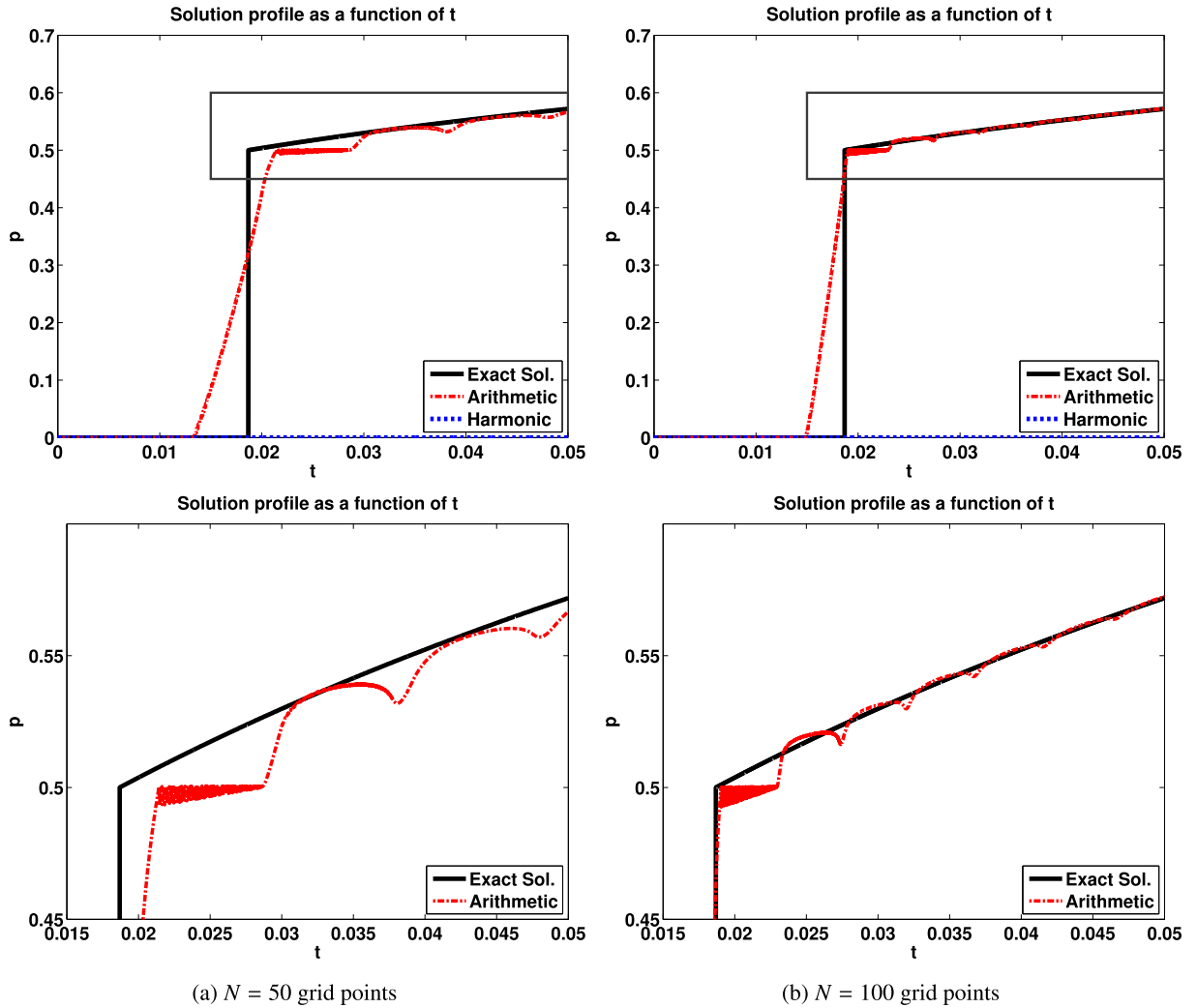


Fig. 12. Temporal profiles at position $x = 0.32$ with $\Delta t = \Delta x^2/32$.

5.2. Integral average

Van der Meer et al. [2] developed the integral average

$$k_{j+1/2}^I = \frac{\int_{p_j}^{p_{j+1}} k(\tilde{p}) d\tilde{p}}{p_{j+1} - p_j}, \quad (5.8)$$

which is effective in reducing the numerical artifacts. The integral average is derived by expressing the coefficient $k(p) = \Phi_p$ in Eqn. (1.1) to obtain

