1. INTRODUCTION

We are concerned with the Cauchy problem for quasilinear anisotropic degenerate parabolic equations of second order with the form

\begin{align*}
\partial_t u + \text{div} f(u) &= \nabla \cdot (A(u) \nabla u), \\
u(0, x) &= u_0(x),
\end{align*}

where \((t, x) \in \mathbb{R}_+ \times \mathbb{R}^d\), \text{div} and \nabla are with respect to \(x \in \mathbb{R}^d\), \(u = u(t, x)\) is the scalar unknown function that is sought,

\begin{equation}
u_0 \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)
\end{equation}
is the initial function, 
\[(1.3) \quad f = (f_1, \ldots, f_d) \in (\text{Lip}_{\text{loc}}(\mathbb{R}))^d \]
is the vector-valued flux function, and 
\[(1.4) \quad A(u) = \sigma^A(u)\sigma^A(u)^\top \geq 0, \quad \text{with} \quad \sigma^A \in (L^\infty_{\text{loc}}(\mathbb{R}))^{d \times K}, \quad 1 \leq K \leq d, \]
is the matrix-valued diffusion function. The symmetric \(d \times d\) matrix \(A(u) = (a_{ij}(u))\) has entries of the form 
\[a_{ij}(u) = \sum_{k=1}^{K} \sigma^A_{ik}(u)\sigma^A_{jk}(u), \quad i, j = 1, \ldots, d.\]
On the space of symmetric matrices, we employ the usual ordering in the sense of quadratic forms. Note that the scalar hyperbolic conservation law \((A \equiv 0)\) is a special case of (1.1).

Nonlinear partial differential equations of type (1.1) model convection-diffusion motions in nature and occur in a variety of applications. Being very selective, we mention here only flow in porous media (see, e.g., [12] and the references therein) and sedimentation-consolidation processes [6]. It is well known that equation (1.1) possesses discontinuous solutions, and weak solutions are not uniquely determined by their initial data; hence (1.1) must be interpreted in the sense of entropy solutions [19, 30, 31]. The uniqueness of entropy solutions was proved in the one-dimensional context by Wu-Yin [32] and Bénilan-Touré [2]. In the multidimensional context with isotropic diffusion (that is, \(A(\cdot) \geq 0\) is a scalar function), a general uniqueness result is much more recent and was proved by Carrillo [7] by using Kružkov’s doubling of variables technique; and various extensions of his result can be found in Bürger-Evje-Karlsen [5], Eymard-Gallouët-Herbin-Michel [15], Karlsen-Risebro [18], Karlsen-Ohlberger [16], Mascia-Porretta-Terracina [24], Michel-Vovelle [25], and Rouvre-Gagneux [28]. Chen-DiBenedetto [8] proved the uniqueness of unbounded entropy solutions by using the doubling of variables technique. Chen-Perthame [9] finally introduced the notion of kinetic solutions and established an \(L^1\) well-posedness theory for the general anisotropic diffusion case by developing a kinetic approach for (1.1). Let us also mention the earlier work by Tassa [29], who proved the uniqueness for piecewise smooth weak solutions. There are also several recent studies concerned with the convergence of various numerical schemes: see [12] for operator splitting methods, [13, 17] for monotone finite difference schemes, [15, 26, 25] for monotone finite volume schemes, and [1, 4] for BGK schemes. All these papers provide the \(L^1\) convergence of approximate solutions without a rate of convergence (an error estimate). As is well known, \(L^1\) error estimates are more desirable for robust scientific computation and prediction, which have been an open problem for the general anisotropic case in numerical analysis.

