where $\varphi_j(\mathbf{x})$ is a finite element test function. S_i is the support of basis function i and $S_{i,j}$ is the shared support of basis functions i and j.

A number of temporal discretizations are considered in this paper. Fully explicit temporal discretizations considered include forward Euler:

$$\mathbf{M}^C \frac{\mathbf{U}^{n+1} - \mathbf{U}^n}{\Delta t} + \mathbf{A}\mathbf{U}^n = \mathbf{b}^n \,, \tag{10}$$

as well as Strong Stability Preserving Runge Kutta (SSPRK) methods that can be expressed in the following form:

$$\hat{\mathbf{U}}^0 = \mathbf{U}^n \,, \tag{11a}$$

$$\hat{\mathbf{U}}^i = \gamma_i \mathbf{U}^n + \zeta_i \left[\hat{\mathbf{U}}^{i-1} + \Delta t \mathbf{G}(t^n + c_i \Delta t, \hat{\mathbf{U}}^{i-1}) \right] , \quad i = 1, \ldots, s \,, \tag{11b}$$

$$\mathbf{U}^{n+1} = \hat{\mathbf{U}}^s \,. \tag{11c}$$

where s is the number of stages, γ_i, ζ_i, and c_i are coefficients that correspond to the particular SSPRK method, and $\mathbf{G}$ represents the right-hand-side function of an ODE

$$\frac{d\mathbf{U}}{dt} = \mathbf{G}(t, \mathbf{U}(t)) \,, \tag{12}$$

which in this case is the following:

$$\mathbf{G}(t, \mathbf{U}(t)) = (\mathbf{M}^C)^{-1} (\mathbf{b}(t) - \mathbf{A}\mathbf{U}(t)) \,. \tag{13}$$

SSPRK methods are a subclass of Runge Kutta methods that offer high-order accuracy while preserving stability [29,30]. The form given in Equation (11) makes it clear that these SSPRK methods can be expressed as a linear combination of steps resembling forward Euler steps, with the only difference being that the explicit time dependence of the source is not necessarily on the old time t^n but instead is on a stage time $t^n + c_i \Delta t$. An example is the 3-stage, 3rd-order accurate SSPRK method has the following coefficients:

$$\gamma = \begin{bmatrix} 0 \\ \frac{3}{4} \\ \frac{1}{3} \end{bmatrix} , \quad \zeta = \begin{bmatrix} 1 \\ \frac{1}{4} \\ \frac{2}{3} \end{bmatrix} , \quad c = \begin{bmatrix} 0 \\ 1 \\ \frac{1}{2} \end{bmatrix} . \tag{14}$$

This work also considers the Theta-family of temporal discretizations:

$$\mathbf{M}^C \frac{\mathbf{U}^{n+1} - \mathbf{U}^n}{\Delta t} + \mathbf{A}((1-\theta)\mathbf{U}^n + \theta \mathbf{U}^{n+1}) = (1-\theta)\mathbf{b}^n + \theta \mathbf{b}^{n+1} \,, \tag{15}$$

where $0 \leq \theta \leq 1$ is the implicitness parameter. For example, θ values of 0, $\frac{1}{2}$, and 1 correspond to forward Euler, Crank–Nicholson, and backward Euler discretizations, respectively.

Finally, in the case of a steady-state solve, we have the following system of equations:

$$\mathbf{A}\mathbf{U} = \mathbf{b} \,. \tag{16}$$

3. FCT methodology applied to particle transport

Recall that the FCT algorithm is built from a low-order scheme and a high-order scheme. Section 3.1 describes the low-order scheme, and Section 3.2 describes the high-order scheme. Section 3.3 describes the FCT scheme combined from these components.

3.1. Low-order scheme

The role of a low-order scheme in the context of the FCT algorithm is to provide a fail-safe solution, which has desirable properties such as positivity-preservation and lack of spurious oscillations. These properties come at the cost of excessive artificial diffusion and thus a lesser degree of accuracy. However, the idea of the FCT algorithm is to undo some of the over-dissipation of the low-order scheme as much as possible without violating some physically-motivated solution bounds.

Here positivity-preservation and monotonicity are achieved by requiring that the matrix of the low-order system satisfies the M-matrix property. M-matrices are a subset of inverse-positive matrices and have the monotone property. For instance, consider the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$; If $\mathbf{A}$ is an M-matrix, then the following property is verified:

$$\text{If } \mathbf{b} \geq 0, \text{ then } \mathbf{x} \geq 0 \,. \tag{17}$$

Please cite this article in press as: J.E. Hansel, J.C. Ragusa, Flux-corrected transport techniques applied to the radiation transport equation discretized with continuous finite elements, J. Comput. Phys. (2017), https://doi.org/10.1016/j.jcp.2017.10.029

ARTICLE IN PRESS
JID:YJCPH AID:7667 /FLA [m3G; v1.224; Prn:7/11/2017; 0:59] P.4 (1-17)
4 J.E. Hansel, J.C. Ragusa / Journal of Computational Physics ••• (••••) •••–•••

where $\varphi_j(\mathbf{x})$ is a finite element test function. S_i is the support of basis function i and $S_{i,j}$ is the shared support of basis functions i and j.

A number of temporal discretizations are considered in this paper. Fully explicit temporal discretizations considered include forward Euler:

$$\mathbf{M}^C \frac{\mathbf{U}^{n+1} - \mathbf{U}^n}{\Delta t} + \mathbf{A}\mathbf{U}^n = \mathbf{b}^n \,, \tag{10}$$

as well as Strong Stability Preserving Runge Kutta (SSPRK) methods that can be expressed in the following form:

$$\hat{\mathbf{U}}^0 = \mathbf{U}^n \,, \tag{11a}$$

$$\hat{\mathbf{U}}^i = \gamma_i \mathbf{U}^n + \zeta_i \left[\hat{\mathbf{U}}^{i-1} + \Delta t \mathbf{G}(t^n + c_i \Delta t, \hat{\mathbf{U}}^{i-1}) \right] , \quad i = 1, \ldots, s \,, \tag{11b}$$

$$\mathbf{U}^{n+1} = \hat{\mathbf{U}}^s \,. \tag{11c}$$

where s is the number of stages, γ_i, ζ_i, and c_i are coefficients that correspond to the particular SSPRK method, and $\mathbf{G}$ represents the right-hand-side function of an ODE

$$\frac{d\mathbf{U}}{dt} = \mathbf{G}(t, \mathbf{U}(t)) \,, \tag{12}$$

which in this case is the following:

$$\mathbf{G}(t, \mathbf{U}(t)) = (\mathbf{M}^C)^{-1} (\mathbf{b}(t) - \mathbf{A}\mathbf{U}(t)) \,. \tag{13}$$

SSPRK methods are a subclass of Runge Kutta methods that offer high-order accuracy while preserving stability [29,30]. The form given in Equation (11) makes it clear that these SSPRK methods can be expressed as a linear combination of steps resembling forward Euler steps, with the only difference being that the explicit time dependence of the source is not necessarily on the old time t^n but instead is on a stage time $t^n + c_i \Delta t$. An example is the 3-stage, 3rd-order accurate SSPRK method has the following coefficients:

$$\gamma = \begin{bmatrix} 0 \\ \frac{3}{4} \\ \frac{1}{3} \end{bmatrix} , \quad \zeta = \begin{bmatrix} 1 \\ \frac{1}{4} \\ \frac{2}{3} \end{bmatrix} , \quad c = \begin{bmatrix} 0 \\ 1 \\ \frac{1}{2} \end{bmatrix} . \tag{14}$$

This work also considers the Theta-family of temporal discretizations:

$$\mathbf{M}^C \frac{\mathbf{U}^{n+1} - \mathbf{U}^n}{\Delta t} + \mathbf{A}((1-\theta)\mathbf{U}^n + \theta \mathbf{U}^{n+1}) = (1-\theta)\mathbf{b}^n + \theta \mathbf{b}^{n+1} \,, \tag{15}$$

where $0 \leq \theta \leq 1$ is the implicitness parameter. For example, θ values of 0, $\frac{1}{2}$, and 1 correspond to forward Euler, Crank–Nicholson, and backward Euler discretizations, respectively.

Finally, in the case of a steady-state solve, we have the following system of equations:

$$\mathbf{A}\mathbf{U} = \mathbf{b} \,. \tag{16}$$

3. FCT methodology applied to particle transport

Recall that the FCT algorithm is built from a low-order scheme and a high-order scheme. Section 3.1 describes the low-order scheme, and Section 3.2 describes the high-order scheme. Section 3.3 describes the FCT scheme combined from these components.

3.1. Low-order scheme

The role of a low-order scheme in the context of the FCT algorithm is to provide a fail-safe solution that has desirable properties such as positivity-preservation and lack of spurious oscillations. These properties come at the cost of excessive artificial diffusion and thus a lesser degree of accuracy. However, the idea of the FCT algorithm is to undo some of the over-dissipation of the low-order scheme as much as possible without violating some physically-motivated solution bounds.