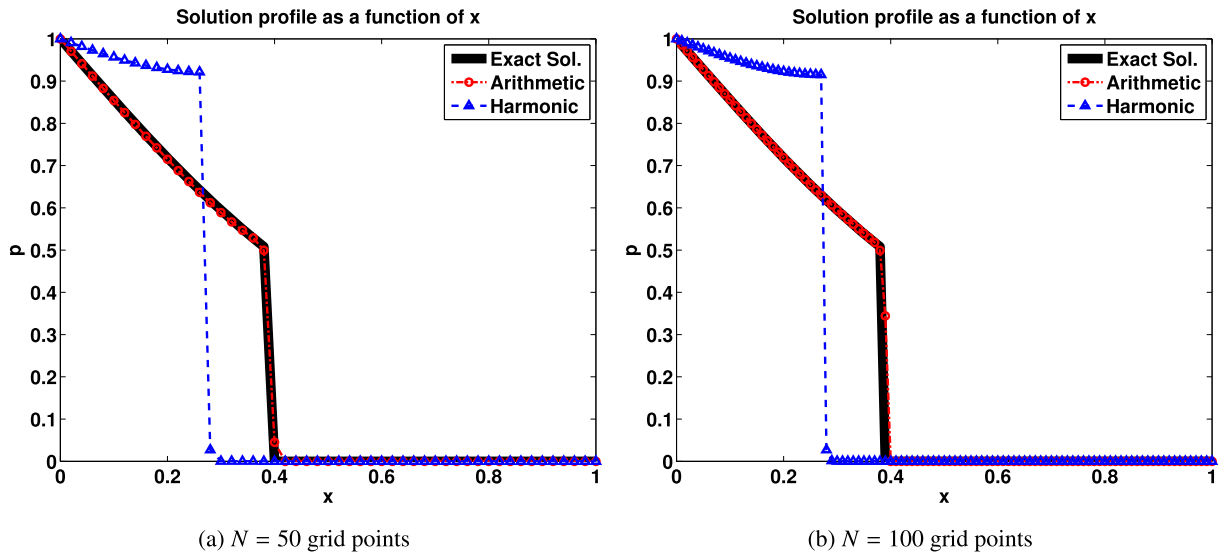
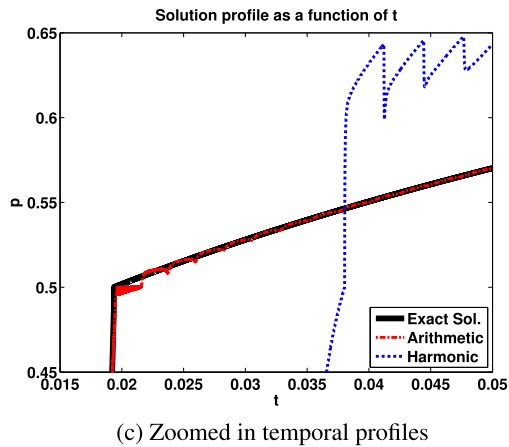
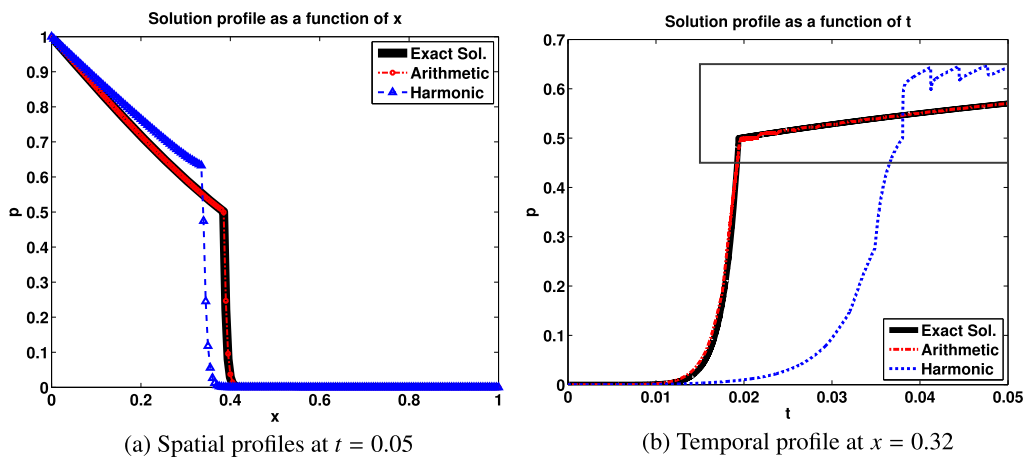
$$p_t = \nabla \cdot (\Phi_p \nabla p).$$

Discretizing Φ_p directly with central differences at the cell face $x_{j+1/2}$ gives

$$k_{j+1/2}^I = \frac{\Phi(p_{j+1}) - \Phi(p_j)}{p_{j+1} - p_j}. \quad (5.9)$$

Using the definition of $\Phi(p) = \int_0^p k(\tilde{p}) d\tilde{p}$, Eqn. (5.9) simplifies to Eqn. (5.8).

The numerical solution with the integral average is provided in Figs. 17 and 18. The improvement with the integral average is clear. It does not introduce high frequency oscillations near p^* . Although the low frequency, grid-dependent oscillations remain, the amplitude of these oscillations is smaller than those with arithmetic averaging. The numerical

Fig. 13. Spatial profiles at time $t = 0.05$ with $\Delta t = \Delta x^2/32$.Fig. 14. Solution profiles with $N = 200$ grid point and $\Delta t = \Delta x^2/32$.

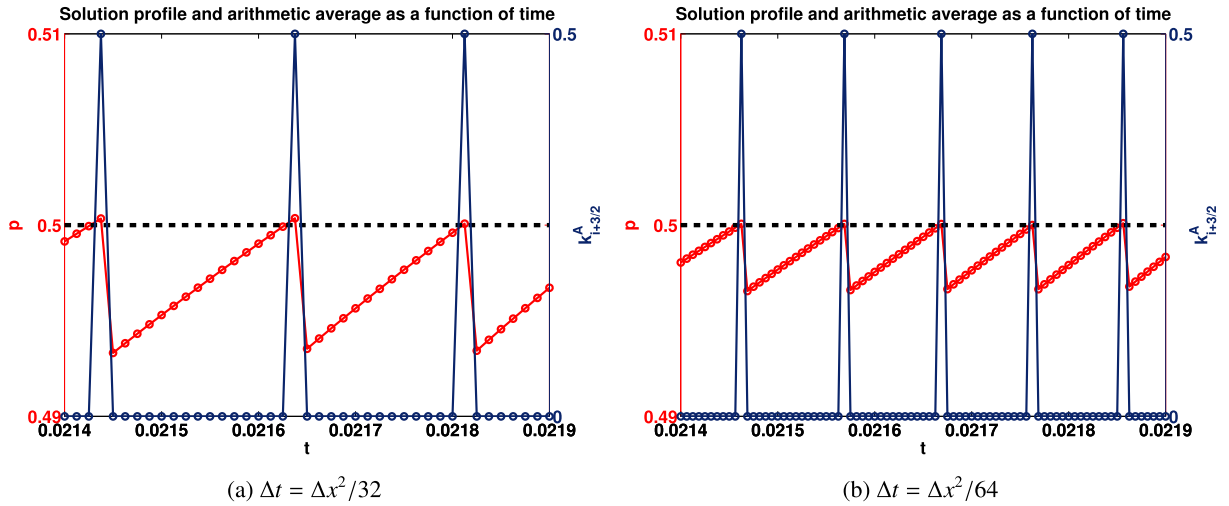


Fig. 15. Zoomed in region of the high frequency temporal oscillations near $p^* = 0.5$ in the left cell p_{i+1} for the arithmetic average $k_{i+3/2}^A$ with $N = 50$ grid points at position $x = 0.32$.

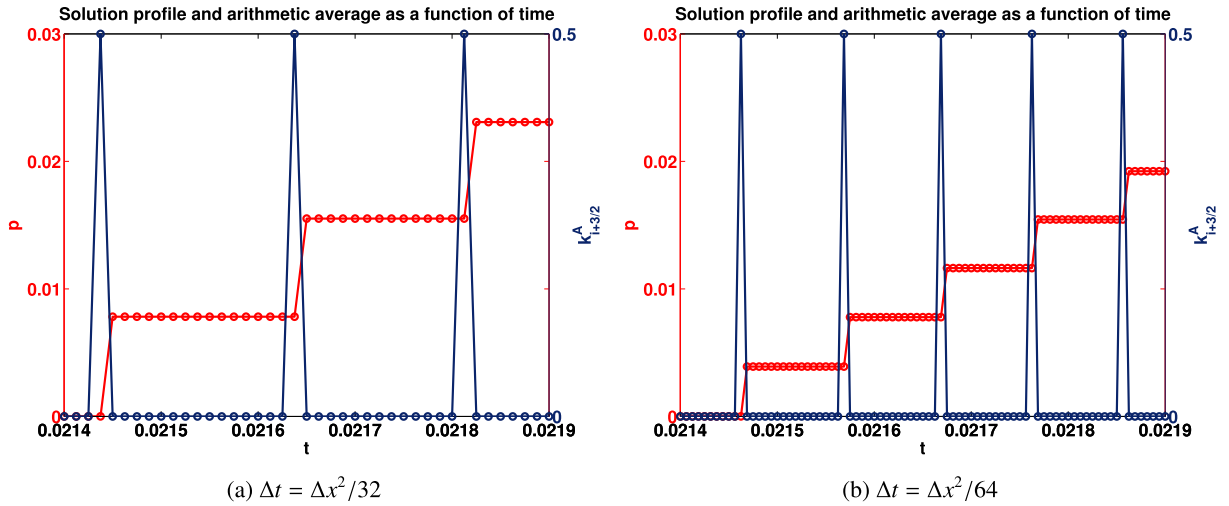


Fig. 16. Zoomed in region of the solution in the right cell p_{i+2} for the arithmetic average $k_{i+3/2}^A$ with $N = 50$ grid points at position $x = 0.34$.

solution is now monotonically increasing in time [2], matching the behavior of the true solution. The diffusive shock profile with the integral average is also illustrated in the spatial results in Fig. 18.

To explain why the integral average outperforms the arithmetic and harmonic averages, we write it in an alternate form than presented in [2]. The integral average $k_{i+1/2}^I$ in Eqn. (5.8) can be broken up at p^* into two separate integrals to obtain

$$k_{i+1/2}^I = \frac{k_{\min}(p_{i+1} - p^*) + k_{\max}(p^* - p_i)}{p_{i+1} - p_i}, \quad (5.10)$$

in the shock interval. It then contains some information about the shock, as encoded in the bounds of the integral. Unlike the harmonic and arithmetic averages, the integral average monotonically increases as the shock advances through the interval $[x_i, x_{i+1}]$.

The corresponding continuous flux is given by

$$F_i^{+I} = F_{i+1}^{-I} = -\frac{k_{\min}(p_{i+1} - p^*) + k_{\max}(p^* - p_i)}{\Delta x}. \quad (5.11)$$

Using this description, we can see that by utilizing p^* , the integral flux avoids computing the undefined gradient of p across the jump, as done in the schemes with arithmetic and harmonic averaging. It is this incorporation of p^* into the scheme that prevents the high frequency oscillations seen with arithmetic averaging, which explains the improved behavior.