Here positivity-preservation and monotonicity are achieved by requiring that the matrix of the low-order system satisfies the M-matrix property. M-matrices are a subset of inverse-positive matrices and have the monotone property. For instance, consider the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$; If $\mathbf{A}$ is an M-matrix, then the following property is verified:

$$\text{If } \mathbf{b} \geq 0, \text{ then } \mathbf{x} \geq 0 \,. \tag{17}$$

UNCORRECTED PROOF

Please cite this article in press as: J.E. Hansel, J.C. Ragusa, Flux-corrected transport techniques applied to the radiation transport equation discretized with continuous finite elements, J. Comput. Phys. (2017), https://doi.org/10.1016/j.jcp.2017.10.029

Fig. 1. Illustration of cell degree of freedom indices $\mathcal{I}_K$.

Hence, given that the linear system matrix is an M-matrix, positivity-preservation is proven by proving positivity of the right-hand-side vector **b**. This monotonicity property of the linear system matrix is also responsible for the satisfaction of a discrete maximum principle [28].

In this section, a first-order viscosity method introduced by Guermond [28] will be adapted to the transport equation given by Equation (4). This method uses an element-wise artificial viscosity definition in conjunction with a graph-theoretic local viscous bilinear form that makes the method valid for arbitrary element shapes and dimensions. These definitions will be shown to ensure that the system matrix is a non-singular M-matrix.

The graph-theoretic local viscous bilinear form has the following definition.

Definition 1 *(Local viscous bilinear form).* The local viscous bilinear form for element K is defined as follows:

$$d_K(\varphi_j, \varphi_i) \equiv \begin{cases} -\frac{1}{n_K - 1} V_K & i \neq j, \quad i, j \in \mathcal{I}_K, \\ V_K & i = j, \quad i, j \in \mathcal{I}_K, \\ 0 & \text{otherwise}, \end{cases} \tag{18}$$

where V_K is the volume of cell K, $\mathcal{I}_K$ is the set of degree of freedom indices such that the corresponding test function has support on cell K, and n_K is the number of indices in that set.

This bilinear form bears resemblance to a standard Laplacian bilinear form: the diagonal entries are positive, the off-diagonal entries are negative, and the row sums are zero. These facts will be invoked in the proof of the M-matrix conditions later in this section.

The element-wise low-order viscosity definition from [28] is adapted to account for the reaction term in the transport equation, Equation (4), but otherwise remains unchanged.

Definition 2 *(Low-order viscosity).* The low-order viscosity for cell K is defined as follows:

$$\nu_K^L \equiv \max_{i \neq j \in \mathcal{I}_K} \frac{\max(0, A_{i,j})}{-\sum_{T \in \mathcal{K}(S_{i,j})} d_T(\varphi_j, \varphi_i)}, \tag{19}$$

where $A_{i,j}$ is the i, j entry of matrix **A** given by Equation (8c), $\mathcal{I}_K$ is the set of degree of freedom indices corresponding to basis functions that have support on cell K (this is illustrated in Fig. 1 – the indicated nodes have degree of freedom indices belonging to $\mathcal{I}_K$), and $\mathcal{K}(S_{i,j})$ is the set of cell indices for which the cell domain and the shared support $S_{i,j}$ overlap.

This viscosity definition is designed to give the minimum amount of artificial diffusion without violating the M-matrix conditions.

Now that the low-order artificial diffusion operator (bilinear form + viscosity definitions) has been provided, we describe the low-order system. Consider a modification of the Galerkin scheme given in Equation (8) which lumps the mass matrix ($\mathbf{M}^C \rightarrow \mathbf{M}^L$) and adds an artificial diffusion operator $\mathbf{D}^L$, hereafter called the low-order diffusion matrix:

$$\mathbf{M}^L \frac{d\mathbf{U}^L}{dt} + (\mathbf{A} + \mathbf{D}^L)\mathbf{U}^L(t) = \mathbf{b}(t), \tag{20}$$

where $\mathbf{U}^L(t)$ denotes the discrete low-order solution values. Defining the low-order steady-state system matrix $\mathbf{A}^L \equiv \mathbf{A} + \mathbf{D}^L$, the low-order system for the steady-state system, explicit Euler system, and Theta system, respectively, are Steady-state:

$$\mathbf{A}^L \mathbf{U}^L = \mathbf{b}. \tag{21a}$$

Explicit Euler:

Please cite this article in press as: J.E. Hansel, J.C. Ragusa, Flux-corrected transport techniques applied to the radiation transport equation discretized with continuous finite elements, J. Comput. Phys. (2017), https://doi.org/10.1016/j.jcp.2017.10.029

ARTICLE IN PRESS
JID:YJCPH AID:7667 /FLA [m3G; v1.224; Prn:7/11/2017; 0:59] P.5 (1-17)
J.E. Hansel, J.C. Ragusa / Journal of Computational Physics ••• (••••) •••–••• 5

Fig. 1. Illustration of cell degree of freedom indices $\mathcal{I}_K$.

Hence, given that the linear system matrix is an M-matrix, positivity-preservation is proven by proving positivity of the right-hand-side vector **b**. This monotonicity property of the linear system matrix is also responsible for the satisfaction of a discrete maximum principle [28].

In this section, a first-order viscosity method introduced by Guermond [28] will be adapted to the transport equation given by Equation (4). This method uses an element-wise artificial viscosity definition in conjunction with a graph-theoretic local viscous bilinear form that makes the method valid for arbitrary element shapes and dimensions. These definitions will be shown to ensure that the system matrix is a non-singular M-matrix.

The graph-theoretic local viscous bilinear form has the following definition.

Definition 1 (Local viscous bilinear form). The local viscous bilinear form for element K is defined as follows:

$$d_K(\varphi_j, \varphi_i) \equiv \begin{cases} -\frac{1}{n_K - 1} V_K & i \neq j, \quad i, j \in \mathcal{I}_K, \\ V_K & i = j, \quad i, j \in \mathcal{I}_K, \\ 0 & \text{otherwise}, \end{cases} \tag{18}$$

where V_K is the volume of cell K, $\mathcal{I}_K$ is the set of degree of freedom indices such that the corresponding test function has support on cell K, and n_K is the number of indices in that set.

This bilinear form bears resemblance to a standard Laplacian bilinear form: the diagonal entries are positive, the off-diagonal entries are negative, and the row sums are zero. These facts will be invoked in the proof of the M-matrix conditions later in this section.

The element-wise low-order viscosity definition from [28] is adapted to account for the reaction term in the transport equation, Equation (4), but otherwise remains unchanged.

Definition 2 (Low-order viscosity). The low-order viscosity for cell K is defined as follows:

$$\nu_K^L \equiv \max_{i \neq j \in \mathcal{I}_K} \frac{\max(0, A_{i,j})}{-\sum_{T \in \mathcal{K}(S_{i,j})} d_T(\varphi_j, \varphi_i)}, \tag{19}$$

where $A_{i,j}$ is the i, j entry of matrix **A** given by Equation (8c), $\mathcal{I}_K$ is the set of degree of freedom indices corresponding to basis functions that have support on cell K (this is illustrated in Fig. 1 – the indicated nodes have degree of freedom indices belonging to $\mathcal{I}_K$), and $\mathcal{K}(S_{i,j})$ is the set of cell indices for which the cell domain and the shared support $S_{i,j}$ overlap.

This viscosity definition is designed to give the minimum amount of artificial diffusion without violating the M-matrix conditions.

Now that the low-order artificial diffusion operator (bilinear form + viscosity definitions) has been provided, we describe the low-order system. Consider a modification of the Galerkin scheme given in Equation (8) which lumps the mass matrix ($\mathbf{M}^C \rightarrow \mathbf{M}^L$) and adds an artificial diffusion operator $\mathbf{D}^L$, hereafter called the low-order diffusion matrix:

$$\mathbf{M}^L \frac{d\mathbf{U}^L}{dt} + (\mathbf{A} + \mathbf{D}^L)\mathbf{U}^L(t) = \mathbf{b}(t), \tag{20}$$

where $\mathbf{U}^L(t)$ denotes the discrete low-order solution values. Defining the low-order steady-state system matrix $\mathbf{A}^L \equiv \mathbf{A} + \mathbf{D}^L$, the low-order system for the steady-state system, explicit Euler system, and Theta system, respectively, are:
Steady-state:

$$\mathbf{A}^L \mathbf{U}^L = \mathbf{b}. \tag{21a}$$

Please cite this article in press as: J.E. Hansel, J.C. Ragusa, Flux-corrected transport techniques applied to the radiation transport equation discretized with continuous finite elements, J. Comput. Phys. (2017), https://doi.org/10.1016/j.jcp.2017.10.029

UNCORRECTED PROOF

$$\mathbf{M}^L\mathbf{U}^{L,n+1} = \mathbf{M}^L\mathbf{U}^n + \Delta t\left(\mathbf{b}^n - \mathbf{A}^L\mathbf{U}^n\right). \tag{21b}$$

Theta scheme:

$$\left(\mathbf{M}^L + \theta\Delta t\mathbf{A}^L\right)\mathbf{U}^{L,n+1} = \mathbf{M}^L\mathbf{U}^n + \Delta t(\mathbf{b}^\theta - \mathbf{A}^L(1-\theta)\mathbf{U}^n), \tag{21c}$$

where $\mathbf{b}^\theta \equiv (1-\theta)\mathbf{b}^n + \theta\mathbf{b}^{n+1}$. The low-order diffusion matrix is assembled element-wise using the local viscous bilinear form and low-order viscosity definitions:

$$D^L_{i,j} \equiv \sum_{K\in\mathcal{K}(S_{i,j})} \nu^L_K d_K(\varphi_j,\varphi_i). \tag{22}$$

Now the low-order scheme has been fully described, some statements will be made on its properties. Firstly the M-matrix property will be shown for the low-order matrix $\mathbf{A}^L$.

Theorem 1 *(M-matrix property). The low-order steady-state system matrix $\mathbf{A}^L$ is a non-singular M-matrix.*

Proof. There are many definitions that can be used to identify a non-singular M-matrix; one definition gives that an M-matrix can be identified by verifying both of the following properties [31]:

1. strict positivity of diagonal entries: $A_{i,i} > 0$, $\forall i$ and
2. non-positivity of off-diagonal entries: $A_{i,j} \le 0$, $\forall i$, $\forall j \ne i$.

First, we show that the off-diagonal elements of the matrix $\mathbf{A}^L$ are non-positive. The diffusion matrix entry $D^L_{i,j}$ is bounded as follows:

$$\begin{aligned} D^L_{i,j} &= \sum_{K\in\mathcal{K}(S_{i,j})} \nu^L_K d_K(\varphi_j,\varphi_i) \\ &= \sum_{K\in\mathcal{K}(S_{i,j})} \max_{k\ne\ell\in\mathcal{I}_K}\left(\frac{\max(0,A_{k,\ell})}{-\sum_{T\in\mathcal{K}(S_{k,\ell})} d_T(\varphi_\ell,\varphi_k)}\right) d_K(\varphi_j,\varphi_i). \end{aligned}$$

For an arbitrary quantity $c_{k,\ell} \ge 0, \forall k \ne \ell \in \mathcal{I}$, the following is true for $i \ne j \in \mathcal{I}$: $\max_{k\ne\ell\in\mathcal{I}} c_{k,\ell} \ge c_{i,j}$, and thus for $a \le 0$, $a \max_{k\ne\ell\in\mathcal{I}} c_{k,\ell} \le a c_{i,j}$. Recall that $d_K(\varphi_j,\varphi_i) < 0$ for $j \ne i$. Thus, we have:

$$\begin{aligned} D^L_{i,j} &\le \sum_{K\in\mathcal{K}(S_{i,j})} \frac{\max(0,A_{i,j})}{-\sum_{T\in\mathcal{K}(S_{i,j})} d_T(\varphi_j,\varphi_i)} d_K(\varphi_j,\varphi_i), \qquad j\ne i, \\ &= -\max(0,A_{i,j}) \frac{\sum_{K\in\mathcal{K}(S_{i,j})} d_K(\varphi_j,\varphi_i)}{\sum_{T\in\mathcal{K}(S_{i,j})} d_T(\varphi_j,\varphi_i)}, \qquad j\ne i, \\ &= -\max(0,A_{i,j}), \qquad j\ne i, \\ &\le -A_{i,j}, \qquad j\ne i. \end{aligned}$$

Then applying this relation to the definition of the low-order steady state matrix gives

$$A^L_{i,j} = A_{i,j} + D^L_{i,j} \le 0.$$

Next it will be shown that the row sums are non-negative. Using the fact that $\sum_j \varphi_j(\mathbf{x}) = 1$ and $\sum_j d_K(\varphi_j,\varphi_i) = 0$,

$$\begin{aligned} \sum_j A^L_{i,j} &= \sum_j \int_{S_{i,j}} \left(\mathbf{f}'(u_h)\cdot\nabla\varphi_j + \sigma\varphi_j\right)\varphi_i dV + \sum_j \sum_{K\in\mathcal{K}(S_{i,j})} \nu^L d_K(\varphi_j,\varphi_i), \\ &= \int_{S_i} \left(\mathbf{f}'(u_h)\cdot\nabla\sum_j \varphi_j(\mathbf{x}) + \sigma(\mathbf{x})\sum_j \varphi_j(\mathbf{x})\right)\varphi_i(\mathbf{x}) dV, \end{aligned}$$

Please cite this article in press as: J.E. Hansel, J.C. Ragusa, Flux-corrected transport techniques applied to the radiation transport equation discretized with continuous finite elements, J. Comput. Phys. (2017), https://doi.org/10.1016/j.jcp.2017.10.029

ARTICLE IN PRESS
JID:YJCPH AID:7667 /FLA [m3G; v1.224; Prn:7/11/2017; 0:59] P.6 (1-17)
6 J.E. Hansel, J.C. Ragusa / Journal of Computational Physics ••• (••••) •••–•••

Explicit Euler:

$$\mathbf{M}^L\mathbf{U}^{L,n+1} = \mathbf{M}^L\mathbf{U}^n + \Delta t\left(\mathbf{b}^n - \mathbf{A}^L\mathbf{U}^n\right). \tag{21b}$$

Theta scheme:

$$\left(\mathbf{M}^L + \theta\Delta t\mathbf{A}^L\right)\mathbf{U}^{L,n+1} = \mathbf{M}^L\mathbf{U}^n + \Delta t(\mathbf{b}^\theta - \mathbf{A}^L(1-\theta)\mathbf{U}^n), \tag{21c}$$

where $\mathbf{b}^\theta \equiv (1-\theta)\mathbf{b}^n + \theta\mathbf{b}^{n+1}$. The low-order diffusion matrix is assembled element-wise using the local viscous bilinear form and low-order viscosity definitions:

$$D^L_{i,j} \equiv \sum_{K\in\mathcal{K}(S_{i,j})} \nu^L_K d_K(\varphi_j,\varphi_i). \tag{22}$$

Now the low-order scheme has been fully described, some statements will be made on its properties. Firstly the M-matrix property will be shown for the low-order matrix $\mathbf{A}^L$.

Theorem 1 *(M-matrix property). The low-order steady-state system matrix $\mathbf{A}^L$ is a non-singular M-matrix.*

Proof. There are many definitions that can be used to identify a non-singular M-matrix; one definition gives that an M-matrix can be identified by verifying both of the following properties [31]:

1. strict positivity of diagonal entries: $A_{i,i} > 0$, $\forall i$ and
2. non-positivity of off-diagonal entries: $A_{i,j} \le 0$, $\forall i$, $\forall j \ne i$.

First, we show that the off-diagonal elements of the matrix $\mathbf{A}^L$ are non-positive. The diffusion matrix entry $D^L_{i,j}$ is bounded as follows:

$$\begin{aligned} D^L_{i,j} &= \sum_{K\in\mathcal{K}(S_{i,j})} \nu^L_K d_K(\varphi_j,\varphi_i) \\ &= \sum_{K\in\mathcal{K}(S_{i,j})} \max_{k\ne\ell\in\mathcal{I}_K}\left(\frac{\max(0,A_{k,\ell})}{-\sum_{T\in\mathcal{K}(S_{k,\ell})} d_T(\varphi_\ell,\varphi_k)}\right) d_K(\varphi_j,\varphi_i). \end{aligned}$$

For an arbitrary quantity $c_{k,\ell} \ge 0, \forall k \ne \ell \in \mathcal{I}$, the following is true for $i \ne j \in \mathcal{I}$: $\max_{k\ne\ell\in\mathcal{I}} c_{k,\ell} \ge c_{i,j}$, and thus for $a \le 0$, $a \max_{k\ne\ell\in\mathcal{I}} c_{k,\ell} \le a c_{i,j}$. Recall that $d_K(\varphi_j,\varphi_i) < 0$ for $j \ne i$. Thus, we have:

$$\begin{aligned} D^L_{i,j} &\le \sum_{K\in\mathcal{K}(S_{i,j})} \frac{\max(0,A_{i,j})}{-\sum_{T\in\mathcal{K}(S_{i,j})} d_T(\varphi_j,\varphi_i)} d_K(\varphi_j,\varphi_i), \qquad j\ne i, \\ &= -\max(0,A_{i,j}) \frac{\sum_{K\in\mathcal{K}(S_{i,j})} d_K(\varphi_j,\varphi_i)}{\sum_{T\in\mathcal{K}(S_{i,j})} d_T(\varphi_j,\varphi_i)}, \qquad j\ne i, \\ &= -\max(0,A_{i,j}), \qquad j\ne i, \\ &\le -A_{i,j}, \qquad j\ne i. \end{aligned}$$