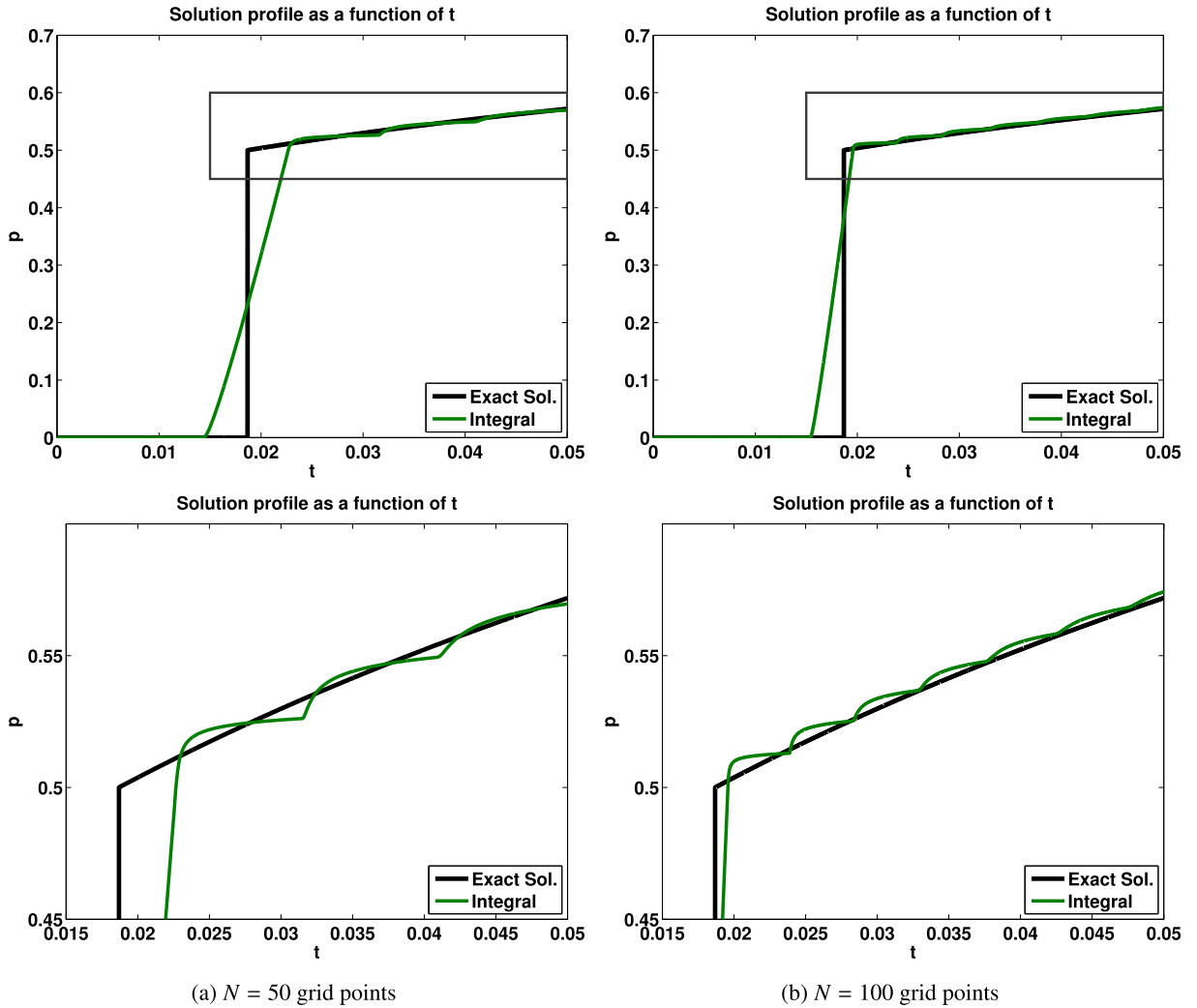


Fig. 17. Temporal profiles at position $x = 0.32$ with $\Delta t = \Delta x^2/32$.

As seen, the integral average does not remove the low frequency oscillations. This is because although p^* is incorporated in the flux in Eqn. (5.11), $x^*(t)$ is not. In the case, where $k_{\min} = 0$, $F_i^{+I} = F_{i+1}^{-I}$ in Eqn. (5.11) appears to be similar to F_i^+ in Eqn. (2.6). The difference occurs in the denominator, where the relative shock position $\Delta x^*(t)$ detected in SAM is replaced with Δx in the integral flux. The integral flux is then assuming that the distance to the shock is fixed of size Δx .

To see what happens with the temporal oscillations under grid refinement, we also implement the scheme with integral averaging and Adaptive Mesh Refinement (AMR) [42]. We define a coarse mesh away from the shock and a fine mesh near the shock. The fine mesh travels with the shock as it moves. In Fig. 19, a coarse mesh size of $\Delta x = 1/100$ is utilized. Within the coarse mesh cells surrounding the shock, a fine inner mesh size of $\Delta x_{\text{inner}} = 1/10$ is defined. Fig. 19 displays that oscillations are present with a smaller period and damped amplitude. The zoomed in results in Fig. 19 illustrate that there are $N_{\text{inner}} \equiv 1/\Delta x_{\text{inner}}$ oscillations in between the coarse grid cells, as expected as the front crosses the $N_{\text{inner}} = 10$ inner grid cells for $\Delta x_{\text{inner}} = 1/10$. The temporal oscillations are spatially dependent, and applying AMR does not remove them.

5.3. SAM in finite volume average form and its connection to the integral average

Writing SAM as an averaged-based method on a uniform Cartesian grid of spatial step size Δx provides additional insight. In particular, we will see another explanation of the temporal oscillations. For $j \neq i$, we have defined $k_{j+1/2}^{\text{SAM}}$ by Eqn. (2.5) in Section 2. The Cartesian and auxiliary grids are the same in the parabolic regions away from the shock. Using Eqns. (2.6)–(2.7), we can write the out-flux F_i^+ of CV_i and the in-flux F_{i+1}^- of CV_{i+1} for SAM in finite volume average form as

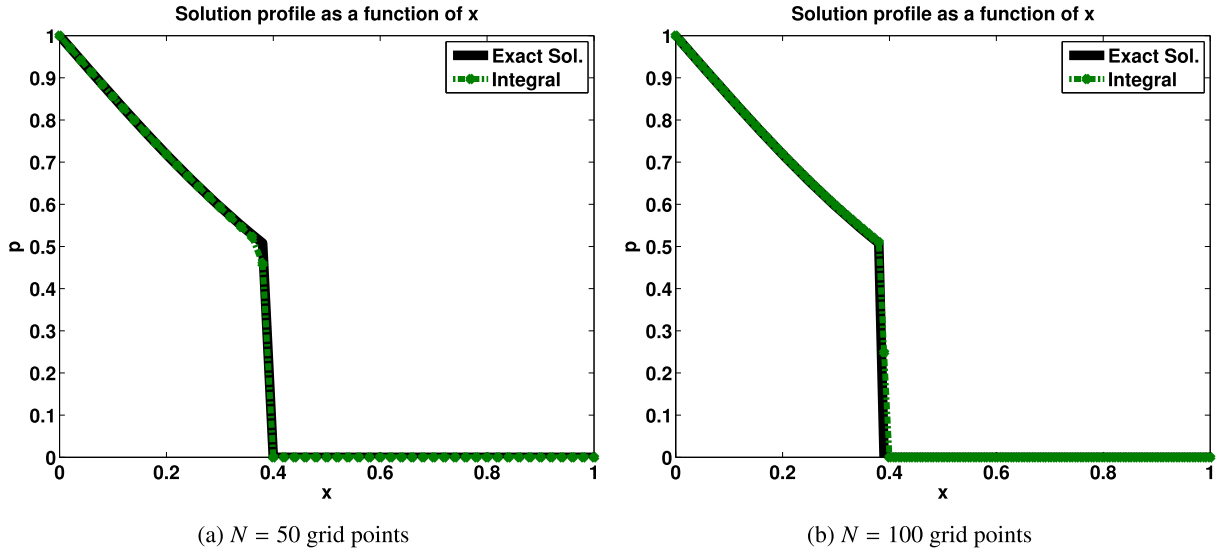


Fig. 18. Spatial profiles at time $t = 0.05$ with $\Delta t = \Delta x^2/32$.