Then applying this relation to the definition of the low-order steady state matrix gives

$$A^L_{i,j} = A_{i,j} + D^L_{i,j} \le 0.$$

Next it will be shown that the row sums are non-negative. Using the fact that $\sum_j \varphi_j(\mathbf{x}) = 1$ and $\sum_j d_K(\varphi_j,\varphi_i) = 0$,

$$\begin{aligned} \sum_j A^L_{i,j} &= \sum_j \int_{S_{i,j}} \left(\mathbf{f}'(u_h)\cdot\nabla\varphi_j + \sigma\varphi_j\right)\varphi_i dV + \sum_j \sum_{K\in\mathcal{K}(S_{i,j})} \nu^L d_K(\varphi_j,\varphi_i), \\ &= \int_{S_i} \left(\mathbf{f}'(u_h)\cdot\nabla\sum_j \varphi_j(\mathbf{x}) + \sigma(\mathbf{x})\sum_j \varphi_j(\mathbf{x})\right)\varphi_i(\mathbf{x}) dV, \end{aligned}$$

Please cite this article in press as: J.E. Hansel, J.C. Ragusa, Flux-corrected transport techniques applied to the radiation transport equation discretized with continuous finite elements, J. Comput. Phys. (2017), https://doi.org/10.1016/j.jcp.2017.10.029

UNCORRECTED PROOF

$$= \int_{S_i} \sigma(\mathbf{x})\varphi_i(\mathbf{x})dV\,,$$
$$\geq 0\,.$$

Remark 1. If incoming flux boundary conditions are weakly imposed, then the steady-state system matrix is modified: $\mathbf{A} \rightarrow \tilde{\mathbf{A}}$, and the low-order viscosity then uses the *modified* steady-state matrix $\tilde{\mathbf{A}}$. The non-positivity property of the off-diagonal elements still holds. The non-negativity property of the row sums also holds, owing to the relation $\tilde{A}_{i,j} \geq A_{i,j}$.

If the support S_i is not entirely vacuum ($\sigma(\mathbf{x}) \geq 0$ with $\sigma(\mathbf{x}) > 0$ for some $\mathbf{x}$), then the row sum is *strictly* positive. Proof of strict positivity of the diagonal elements directly follows from proof of non-positivity of the off-diagonal elements and strict positivity of the row sums. Thus both conditions for the non-singular M-matrix property have been met. □

Thus far, we have been proven that the system matrix for the low-order steady-state system is an M-matrix, and it remains to demonstrate the same for each of the transient systems. For the explicit Euler/SSPRK systems, the system matrix is just the lumped mass matrix $\mathbf{M}^L$, which is easily shown to be an M-matrix since it is a positive, diagonal matrix. For the θ temporal discretization, the system matrix is a linear combination of the lumped mass matrix and the low-order steady-state system matrix; this linear combination is also an M-matrix since it is a combination of two M-matrices with non-negative combination coefficients.

To complete the proof of positivity preservation for the low-order scheme, we need to show that the system right-hand-side vectors for each temporal discretization are non-negative.

This is immediate for the steady-state case due to the assumption that the source q is non-negative. The following theorem gives that the system right-hand-side vector for the theta system is non-negative. This theorem extends to explicit Euler discretization since explicit Euler is a special case of the Theta discretization.

Theorem 2 *(Non-negativity of the theta low-order system right-hand-side). If the old solution* $\mathbf{U}^n$ *is non-negative and the time step size* Δt *satisfies*

$$\Delta t \leq \frac{M^L_{i,i}}{(1-\theta)A^L_{i,i}}, \quad \forall i\,, \tag{23}$$

then the new solution $\mathbf{U}^{L,n+1}$ *of the Theta low-order system given by Equation* (21c) *is non-negative, i.e.,* $U_i^{L,n+1} \geq 0$, $\forall i$.

Proof. The right-hand-side vector $\mathbf{y}$ of this system has the entries

$$y_i = \Delta t b_i^\theta + \left(M^L_{i,i} - (1-\theta)\Delta t A^L_{i,i}\right) U_i^n - (1-\theta)\Delta t \sum_{j\neq i} A^L_{i,j} U_j^n\,.$$

As stated previously, the source function q is non-negative and thus $b_i^\theta \geq 0$. Due to the time step size assumption given by Equation (23),

$$M^L_{i,i} - (1-\theta)\Delta t A^L_{i,i} \geq 0\,,$$

and because the off-diagonal terms of $\mathbf{A}^L$ are non-positive, the off-diagonal sum term is non-negative. Thus y_i is a sum of non-negative terms, and the theorem is proven. □

It can also be shown that the described low-order scheme satisfies a local discrete maximum principle, which is easily shown given the M-matrix property. One may decide to use these bounds as the imposed bounds in the FCT algorithm; however, this approach has been found to yield less accurate solutions than the approach to be outlined in Section 3.3 and is thus not discussed here for brevity.

3.2. High-order scheme

This section describes the entropy viscosity method applied to the scalar conservation law given by Equation (4). Recall that the entropy viscosity method is to be used as the high-order scheme in the FCT algorithm, instead of the standard Galerkin method as has been used previously in FCT-FEM schemes; for Galerkin FCT-FEM examples, see, for instance, [22, 24,21,25,26]. Usage of the entropy viscosity method in the FCT algorithm ensures convergence to the entropy solution [27].

The entropy viscosity method has been applied to a number of PDEs such as general scalar conservation laws of the form

$$\frac{\partial u}{\partial t} + \nabla \cdot \mathbf{f}(u) = 0\,, \tag{24}$$

Please cite this article in press as: J.E. Hansel, J.C. Ragusa, Flux-corrected transport techniques applied to the radiation transport equation discretized with continuous finite elements, J. Comput. Phys. (2017), https://doi.org/10.1016/j.jcp.2017.10.029

ARTICLE IN PRESS
JID:YJCPH AID:7667 /FLA [m3G; v1.224; Prn:7/11/2017; 0:59] P.7 (1-17)
J.E. Hansel, J.C. Ragusa / Journal of Computational Physics ••• (••••) •••–••• 7

$$= \int_{S_i} \sigma(\mathbf{x})\varphi_i(\mathbf{x})dV\,,$$
$$\geq 0\,.$$

Remark 1. If incoming flux boundary conditions are weakly imposed, then the steady-state system matrix is modified: $\mathbf{A} \rightarrow \tilde{\mathbf{A}}$, and the low-order viscosity then uses the *modified* steady-state matrix $\tilde{\mathbf{A}}$. The non-positivity property of the off-diagonal elements still holds. The non-negativity property of the row sums also holds, owing to the relation $\tilde{A}_{i,j} \geq A_{i,j}$.

If the support S_i is not entirely vacuum ($\sigma(\mathbf{x}) \geq 0$ with $\sigma(\mathbf{x}) > 0$ for some $\mathbf{x}$), then the row sum is *strictly* positive. Proof of strict positivity of the diagonal elements directly follows from proof of non-positivity of the off-diagonal elements and strict positivity of the row sums. Thus both conditions for the non-singular M-matrix property have been met. □

Thus far, we have proven that the system matrix for the low-order steady-state system is an M-matrix, and it remains to demonstrate the same for each of the transient systems. For the explicit Euler/SSPRK systems, the system matrix is just the lumped mass matrix $\mathbf{M}^L$, which is easily shown to be an M-matrix since it is a positive, diagonal matrix. For the θ temporal discretization, the system matrix is a linear combination of the lumped mass matrix and the low-order steady-state system matrix; this linear combination is also an M-matrix since it is a combination of two M-matrices with non-negative combination coefficients.

To complete the proof of positivity preservation for the low-order scheme, we need to show that the system right-hand-side vectors for each temporal discretization are non-negative.

This is immediate for the steady-state case due to the assumption that the source q is non-negative. The following theorem gives that the system right-hand-side vector for the theta system is non-negative. This theorem extends to explicit Euler discretization since explicit Euler is a special case of the Theta discretization.

Theorem 2 *(Non-negativity of the theta low-order system right-hand-side). If the old solution* $\mathbf{U}^n$ *is non-negative and the time step size* Δt *satisfies*

$$\Delta t \leq \frac{M^L_{i,i}}{(1-\theta)A^L_{i,i}}, \quad \forall i\,, \tag{23}$$

then the new solution $\mathbf{U}^{L,n+1}$ *of the Theta low-order system given by Equation* (21c) *is non-negative, i.e.,* $U_i^{L,n+1} \geq 0$, $\forall i$.

Proof. The right-hand-side vector $\mathbf{y}$ of this system has the entries

$$y_i = \Delta t b_i^\theta + \left(M^L_{i,i} - (1-\theta)\Delta t A^L_{i,i}\right) U_i^n - (1-\theta)\Delta t \sum_{j\neq i} A^L_{i,j} U_j^n\,.$$

As stated previously, the source function q is non-negative and thus $b_i^\theta \geq 0$. Due to the time step size assumption given by Equation (23),

$$M^L_{i,i} - (1-\theta)\Delta t A^L_{i,i} \geq 0\,,$$

and because the off-diagonal terms of $\mathbf{A}^L$ are non-positive, the off-diagonal sum term is non-negative. Thus y_i is a sum of non-negative terms, and the theorem is proven. □

It can also be shown that the described low-order scheme satisfies a local discrete maximum principle, which is easily shown given the M-matrix property. One may decide to use these bounds as the imposed bounds in the FCT algorithm; however, this approach has been found to yield less accurate solutions than the approach to be outlined in Section 3.3 and is thus not discussed here for brevity.

3.2. High-order scheme

This section describes the entropy viscosity method applied to the scalar conservation law given by Equation (4). Recall that the entropy viscosity method is to be used as the high-order scheme in the FCT algorithm, instead of the standard Galerkin method as has been used previously in FCT-FEM schemes; for Galerkin FCT-FEM examples, see, for instance, [22, 24,21,25,26]. Usage of the entropy viscosity method in the FCT algorithm ensures convergence to the entropy solution [27].

The entropy viscosity method has been applied to a number of PDEs such as general scalar conservation laws of the form

$$\frac{\partial u}{\partial t} + \nabla \cdot \mathbf{f}(u) = 0\,, \tag{24}$$

Please cite this article in press as: J.E. Hansel, J.C. Ragusa, Flux-corrected transport techniques applied to the radiation transport equation discretized with continuous finite elements, J. Comput. Phys. (2017), https://doi.org/10.1016/j.jcp.2017.10.029

UNCORRECTED PROOF

the inviscid Euler equations [1,32], and the two-phase seven-equation fluid model [33]. The scalar model studied in this paper does not fit into the general form given by Equation (24), due to the addition of the reaction term σu and source term q. Application of entropy viscosity method to the transport equation model is novel and it is described below.

Since the weak form of the problem does not have a unique solution, one must supply additional conditions called *admissibility* conditions or *entropy* conditions to filter out spurious weak solutions, leaving only the physical weak solution, often called the entropy solution. A number of entropy conditions are valid, but usually the most convenient entropy condition for use in numerical methods takes the form of an *entropy inequality*, such as the following, which is valid for the general scalar conservation law given by Equation (24):

$$\frac{\partial \eta(u)}{\partial t} + \nabla \cdot \boldsymbol{\Psi}(u) \leq 0\,, \tag{25}$$

which holds for any convex entropy function $\eta(u)$ and associated entropy flux $\boldsymbol{\Psi}(u) \equiv \int \eta'(u)\mathbf{f}'(u)du$. If one can show that this inequality holds for an arbitrary convex entropy function, then one proves it holds for all convex entropy functions [34,1]. For the scalar PDE considered in this paper, the entropy inequality becomes the following:

$$\frac{\partial \eta(u)}{\partial t} + \nabla \cdot \boldsymbol{\Psi}(u) + \eta'(u)\sigma u - \eta'(u)q \leq 0\,. \tag{26}$$

One can verify this inequality by multiplying the governing PDE by $\eta'(u)$ and applying reverse chain rule.

The entropy viscosity method enforces the entropy inequality by measuring local entropy production and dissipating accordingly. In practice, one defines the entropy residual:

$$\mathcal{R}(u) \equiv \frac{\partial \eta(u)}{\partial t} + \nabla \cdot \boldsymbol{\Psi}(u) + \eta'(u)\sigma u - \eta'(u)q\,, \tag{27}$$

which can be viewed as the amount of violation of the entropy inequality. The entropy viscosity for an element K is then defined to be proportional to this violation, for example:

$$\nu_K^{\eta} = \frac{c_{\mathcal{R}} \|\mathcal{R}(u_h)\|_{L^\infty(K)}}{\hat{\eta}_K}\,, \tag{28}$$

where $\hat{\eta}_K$ is a normalization constant with the units of entropy, $c_{\mathcal{R}}$ is a proportionality constant that can be modulated for each problem, and $\|\mathcal{R}(u_h)\|_{L^\infty(K)}$ is the maximum of the entropy residual on element K, which can be approximated as the maximum over the quadrature points of element K. Designing a universally appropriate normalization constant $\hat{\eta}_K$ remains a challenge for the entropy viscosity method (see [32] for an alternate normalization for low-Mach flows). The objective of this normalization coefficient is to prevent the user from needing to make significant adjustments to the tuning parameter $c_{\mathcal{R}}$ for different problems. A definition that produces reasonable results for a large number of problems is the following:

$$\hat{\eta}_K \equiv \|\eta - \bar{\eta}\|_{L^\infty(\mathcal{D})}\,, \tag{29}$$

where $\bar{\eta}$ is the average entropy over the entire computational domain.

In addition to the entropy residual, it can also be beneficial to measure the jump in the gradient of the entropy flux across cell interfaces. Note that given the definition of the entropy flux, the gradient of the entropy flux is $\nabla\boldsymbol{\Psi}(u) = \nabla\eta(u)\mathbf{f}'(u)$. Then let $\mathcal{J}_F$ denote the jump of the normal component of the entropy flux gradient across face F:

$$\mathcal{J}_F \equiv |\mathbf{f}'(u)\cdot\mathbf{n}_F|[\![\nabla\eta(u)\cdot\mathbf{n}_F]\!]\,, \tag{30}$$

where the double square brackets denote a jump quantity. Then we define the maximum jump on a cell:

$$\mathcal{J}_K \equiv \max_{F\in\mathcal{F}(K)} |\mathcal{J}_F|\,. \tag{31}$$

Finally, putting everything together, one can define the entropy viscosity for a cell K to be

$$\nu_K^{\eta} = \frac{c_{\mathcal{R}} \|\mathcal{R}(u_h)\|_{L^\infty(K)} + c_{\mathcal{J}}\mathcal{J}_K}{\hat{\eta}_K}\,. \tag{32}$$

However, it is known that the low-order viscosity for an element, as computed in Section 3.1, gives enough local artificial diffusion for regularization; any amount of viscosity larger than this low-order viscosity would be excessive. Thus, the low-order viscosity for an element is imposed as the upper bound for the high-order viscosity:

$$\nu_K^{H} \equiv \min(\nu_K^{L}, \nu_K^{\eta})\,. \tag{33}$$

One can note that, in smooth regions, this high-order viscosity will be small, and, in regions of strong gradients or discontinuities, the entropy viscosity can be relatively large.

Finally, the high-order system of equations for the various time discretizations are as follows: Steady-state:

Please cite this article in press as: J.E. Hansel, J.C. Ragusa, Flux-corrected transport techniques applied to the radiation transport equation discretized with continuous finite elements, J. Comput. Phys. (2017), https://doi.org/10.1016/j.jcp.2017.10.029

ARTICLE IN PRESS
JID:YJCPH AID:7667 /FLA [m3G; v1.224; Prn:7/11/2017; 0:59] P.8 (1-17)
8 J.E. Hansel, J.C. Ragusa / Journal of Computational Physics ••• (••••) •••–•••

the inviscid Euler equations [1,32], and the two-phase seven-equation fluid model [33]. The scalar model studied in this paper does not fit into the general form given by Equation (24), due to the addition of the reaction term σu and source term q. Application of entropy viscosity method to the transport equation model is novel and it is described below.

Since the weak form of the problem does not have a unique solution, one must supply additional conditions called *admissibility* conditions or *entropy* conditions to filter out spurious weak solutions, leaving only the physical weak solution, often called the entropy solution. A number of entropy conditions are valid, but usually the most convenient entropy condition for use in numerical methods takes the form of an *entropy inequality*, such as the following, which is valid for the general scalar conservation law given by Equation (24):

$$\frac{\partial \eta(u)}{\partial t} + \nabla \cdot \boldsymbol{\Psi}(u) \leq 0\,, \tag{25}$$

which holds for any convex entropy function $\eta(u)$ and associated entropy flux $\boldsymbol{\Psi}(u) \equiv \int \eta'(u)\mathbf{f}'(u)du$. If one can show that this inequality holds for an arbitrary convex entropy function, then one proves it holds for all convex entropy functions [34,1]. For the scalar PDE considered in this paper, the entropy inequality becomes the following:

$$\frac{\partial \eta(u)}{\partial t} + \nabla \cdot \boldsymbol{\Psi}(u) + \eta'(u)\sigma u - \eta'(u)q \leq 0\,. \tag{26}$$

One can verify this inequality by multiplying the governing PDE by $\eta'(u)$ and applying reverse chain rule.