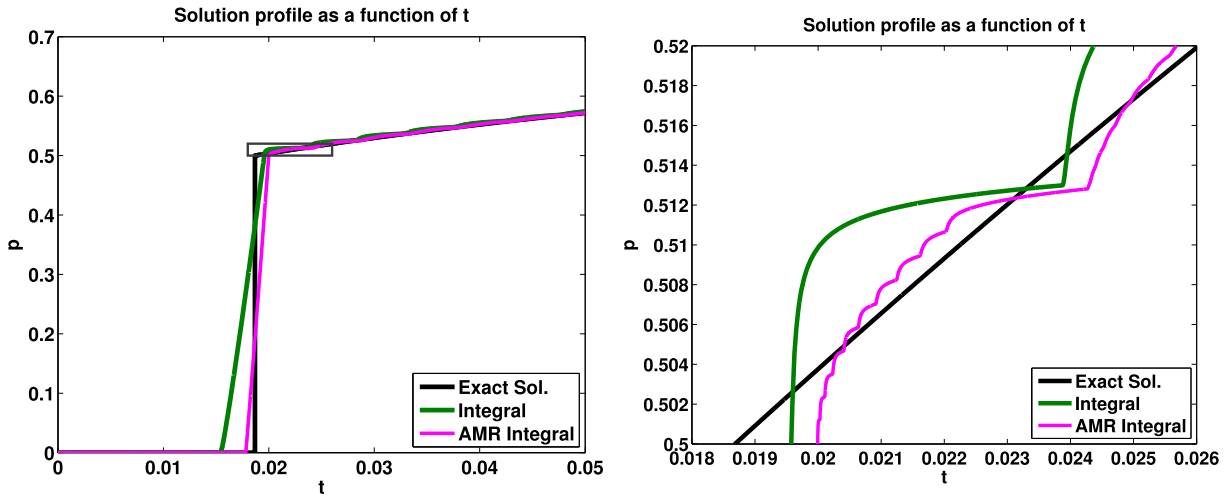


Fig. 19. Temporal profiles at position $x = 0.32$ with $\Delta t = \Delta x^2/32$ and $N = 100$ coarse grid points and 10 inner grid points.

$$F_i^+ = - \underbrace{\left(k_{\max} \frac{\Delta x}{\Delta x^*} \frac{p^* - p_i}{p_{i+1} - p_i} \right)}_{k_{i+1/2}^{SAM+}} \frac{p_{i+1} - p_i}{\Delta x}, \quad (5.12)$$

and

$$F_{i+1}^- = - \underbrace{\left(k_{\min} \frac{\Delta x}{\Delta x - \Delta x^*} \frac{p_{i+1} - p^*}{p_{i+1} - p_i} \right)}_{k_{i+1/2}^{SAM-}} \frac{p_{i+1} - p_i}{\Delta x}, \quad (5.13)$$

respectively. As opposed to the other averaging approaches, now $F_i^+ \neq F_{i+1}^-$: conservation is honored on the auxiliary grid, defined in Section 2.1.

We can express these SAM averages as a weighted linear combination

$$c_1 \frac{k_{\min}(p_{i+1} - p^*)}{p_{i+1} - p_i} + c_2 \frac{k_{\max}(p^* - p_i)}{p_{i+1} - p_i},$$

of the integral average $k_{i+1/2}^I$ in Eqn. (5.10). For $c_1 = 0$ and $c_2 = \Delta x / \Delta x^*$, $k_{i+1/2}^{SAM+}$ in Eqn. (5.12) is recovered, while for $c_1 = \Delta x / (\Delta x - \Delta x^*)$ and $c_2 = 0$, $k_{i+1/2}^{SAM-}$ in Eqn. (5.13) is recovered. Conversely, we can also express the integral flux in terms of the SAM fluxes surrounding the shock cell. Rewriting Eqn. (5.11), we have

$$\begin{aligned} F_i^{+I} &= - \left[k_{\min} \frac{p_{i+1} - p^*}{\Delta x - \Delta x^*} \frac{\Delta x - \Delta x^*}{\Delta x} + k_{\max} \frac{p^* - p_i}{\Delta x^*} \frac{\Delta x^*}{\Delta x} \right] \\ &= F_{i+1}^-(1 - y) + F_i^+ y, \end{aligned} \quad (5.14)$$

where $y \equiv \Delta x^* / \Delta x$. The shock moves from left to right in the interval $[x_i, x_{i+1}]$, and $0 < y < 1$.

As discussed in Section 2.1, the SAM flux is a first order approximation of the analytical flux, and the jump condition is satisfied. The integral flux, on the other hand, does not converge to the analytical flux as $\Delta x \rightarrow 0$, and does not satisfy the jump condition. Without loss of generality, we let the shock be to the right of the cell face $x_{i+1/2}$ by $\epsilon \Delta x$ for some small $\epsilon > 0$. Then $y = 1/2 + \epsilon$. Using Taylor series expansion in the parabolic region to the left of the shock gives

$$F_i^+ = F^L[p(x_{i+1/2+\epsilon})] + \mathcal{O}(\Delta x), \quad (5.15)$$

where $F^L[p(x)] = -k_{\max} \nabla p(x)$. Similarly, Taylor series expansion in the parabolic region to the right of the shock gives

$$F_{i+1}^- = F^R[p(x_{i+1/2+\epsilon})] + \mathcal{O}(\Delta x), \quad (5.16)$$

where $F^R[p(x)] = -k_{\min} \nabla p(x)$. Substituting Eqns. (5.15)–(5.16) into Eqn. (5.14) gives the following

$$F_i^{+I} = F^R[p(x_{i+1/2+\epsilon})](0.5 - \epsilon) + F^L[p(x_{i+1/2+\epsilon})](0.5 + \epsilon) + \mathcal{O}(\Delta x).$$

Taking the limit as $\epsilon \rightarrow 0$, we obtain

$$F_i^{+I} = \frac{F^R[p(x_{i+1/2})] + F^L[p(x_{i+1/2})]}{2} + \mathcal{O}(\Delta x),$$

as the integral flux at the interface $x_{i+1/2}$. Since $x_{i+1/2}$ is to the left of the shock, the analytical flux at $x_{i+1/2}$ is given by the left flux $F^L[p(x_{i+1/2})]$. The integral flux is then averaging the left and right fluxes across the jump. The right flux $F^R[p(x_{i+1/2})]$ should not have any impact on the value across the jump. This jump condition violation is exactly what results in the temporal oscillations. SAM does satisfy the jump condition and as the results in Sections 3 and 4 show, it does remove the artifacts entirely.