The entropy viscosity method enforces the entropy inequality by measuring local entropy production and dissipating accordingly. In practice, one defines the entropy residual:

$$\mathcal{R}(u) \equiv \frac{\partial \eta(u)}{\partial t} + \nabla \cdot \boldsymbol{\Psi}(u) + \eta'(u)\sigma u - \eta'(u)q\,, \tag{27}$$

which can be viewed as the amount of violation of the entropy inequality. The entropy viscosity for an element K is then defined to be proportional to this violation, for example:

$$\nu_K^{\eta} = \frac{c_{\mathcal{R}} \|\mathcal{R}(u_h)\|_{L^\infty(K)}}{\hat{\eta}_K}\,, \tag{28}$$

where $\hat{\eta}_K$ is a normalization constant with the units of entropy, $c_{\mathcal{R}}$ is a proportionality constant that can be modulated for each problem, and $\|\mathcal{R}(u_h)\|_{L^\infty(K)}$ is the maximum of the entropy residual on element K, which can be approximated as the maximum over the quadrature points of element K. Designing a universally appropriate normalization constant $\hat{\eta}_K$ remains a challenge for the entropy viscosity method (see [32] for an alternate normalization for low-Mach flows). The objective of this normalization coefficient is to prevent the user from needing to make significant adjustments to the tuning parameter $c_{\mathcal{R}}$ for different problems. A definition that produces reasonable results for a large number of problems is the following:

$$\hat{\eta}_K \equiv \|\eta - \bar{\eta}\|_{L^\infty(\mathcal{D})}\,, \tag{29}$$

where $\bar{\eta}$ is the average entropy over the entire computational domain.

In addition to the entropy residual, it can also be beneficial to measure the jump in the gradient of the entropy flux across cell interfaces. Note that given the definition of the entropy flux, the gradient of the entropy flux is $\nabla\boldsymbol{\Psi}(u) = \nabla\eta(u)\mathbf{f}'(u)$. Then let $\mathcal{J}_F$ denote the jump of the normal component of the entropy flux gradient across face F:

$$\mathcal{J}_F \equiv |\mathbf{f}'(u)\cdot\mathbf{n}_F|[\![\nabla\eta(u)\cdot\mathbf{n}_F]\!]\,, \tag{30}$$

where the double square brackets denote a jump quantity. Then we define the maximum jump on a cell:

$$\mathcal{J}_K \equiv \max_{F\in\mathcal{F}(K)} |\mathcal{J}_F|\,. \tag{31}$$

Finally, putting everything together, one can define the entropy viscosity for a cell K to be

$$\nu_K^{\eta} = \frac{c_{\mathcal{R}} \|\mathcal{R}(u_h)\|_{L^\infty(K)} + c_{\mathcal{J}}\mathcal{J}_K}{\hat{\eta}_K}\,. \tag{32}$$

However, it is known that the low-order viscosity for an element, as computed in Section 3.1, gives enough local artificial diffusion for regularization; any amount of viscosity larger than this low-order viscosity would be excessive. Thus, the low-order viscosity for an element is imposed as the upper bound for the high-order viscosity:

$$\nu_K^{H} \equiv \min(\nu_K^{L}, \nu_K^{\eta})\,. \tag{33}$$

One can note that, in smooth regions, this high-order viscosity will be small, and, in regions of strong gradients or discontinuities, the entropy viscosity can be relatively large.

Finally, the high-order system of equations for the steady-state system, explicit Euler system, and Theta system, respectively, are as follows:
Steady-state:

Please cite this article in press as: J.E. Hansel, J.C. Ragusa, Flux-corrected transport techniques applied to the radiation transport equation discretized with continuous finite elements, J. Comput. Phys. (2017), https://doi.org/10.1016/j.jcp.2017.10.029

UNCORRECTED PROOF

$$\mathbf{A}^H\mathbf{U}^H = \mathbf{b}\,. \tag{34a}$$

Explicit Euler:

$$\mathbf{M}^C\frac{\mathbf{U}^{H,n+1}-\mathbf{U}^n}{\Delta t}+\mathbf{A}^{H,n}\mathbf{U}^n=\mathbf{b}^n\,. \tag{34b}$$

Theta scheme:

$$\mathbf{M}^C\frac{\mathbf{U}^{H,n+1}-\mathbf{U}^n}{\Delta t}+\theta\mathbf{A}^{H,n+1}\mathbf{U}^{H,n+1}+(1-\theta)\mathbf{A}^{H,n}\mathbf{U}^n=\mathbf{b}^\theta\,, \tag{34c}$$

where the high-order steady-state system matrix is defined as $\mathbf{A}^H \equiv \mathbf{A}+\mathbf{D}^H$, and the high-order diffusion matrix $\mathbf{D}^H$ is defined similarly to the low-order case but using the high-order viscosity instead of the low-order viscosity:

$$D^H_{i,j} \equiv \sum_{K\in\mathcal{K}(S_{i,j})} \nu^H_K d_K(\varphi_j,\varphi_i)\,. \tag{35}$$

Note that unlike the low-order scheme, the high-order scheme does not lump the mass matrix.

Remark 2. Note that due to the nonlinearity of the entropy viscosity, the entropy viscosity scheme is nonlinear for implicit and steady-state temporal discretizations, and thus some nonlinear solution technique must be utilized. For the results in this paper, a simple fixed-point iteration scheme is used. An alternative such as Newton's method is likely to be advantageous in terms of the number of nonlinear iterations; however, fixed-point is used here for comparison with the nonlinear FCT scheme to be described in Section 3.3.

3.3. FCT scheme

3.3.1. The FCT system

The entropy viscosity method described in Section 3.2 enforces the entropy condition and thus produces numerical approximations that converge to the entropy solution. However, numerical solutions may still contain spurious oscillations and negativities, although these effects are smaller in magnitude than for the corresponding Galerkin solution. In this paper, the flux-corrected transport (FCT) algorithm is used to further mitigate the formation of spurious oscillations and to guarantee the absence of negativities.

The first ingredient of the FCT algorithm is the definition of the antidiffusive fluxes. To arrive at this definition, the low-order systems, given by Equations (21a), (21b), and (21c) for each temporal discretization, are augmented with the addition of the *antidiffusion source* $\mathbf{p}$, which now, instead of producing the low-order solution $\mathbf{U}^L$, produces the high-order solution $\mathbf{U}^H$:

$$\mathbf{A}^L\mathbf{U}^H = \mathbf{b}+\mathbf{p}\,, \tag{36a}$$

$$\mathbf{M}^L\frac{\mathbf{U}^H-\mathbf{U}^n}{\Delta t}+\mathbf{A}^L\mathbf{U}^n=\mathbf{b}^n+\mathbf{p}^n\,, \tag{36b}$$

$$\mathbf{M}^L\frac{\mathbf{U}^H-\mathbf{U}^n}{\Delta t}+\mathbf{A}^L\left(\theta\mathbf{U}^H+(1-\theta)\mathbf{U}^n\right)=\mathbf{b}^\theta+\mathbf{p}^\theta\,. \tag{36c}$$

Then the corresponding high-order systems, given by Equations (34a), (34b), (34c) are subtracted from these equations to give the following definitions for $\mathbf{p}$:

$$\mathbf{p} \equiv \left(\mathbf{D}^L-\mathbf{D}^H\right)\mathbf{U}^H\,, \tag{37a}$$

$$\mathbf{p}^n \equiv -\left(\mathbf{M}^C-\mathbf{M}^L\right)\frac{\mathbf{U}^H-\mathbf{U}^n}{\Delta t}+\left(\mathbf{D}^L-\mathbf{D}^H\right)\mathbf{U}^n\,, \tag{37b}$$

$$\mathbf{p}^\theta \equiv -\left(\mathbf{M}^C-\mathbf{M}^L\right)\frac{\mathbf{U}^H-\mathbf{U}^n}{\Delta t}+(1-\theta)\left(\mathbf{D}^L-\mathbf{D}^{H,n}\right)\mathbf{U}^n+\theta\left(\mathbf{D}^L-\mathbf{D}^{H,n+1}\right)\mathbf{U}^H\,. \tag{37c}$$

The next step is to decompose each antidiffusive source p_i into a sum of antidiffusive fluxes: $p_i=\sum_j P_{i,j}$. Because the matrices $\mathbf{M}^C-\mathbf{M}^L$ and $\mathbf{D}^L-\mathbf{D}^H$ are symmetric and feature row sums of zero, the following are valid antidiffusive flux decompositions:

$$P_{i,j} = \left(D^L_{i,j}-D^H_{i,j}\right)\left(U^H_j-U^H_i\right)\,, \tag{38a}$$

$$P^n_{i,j} = -M^C_{i,j}\left(\frac{U^H_j-U^n_j}{\Delta t}-\frac{U^H_i-U^n_i}{\Delta t}\right)+\left(D^L_{i,j}-D^{H,n}_{i,j}\right)\left(U^n_j-U^n_i\right)\,, \tag{38b}$$

Please cite this article in press as: J.E. Hansel, J.C. Ragusa, Flux-corrected transport techniques applied to the radiation transport equation discretized with continuous finite elements, J. Comput. Phys. (2017), https://doi.org/10.1016/j.jcp.2017.10.029

ARTICLE IN PRESS
JID:YJCPH AID:7667 /FLA [m3G; v1.224; Prn:7/11/2017; 0:59] P.9 (1-17)
J.E. Hansel, J.C. Ragusa / Journal of Computational Physics ••• (••••) •••–••• 9

$$\mathbf{A}^H\mathbf{U}^H = \mathbf{b}\,, \tag{34a}$$

Explicit Euler:

$$\mathbf{M}^C\frac{\mathbf{U}^{H,n+1}-\mathbf{U}^n}{\Delta t}+\mathbf{A}^{H,n}\mathbf{U}^n=\mathbf{b}^n\,, \tag{34b}$$

Theta scheme:

$$\mathbf{M}^C\frac{\mathbf{U}^{H,n+1}-\mathbf{U}^n}{\Delta t}+\theta\mathbf{A}^{H,n+1}\mathbf{U}^{H,n+1}+(1-\theta)\mathbf{A}^{H,n}\mathbf{U}^n=\mathbf{b}^\theta\,, \tag{34c}$$

where the high-order steady-state system matrix is defined as $\mathbf{A}^H \equiv \mathbf{A}+\mathbf{D}^H$, and the high-order diffusion matrix $\mathbf{D}^H$ is defined similarly to the low-order case but using the high-order viscosity instead of the low-order viscosity:

$$D^H_{i,j} \equiv \sum_{K\in\mathcal{K}(S_{i,j})} \nu^H_K d_K(\varphi_j,\varphi_i)\,. \tag{35}$$

Note that unlike the low-order scheme, the high-order scheme does not lump the mass matrix.

Remark 2. Note that due to the nonlinearity of the entropy viscosity, the entropy viscosity scheme is nonlinear for implicit and steady-state temporal discretizations, and thus some nonlinear solution technique must be utilized. For the results in this paper, a simple fixed-point iteration scheme is used. An alternative such as Newton's method is likely to be advantageous in terms of the number of nonlinear iterations; however, fixed-point is used here for comparison with the nonlinear FCT scheme to be described in Section 3.3.

3.3. FCT scheme

3.3.1. The FCT system

The entropy viscosity method described in Section 3.2 enforces the entropy condition and thus produces numerical approximations that converge to the entropy solution. However, numerical solutions may still contain spurious oscillations and negativities, although these effects are smaller in magnitude than for the corresponding Galerkin solution. In this paper, the flux-corrected transport (FCT) algorithm is used to further mitigate the formation of spurious oscillations and to guarantee the absence of negativities.

The first ingredient of the FCT algorithm is the definition of the antidiffusive fluxes. To arrive at this definition, the low-order systems, given by Equations (21a), (21b), and (21c) for each temporal discretization, are augmented with the addition of the *antidiffusion source* $\mathbf{p}$, which now, instead of producing the low-order solution $\mathbf{U}^L$, produces the high-order solution $\mathbf{U}^H$:

$$\mathbf{A}^L\mathbf{U}^H = \mathbf{b}+\mathbf{p}\,, \tag{36a}$$

$$\mathbf{M}^L\frac{\mathbf{U}^H-\mathbf{U}^n}{\Delta t}+\mathbf{A}^L\mathbf{U}^n=\mathbf{b}^n+\mathbf{p}^n\,, \tag{36b}$$

$$\mathbf{M}^L\frac{\mathbf{U}^H-\mathbf{U}^n}{\Delta t}+\mathbf{A}^L\left(\theta\mathbf{U}^H+(1-\theta)\mathbf{U}^n\right)=\mathbf{b}^\theta+\mathbf{p}^\theta\,. \tag{36c}$$

Then the corresponding high-order systems, given by Equations (34a), (34b), (34c) are subtracted from these equations to give the following definitions for $\mathbf{p}$:

$$\mathbf{p} \equiv \left(\mathbf{D}^L-\mathbf{D}^H\right)\mathbf{U}^H\,, \tag{37a}$$

$$\mathbf{p}^n \equiv -\left(\mathbf{M}^C-\mathbf{M}^L\right)\frac{\mathbf{U}^H-\mathbf{U}^n}{\Delta t}+\left(\mathbf{D}^L-\mathbf{D}^H\right)\mathbf{U}^n\,, \tag{37b}$$

$$\mathbf{p}^\theta \equiv -\left(\mathbf{M}^C-\mathbf{M}^L\right)\frac{\mathbf{U}^H-\mathbf{U}^n}{\Delta t}+(1-\theta)\left(\mathbf{D}^L-\mathbf{D}^{H,n}\right)\mathbf{U}^n+\theta\left(\mathbf{D}^L-\mathbf{D}^{H,n+1}\right)\mathbf{U}^H\,. \tag{37c}$$

The next step is to decompose each antidiffusive source p_i into a sum of antidiffusive fluxes: $p_i=\sum_j P_{i,j}$. Because the matrices $\mathbf{M}^C-\mathbf{M}^L$ and $\mathbf{D}^L-\mathbf{D}^H$ are symmetric and feature row sums of zero, the following are valid antidiffusive flux decompositions:

$$P_{i,j} = \left(D^L_{i,j}-D^H_{i,j}\right)\left(U^H_j-U^H_i\right)\,, \tag{38a}$$

$$P^n_{i,j} = -M^C_{i,j}\left(\frac{U^H_j-U^n_j}{\Delta t}-\frac{U^H_i-U^n_i}{\Delta t}\right)+\left(D^L_{i,j}-D^{H,n}_{i,j}\right)\left(U^n_j-U^n_i\right)\,, \tag{38b}$$

Please cite this article in press as: J.E. Hansel, J.C. Ragusa, Flux-corrected transport techniques applied to the radiation transport equation discretized with continuous finite elements, J. Comput. Phys. (2017), https://doi.org/10.1016/j.jcp.2017.10.029

UNCORRECTED PROOF

and

$$U_i^+ \equiv \begin{cases} U_{\max,i}^n e^{-v\Delta t\sigma_{\min,i}} + \frac{q_{\max,i}}{\sigma_{\min,i}}(1 - e^{-v\Delta t\sigma_{\min,i}}), & \sigma_{\min,i} \neq 0 \\ U_{\max,i}^n + v\Delta t q_{\max,i}, & \sigma_{\min,i} = 0. \end{cases} \tag{46c}$$

The other quantities used in the above expressions are:

$$U_{\max,i}^n \equiv \max_{j\in\mathcal{I}(S_i)} U_j^n, \quad U_{\min,i}^n \equiv \min_{j\in\mathcal{I}(S_i)} U_j^n, \tag{46d}$$

$$\sigma_{\max,i} \equiv \max_{\mathbf{x}\in S_i} \sigma(\mathbf{x}), \quad \sigma_{\min,i} \equiv \min_{\mathbf{x}\in S_i} \sigma(\mathbf{x}), \tag{46e}$$

$$q_{\max,i} \equiv \max_{\mathbf{x}\in S_i} q(\mathbf{x}), \quad q_{\min,i} \equiv \min_{\mathbf{x}\in S_i} q(\mathbf{x}). \tag{46f}$$

Note the time step size condition given by Equation (45) implies that when using CFL numbers greater than 1 with implicit time discretizations, these bounds no longer apply. Similar bounds can be derived for $v\Delta t > h_{min}$; however, these bounds for a node i will no longer only depend on the solution values of the immediate neighbors of i; instead, a larger neighborhood must be used in the bounds, making the local solution bounds wider and thus less restrictive and arguably less useful in the FCT algorithm. This represents a significant disadvantage for implicit FCT, not only because the converged FCT solution could contain more undesirable features but also because the wider bounds typically result in a greater number of nonlinear iterations because of the increased freedom in the limiting coefficients.

Steady-state FCT solution bounds can be inferred from Equation (46) by making the substitution $v\Delta t \to s$, where $0 \leq s \leq h_{min}$. This restriction of s similarly ensures that only the nearest neighbors of i are needed for the solution bounds of i. Steady-state FCT unfortunately suffers many of the same drawbacks as implicit FCT because like implicit FCT, its solution bounds are implicit and thus change with each iteration.

3.3.3. Antidiffusion bounds

Bounds imposed on a solution value i, such as the bounds described in Section 3.3.2, directly translate into bounds on the limited antidiffusion source $\hat{p}_i$. These antidiffusion bounds $\hat{p}_i^{\pm}$ for steady-state, explicit Euler, and Theta discretization are respectively derived by solving Equations (39a), (39b), and (39c) for $\hat{p}_i$ and manipulating the inequality $U_i^- \leq U_i \leq U_i^+$. This yields:

$$\hat{p}_i^{\pm} \equiv A_{i,i}^L U_i^{\pm} + \sum_{j\neq i} A_{i,j}^L U_j - b_i, \tag{47a}$$

$$\hat{p}_i^{\pm} \equiv M_{i,i}^L \frac{U_i^{\pm} - U_i^n}{\Delta t} + \sum_j A_{i,j}^L U_j^n - b_i^n, \tag{47b}$$

$$\hat{p}_i^{\pm} \equiv \left(\frac{M_{i,i}^L}{\Delta t} + \theta A_{i,i}^L\right) U_i^{\pm} + \left((1-\theta)A_{i,i}^L - \frac{M_{i,i}^L}{\Delta t}\right) U_i^n + \sum_{j\neq i} A_{i,j}^L U_j^{\theta} - b_i^{\theta}. \tag{47c}$$

We note that, if the limiting coefficients $L_{i,j}$ are selected such that $\hat{p}_i^- \leq \hat{p}_i \leq \hat{p}_i^+$, then the solution bounds are satisfied: $U_i^- \leq U_i \leq U_i^+$.