6. Conclusions and future work

This paper explains what causes the mysterious artifacts in discretizations of the discontinuous Generalized Porous Medium Equation (GPME), and suggests an alternative method that results in numerical solutions without these artifacts. The FTCS scheme with integral averaging performs better than the schemes with harmonic and arithmetic averaging because it has information about the shock value p^* . By rewriting the integral average in the shock cell, it can be seen that full removal of the numerical artifacts requires more than p^* , and that the shock location $x^*(t)$ must also be included in the numerical scheme. The Shock-Based Averaging Method (SAM) incorporates both $x^*(t)$ and p^* , and satisfies the jump condition for integral conservation laws to result in numerical solutions with accurate and smooth temporal profiles. Casting SAM in the finite-volume framework helps provide understanding on why the numerical artifacts were occurring with finite volume averaged-based methods.

Future work includes extensions to higher dimensions and to the Porous Medium Equation (PME) subclass. Since the velocity for the GPME is given for arbitrary dimensions, it can be used as input to a level set implementation for extensions to higher dimensions, as discussed in Section 4. Another future direction would be to combine the approach from this paper with our previous paper [1], discussing the PME subclass of the GPME with continuous coefficients. For the PME, there is no known specified p value at the shock, as given by p^* in this paper. With the additional control volume around the shock in the auxiliary finite volume grid from Section 2, $p^*(t)$ can be solved for as an additional degree of freedom, and an approach similar to SAM can be taken.

Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-114747. We would also like to thank Giuseppe Domenico Cervelli for his guidance.

Appendix A. Exact solution derivation for arbitrary parameters

For the test problem with the initial condition in Fig. 5, the exact solution is known and is used for testing the numerical methods. We give the full derivation, which is similar to [43], and is generalized for arbitrary k_{\max} . The domain is partitioned into $\Omega_1 = (0, x^*(t))$ and $\Omega_2 = [x^*(t), \infty)$. The problem can be subdivided into two constant coefficient heat equations [43,44]. The solution is monotonically non-increasing and so in Ω_1 , $p_1(x, t) > p^*$ and in Ω_2 , $p_2(x, t) \leq p^*$. It can be verified that $p_1(x, t) = 1 - c_1 \Phi(x/(2\sqrt{k_{\max}t}))$ and $p_2(x, t) = c_2 (1 - \Phi(x/(2\sqrt{k_{\min}t})))$ for some constants c_1, c_2 to be determined. $\Phi(x) = \text{erf}(z) = \int_0^z \phi(y)dy$ is the standard Gaussian error function, where $\phi(y) = \frac{2}{\sqrt{\pi}} \exp(-y^2)$.

The unknown shock location ($x^*(t)$) needs to be computed. To do so, two additional boundary conditions are required at the shock $x = x^*(t)$:

1. $p_1(x^*(t), t) = p_2(x^*(t), t) = p^*$, $\forall t$ (Continuity)
2. $k_{\max}(\partial p_1(x^*(t), t)/\partial x) = k_{\min}(\partial p_2(x^*(t), t)/\partial x)$, $\forall t$ (Flux Continuity).

Since Condition (1) must hold for all t , $x^*(t) = \alpha\sqrt{t}$,

$$c_1 = \frac{1 - p^*}{\Phi(\alpha/(2\sqrt{k_{\max}}))}, \quad c_2 = \frac{p^*}{1 - \Phi(\alpha/(2\sqrt{k_{\min}}))}.$$

Condition (2) is used to derive a nonlinear solve for the remaining unknown, α . Substituting in the expressions for the derivatives and simplifying leads to

$$c_1 \sqrt{k_{\max}} \phi\left(\frac{\alpha}{2\sqrt{k_{\max}}}\right) = c_2 \sqrt{k_{\min}} \phi\left(\frac{\alpha}{2\sqrt{k_{\min}}}\right). \quad (\text{A.1})$$

Let $z_1 = \alpha/(2\sqrt{k_{\max}})$ and $z_2 = \alpha/(2\sqrt{k_{\min}})$. We substitute the expressions for c_1 and c_2 into Eqn. (A.1), multiply both sides by $\frac{\alpha}{2}$ and simplify to obtain

$$\begin{aligned} \frac{1 - p^*}{\Phi(z_1)} \sqrt{k_{\max}} \phi(z_1) &= \frac{p^*}{1 - \Phi(z_2)} \sqrt{k_{\min}} \phi(z_2) \iff \\ (1 - p^*)(1 - \Phi(z_2)) \exp(z_2^2) \frac{\alpha}{2\sqrt{k_{\min}}} &= p^* \frac{\alpha}{2\sqrt{k_{\max}}} \Phi(z_1) \exp(z_1^2) \iff \\ (1 - p^*)(1 - \Phi(z_2)) \exp(z_2^2) z_2 &= p^* \Phi(z_1) z_1 \exp(z_1^2). \end{aligned} \quad (\text{A.2})$$

By a series expansion from integration by parts,

$$1 - \Phi(z_2) = \frac{\exp(-z_2^2)}{z_2 \sqrt{\pi}} \left(1 - \frac{1}{2z_2^2} + \frac{3}{4z_2^4} - \dots \right).$$

Using the above expression, Eqn. (A.2) simplifies to

$$\frac{1 - p^*}{\sqrt{\pi}} \left(1 - \frac{1}{2z_2^2} + \frac{3}{4z_2^4} - \dots \right) = p^* \Phi(z_1) z_1 \exp(z_1^2). \quad (\text{A.3})$$

In our application and most of the numerical tests, $k_{\min} = 0$. This implies that $z_2 \rightarrow \infty$. The limit as $z_2 \rightarrow \infty$ of the left hand side of Eqn. (A.3) is simply $(1 - p^*)/\sqrt{\pi}$. This is an advantage of computing the series form, since we can easily take this limit as $k_{\min} \rightarrow 0$. This form is also preferred numerically when k_{\min} is small to avoid multiplication of the large $\exp(z_2^2) z_2$ terms. For the $k_{\min} = 0$ case, the unknown (α) only appears in z_1 on the right hand side of the equation, given by

$$\frac{1 - p^*}{\sqrt{\pi}} = p^* \Phi(z_1) z_1 \exp(z_1^2). \quad (\text{A.4})$$

A simple one-dimensional nonlinear equation solver can be used to solve this equation for z_1 , where $\alpha = 2\sqrt{k_{\max} z_1}$.

Appendix B. Error tables and convergence study

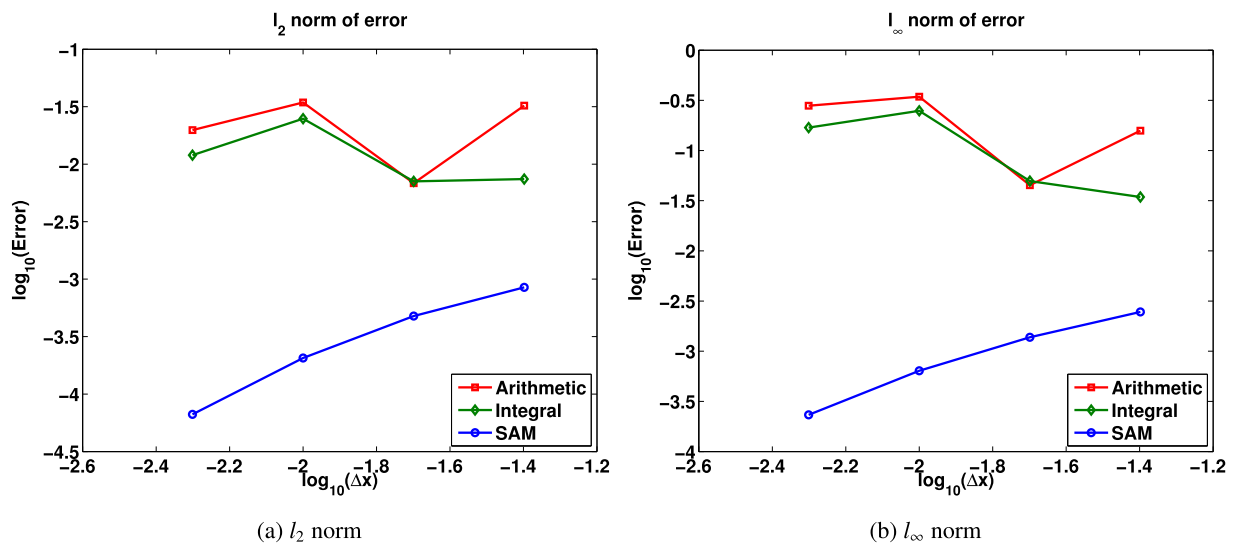
This appendix contains the convergence results for the model problem tested with the same time step size $\Delta t = \Delta x^2/32$ as in the prior sections. The l_2 and l_∞ error norms are calculated with respect to the exact solution on the corresponding grid. The error is measured globally, and includes the points surrounding the shock. We compare the errors of the FTCS with arithmetic and integral averaging to those of SAM. The errors of the FTCS with harmonic averaging are not shown, since the solution locks and does not converge. For the methods with arithmetic and integral averaging, even on finer grids, the asymptotic region is not yet reached. Tables B.1–B.2 show that SAM has approximately first order convergence and errors

Table B.1
 l_2 norm errors.

	$N = 25$	$N = 50$	$N = 100$	$N = 200$	Order
Arithmetic	3.2335e-02	6.8351e-03	3.4431e-02	1.9773e-02	N/A
Integral	7.4250e-03	7.0866e-03	2.4898e-02	1.1974e-02	N/A
SAM	8.4813e-04	4.7762e-04	2.0583e-04	6.6594e-05	1.2229

Table B.2
 l_∞ norm errors.

	$N = 25$	$N = 50$	$N = 100$	$N = 200$	Order
Arithmetic	4.8237e-02	1.5223e-02	3.5742e-02	2.0394e-02	N/A
Integral	1.4104e-02	9.1485e-03	2.9971e-02	1.6252e-02	N/A
SAM	2.1859e-03	1.7763e-03	1.0586e-03	4.5134e-04	1.1336

**Fig. B.20.** Loglog convergence plots for the Stefan problem.

that are orders of magnitude lower than those for the averaged-based methods. The first order convergence of SAM is also shown in Figs. B.20a–B.20b. We do not consider higher order methods because of the dominance of the shock.

References

- [1] D. Maddix, L. Sampaio, M. Gerritsen, Numerical artifacts in the generalized porous medium equation: why harmonic averaging itself is not to blame, *J. Comput. Phys.* 361 (2018) 280–298.
- [2] J. van der Meer, J. Kraaijevanger, M. Möller, J. Jansen, Temporal oscillations in the simulation of foam enhanced oil recovery, in: *ECMOR XV – 15th European Conference on the Mathematics of Oil Recovery*, 2016, pp. 1–20.
- [3] K. Lipnikov, G. Manzini, D. Moulton, M. Shashkov, The mimetic finite difference method for elliptic and parabolic problems with a staggered discretization of diffusion coefficient, *J. Comput. Phys.* 305 (2016) 111–126.
- [4] J. Stefan, Über die Theorie der Eisbildung inbesondere über die Eisbindung im Polarmeere, *Ann. Phys. U. Chem.* 42 (1891) 269–286.
- [5] L. Rubinstein, *The Stefan Problem*, 17 edn., Translations of Math. Monographs, AMS, Providence, Rhode Island, 1971.
- [6] H. Meirmanov, *The Stefan Problem* (Translated from the Russian), 3 edn., De Gruyter Expositions in Mathematics, Walter de Gruyter & Co., Berlin, 1992.
- [7] J. Vázquez, *The Porous Medium Equation, Mathematical Theory*, The Clarendon Press, Oxford University Press, Oxford, 2007.
- [8] J. Sethian, S. Osher, Fronts propagating with curvature-dependent speed: algorithms based on Hamilton–Jacobi formulations, *J. Comput. Phys.* 79 (1988) 12–49.
- [9] S. Chen, B. Merriman, S. Osher, P. Smereka, A simple level set method for solving Stefan problems, *J. Comput. Phys.* 135 (1997) 8–29.
- [10] T. Myers, F. Font, On the one-phase reduction of the Stefan problem with a variable change temperature, *Int. Commun. Heat Mass Transf.* 61 (2015) 37–41.
- [11] K. Brattkus, D. Meiron, Numerical simulations of unsteady crystal growth, *SIAM J. Appl. Math.* 52 (5) (1992) 1303–1320.
- [12] J. Sethian, J. Strain, Crystal growth and dendritic solidification, *J. Comput. Phys.* 98 (1992) 231–253.
- [13] C. Ngo, W. Huang, A study on moving mesh finite element solution of the porous medium equation, *J. Comput. Phys.* 331 (2017) 357–380, arXiv: 1605.03570.
- [14] J. Fischer, G. Günther, Finite speed of propagation and waiting times for the stochastic porous medium equation: a unifying approach, *SIAM J. Math. Anal.* 47 (1) (2015) 825–854.
- [15] S. Angenent, Analyticity of the interface of the porous media equation after the waiting time, *Proc. Am. Math. Soc.* 102 (2) (1988) 329–336.