Limiters such as the Zalesak limiter described in Section 3.3.4 are algebraic operators, taking as input the antidiffusion bounds $\hat{p}_i^{\pm}$ and the antidiffusive fluxes $P_{i,j}$ and returning as output the limiting coefficients $L_{i,j}$. It is important to note that most limiters, including the limiter described in this paper, assume the following: $\hat{p}_i^- \leq 0$, $\hat{p}_i^+ \geq 0$; the reasoning for this assumption is as follows. Recall that FCT starts from the low-order scheme, which is equivalent to the solution with $\hat{p}_i = 0$. The limiter should start from this point so that there is a fail-safe solution for the FCT algorithm: the low-order solution. Otherwise, there is no guarantee that any combination of values of limiting coefficients will achieve the desired condition $\hat{p}_i^- \leq \hat{p}_i \leq \hat{p}_i^+$. If $\hat{p}_i^- > 0$ or $\hat{p}_i^+ < 0$, then the starting state, the low-order solution, with $\hat{p}_i = 0$ is an invalid solution of the FCT algorithm. Some solution bounds automatically satisfy $\hat{p}_i^- \leq 0$ and $\hat{p}_i^+ \geq 0$, but in general these conditions must be enforced. In this paper, the solution bounds are possibly widened by directly enforcing these assumptions:

$$\hat{p}_i^- \leftarrow \min(0, \hat{p}_i^-), \tag{48}$$

$$\hat{p}_i^+ \leftarrow \max(0, \hat{p}_i^+). \tag{49}$$

We have noted that omitting this step can lead to poor results. Without this step, the assumptions of the limiter are violated, and thus limiting coefficients that do not satisfy the imposed solution bounds may be generated.

3.3.4. Limiting coefficients

The results in this paper use the classic multi-dimensional limiter introduced by Zalesak [19]:

Please cite this article in press as: J.E. Hansel, J.C. Ragusa, Flux-corrected transport techniques applied to the radiation transport equation discretized with continuous finite elements, J. Comput. Phys. (2017), https://doi.org/10.1016/j.jcp.2017.10.029

ARTICLE IN PRESS
JID:YJCPH AID:7667 /FLA [m3G; v1.224; Prn:7/11/2017; 0:59] P.11 (1-17)
J.E. Hansel, J.C. Ragusa / Journal of Computational Physics ••• (••••) •••–••• 11
and
$$U_i^+ \equiv \begin{cases} U_{\max,i}^n e^{-v\Delta t\sigma_{\min,i}} + \frac{q_{\max,i}}{\sigma_{\min,i}}(1 - e^{-v\Delta t\sigma_{\min,i}}), & \sigma_{\min,i} \neq 0 \\ U_{\max,i}^n + v\Delta t q_{\max,i}, & \sigma_{\min,i} = 0. \end{cases} \tag{46c}$$
The other quantities used in the above expressions are:
$$U_{\max,i}^n \equiv \max_{j\in\mathcal{I}(S_i)} U_j^n, \quad U_{\min,i}^n \equiv \min_{j\in\mathcal{I}(S_i)} U_j^n, \tag{46d}$$
$$\sigma_{\max,i} \equiv \max_{\mathbf{x}\in S_i} \sigma(\mathbf{x}), \quad \sigma_{\min,i} \equiv \min_{\mathbf{x}\in S_i} \sigma(\mathbf{x}), \tag{46e}$$
$$q_{\max,i} \equiv \max_{\mathbf{x}\in S_i} q(\mathbf{x}), \quad q_{\min,i} \equiv \min_{\mathbf{x}\in S_i} q(\mathbf{x}). \tag{46f}$$
Note the time step size condition given by Equation (45) implies that when using CFL numbers greater than 1 with implicit time discretizations, these bounds no longer apply. Similar bounds can be derived for $v\Delta t > h_{min}$; however, these bounds for a node i will no longer only depend on the solution values of the immediate neighbors of i; instead, a larger neighborhood must be used in the bounds, making the local solution bounds wider and thus less restrictive and arguably less useful in the FCT algorithm. This represents a significant disadvantage for implicit FCT, not only because the converged FCT solution could contain more undesirable features but also because the wider bounds typically result in a greater number of nonlinear iterations because of the increased freedom in the limiting coefficients.
Steady-state FCT solution bounds can be inferred from Equation (46) by making the substitution $v\Delta t \to s$, where $0 \leq s \leq h_{min}$. This restriction of s similarly ensures that only the nearest neighbors of i are needed for the solution bounds of i. Steady-state FCT unfortunately suffers many of the same drawbacks as implicit FCT because like implicit FCT, its solution bounds are implicit and thus change with each iteration.
3.3.3. Antidiffusion bounds
Bounds imposed on a solution value U_i, such as the bounds described in Section 3.3.2, directly translate into bounds on the limited antidiffusion source $\hat{p}_i$. These antidiffusion bounds $\hat{p}_i^{\pm}$ for steady-state, explicit Euler, and Theta discretization are respectively derived by solving Equations (39a), (39b), and (39c) for $\hat{p}_i$ and manipulating the inequality $U_i^- \leq U_i \leq U_i^+$. This yields:
$$\hat{p}_i^{\pm} \equiv A_{i,i}^L U_i^{\pm} + \sum_{j\neq i} A_{i,j}^L U_j - b_i, \tag{47a}$$
$$\hat{p}_i^{\pm} \equiv M_{i,i}^L \frac{U_i^{\pm} - U_i^n}{\Delta t} + \sum_j A_{i,j}^L U_j^n - b_i^n, \tag{47b}$$
$$\hat{p}_i^{\pm} \equiv \left(\frac{M_{i,i}^L}{\Delta t} + \theta A_{i,i}^L\right) U_i^{\pm} + \left((1-\theta)A_{i,i}^L - \frac{M_{i,i}^L}{\Delta t}\right) U_i^n + \sum_{j\neq i} A_{i,j}^L U_j^{\theta} - b_i^{\theta}. \tag{47c}$$
We note that, if the limiting coefficients $L_{i,j}$ are selected such that $\hat{p}_i^- \leq \hat{p}_i \leq \hat{p}_i^+$, then the solution bounds are satisfied: $U_i^- \leq U_i \leq U_i^+$.
Limiters such as the Zalesak limiter described in Section 3.3.4 are algebraic operators, taking as input the antidiffusion bounds $\hat{p}_i^{\pm}$ and the antidiffusive fluxes $P_{i,j}$ and returning as output the limiting coefficients $L_{i,j}$. It is important to note that most limiters, including the limiter described in this paper, assume the following: $\hat{p}_i^- \leq 0$, $\hat{p}_i^+ \geq 0$; the reasoning for this assumption is as follows. Recall that FCT starts from the low-order scheme, which is equivalent to the solution with $\hat{p}_i = 0$. The limiter should start from this point so that there is a fail-safe solution for the FCT algorithm: the low-order solution. Otherwise, there is no guarantee that any combination of values of limiting coefficients will achieve the desired condition $\hat{p}_i^- \leq \hat{p}_i \leq \hat{p}_i^+$. If $\hat{p}_i^- > 0$ or $\hat{p}_i^+ < 0$, then the starting state, the low-order solution, with $\hat{p}_i = 0$ is an invalid solution of the FCT algorithm. Some solution bounds automatically satisfy $\hat{p}_i^- \leq 0$ and $\hat{p}_i^+ \geq 0$, but in general these conditions must be enforced. In this paper, the solution bounds are possibly widened by directly enforcing these assumptions:
$$\hat{p}_i^- \leftarrow \min(0, \hat{p}_i^-), \tag{48}$$
$$\hat{p}_i^+ \leftarrow \max(0, \hat{p}_i^+). \tag{49}$$
We have noted that omitting this step can lead to poor results. Without this step, the assumptions of the limiter are violated, and thus limiting coefficients that do not satisfy the imposed solution bounds may be generated.
3.3.4. Limiting coefficients
The results in this paper use the classic multi-dimensional limiter introduced by Zalesak [19]:
Please cite this article in press as: J.E. Hansel, J.C. Ragusa, Flux-corrected transport techniques applied to the radiation transport equation discretized with continuous finite elements, J. Comput. Phys. (2017), https://doi.org/10.1016/j.jcp.2017.10.029