- [16] S. Antontsev, S. Shmarev, *Evolution PDEs with Nonstandard Growth Conditions: Existence, Uniqueness, Localization, Blow-up*, Atlantis Press, Amsterdam, Paris, 2015.
- [17] J. Vazquez, Numerical simulations of unsteady crystal growth, *SIAM J. Appl. Math.* 52 (5) (1992) 1303–1320.
- [18] P. Lax, Hyperbolic systems of conservation laws II, *Commun. Pure Appl. Math.* X (1957) 537–566.
- [19] M. Westdickenberg, J. Wilkening, Variational particle schemes for the porous medium equation and for the system of isentropic Euler equations, *ESAIM: Math. Model. Numer. Anal.* 44 (2010) 133–166.
- [20] F. Otto, The geometry of dissipative evolution equations: the porous medium equation, *Commun. Partial Differ. Equ.* 26 (2001) 101–174.
- [21] C. Villani, *Topics in Optimal Transportation*, 58 edn., Graduate Studies in Mathematics, American Mathematical Society, Providence, USA, 2003.
- [22] R. Rossi, G. Savaré, Existence and approximation results for gradient flows, *Rend. Mat. Accad. Lincei* 15 (2004) 183–196.
- [23] A. Berger, H. Brezis, J. Rogers, A numerical method for solving the problem $u_t - \Delta f(u) = 0$, *RAIRO. Anal. Numér.* 13 (4) (1979) 297–312.
- [24] A. Kharab, A numerical solution of a free-boundary problem on the unsaturated flow of liquids in porous media, *Comput. Math. Appl.* 12A (12) (1986) 1193–1200.
- [25] Rizwan-uddin, An approximate-solution-based numerical scheme for Stefan problem with time-dependent boundary conditions, *Numer. Heat Transf., Part B* 33 (3) (1998) 269–285.
- [26] J. Caldwell, S. Savović, Y. Kwan, Nodal integral and finite difference solution of one-dimensional Stefan problem, *J. Heat Transf.* 125 (2003) 523–527.
- [27] M. Rose, An implicit enthalpy scheme for one-phase Stefan problems, *J. Sci. Comput.* 5 (2) (1990) 169–185.
- [28] M. Rose, An enthalpy scheme for Stefan problems in several dimensions, *Appl. Numer. Math.* 12 (1993) 229–238.
- [29] Y. Kim, N. Goldenfeld, J. Dantzig, Computation of dendritic microstructures using a level set method, *Phys. Rev. E* 62 (2) (2000) 2471–2474.
- [30] L. Tan, N. Zabaras, A level set simulation of dendritic solidification of multi-component alloys, *J. Comput. Phys.* 221 (1) (2007) 9–40.
- [31] N. Al-Rawahi, G. Tryggvason, Numerical simulation of dendritic solidification with convection: two-dimensional geometry, *J. Comput. Phys.* 180 (2) (2002) 471–496.
- [32] C. Li, S. Garimella, J. Simpson, Fixed-grid front-tracking algorithm for solidification problems, part I: method and validation, *Numer. Heat Transf., Part B* 43 (2) (2003) 117–141.
- [33] M. Bernauer, R. Herzog, Implementation of an X-FEM solver for the classical two-phase Stefan problem, *J. Sci. Comput.* 52 (2012) 271–293.
- [34] L. Salvatori, N. Tosi, Stefan problem through extended finite elements: review and further investigations, *Algorithms* 2 (2009) 1177–1220.
- [35] Y. Zhao, C. Zhao, Z. Xu, H. Xu, Modeling metal foam enhanced phase change heat transfer in thermal energy storage by using phase field method, *Int. J. Heat Mass Transf.* 99 (2016) 170–181.
- [36] J. Yang, L. Yang, C. Xu, X. Du, Numerical analysis on thermal behavior of solid-liquid phase change within copper foam with varying porosity, *Int. J. Heat Mass Transf.* 84 (2015) 1008–1018.
- [37] X. Han, Y. Tian, C. Zhao, An effectiveness study of enhanced heat transfer in phase change materials (PCMs), *Int. J. Heat Mass Transf.* 60 (2013) 459–468.
- [38] S. Jaisankar, S.R. Rao, A central Rankine–Hugoniot solver for hyperbolic conservation laws, *J. Comput. Phys.* 228 (2009) 770–798.
- [39] D. Chopp, Another look at velocity extensions in the level set method, *SIAM J. Sci. Comput.* 31 (5) (2009) 3255–3273.
- [40] M. Namdar Zanganeh, J. Kraaijevanger, H. Burrman, J. Jansen, W. Rossen, Challenges in adjoint-based optimization of a foam EOR process, *Comput. Geosci.* 18 (3–4) (2014) 563–577.
- [41] W. Rossen, S. Zeilinger, J.-X. Shi, M. Lim, Simplified mechanistic simulation of foam processes in porous media, *SPE J.* 4 (3) (1999), SPE-57678-PA.
- [42] M. Berger, J. Olinger, Adaptive mesh refinement for hyperbolic partial differential equations, *J. Comput. Phys.* 53 (3) (1984) 484–512.
- [43] J. Kraaijevanger, *On the Analytical Solution of a Quasi-Linear 1-Dimensional Heat Equation*, 2011.
- [44] J. Crank, *Free and Moving Boundary Problems*, Clarendon Press, Oxford, 1984.