In the hyperbolic context (i.e., \(A \equiv 0\)), error estimates for the vanishing isotropic viscosity method were derived first in Kuznetsov [20] and more recently in Cockburn-Gremaud [10] and Bouchut-Perthame [3, 27], while various estimates for continuous dependence on the nonlinearity (i.e., the flux function \(f\)) were obtained first in Lucier [22] and later in Bouchut-Perthame [3]. Regarding degenerate parabolic problems with isotropic diffusion (that is, \(A(\cdot)\) is a scalar function), continuous dependence estimates for semigroup solutions, and hence also error estimates for the vanishing isotropic viscosity method, were obtained by Cockburn-Gripenberg [11]; see also [18, 14] for a different approach for the case that the flux function \(f\) also depends on \((t,x)\).
We are concerned with explicit estimates for continuous dependence on the nonlinearities and error estimates for the vanishing anisotropic viscosity method for (1.1). We mention that, even in the isotropic case, continuous dependence estimates have never been derived directly for entropy solutions. The purpose of this paper is to use the Chen-Perthame kinetic approach [9] to develop an abstract $L^1$-framework for continuous dependence and error estimates for (1.1) and to present several applications of this framework.

More precisely, we are interested in comparing an entropy solution $u = u(t, x)$ of (1.1) with an entropy solution $v = v(t, x)$ of

\[ \partial_t v + \text{div} g(v) = \nabla \cdot (B(v) \nabla v) + \text{error terms}, \quad v(0, x) = v_0(x), \]

where

\[ v_0 \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d), \]

\[ g = (g_1, \ldots, g_d) \in (\text{Lip}_{\text{loc}}(\mathbb{R}))^d, \]

and

\[ B = \sigma^B(v)\sigma^B(v)^\top \geq 0, \quad \sigma^B \in (L^\infty_{\text{loc}}(\mathbb{R}))^{d \times K}, \quad 1 \leq K \leq d. \]

The symmetric $d \times d$ matrix $B = B(v) = (b_{ij}(v))$ has entries

\[ b_{ij}(v) = \sum_{k=1}^K \sigma^B_{ik}(v)\sigma^B_{jk}(v), \quad i, j = 1, \ldots, d. \]

Similar to the treatment of hyperbolic problems [3, 27], the error terms will take the form of “partial derivatives” for applications, which will be specified later in Section 3.

The first application of our general $L^1$–framework is an explicit estimate for continuous dependence on the nonlinearities in (1.1). If $g \equiv f$ (see Section 4 for the general case), $u_0 \in BV(\mathbb{R}^d)$, and the error terms are zero in (1.5), we obtain that, for any $t > 0$,

\[ \|u(t, \cdot) - v(t, \cdot)\|_{L^1(\mathbb{R}^d)} \leq \|u_0 - v_0\|_{L^1(\mathbb{R}^d)} + C \sqrt{t} \sqrt{\|A - B\|_\infty}, \]

where the $\infty$ - norm is taken componentwise (see Section 3 for the precise definition). We must emphasize that the proof of a result like (1.9) depends in a fundamental way on using the parabolic dissipation/defect measure identified in Chen-Perthame [9], which is also the cornerstone of the uniqueness proof in [9].

The second application of our $L^1$–framework is an error estimate for the vanishing anisotropic viscosity method for (1.1):

\[ \partial_t v + \text{div} f(v) = \nabla \cdot (A(v) \nabla v) + \mu \nabla \cdot (B(v) \nabla v), \quad v(0, x) = v_0(x), \]

where the matrix $B(v) > 0$ is of the same type as in (1.8). If $u_0 \in BV(\mathbb{R}^d)$, we prove that, for any $t > 0$,

\[ \|u(t, \cdot) - v(t, \cdot)\|_{L^1(\mathbb{R}^d)} \leq \|u_0 - v_0\|_{L^1(\mathbb{R}^d)} + C \sqrt{t} \mu, \]

where $C$ depends only on the $L^\infty$ norms of the matrices $A$ and $B$.

Within our $L^1$–framework, there are two ways to obtain an $L^1$ estimate for $u - v$. A traditional way is to view the equation for the anisotropic viscous approximate solutions as
the original equation perturbed by the error terms taking the form of partial derivatives. If \( v \) is uniformly \( BV \) bounded in space variables, one obtains the optimal \( \frac{1}{2} \) rate of convergence. However, if \( v \) is not \( BV \) bounded, only a sub-optimal rate of convergence can be obtained in this way. The more efficient way is to derive the optimal rate of convergence from an estimate like (1.9) for continuous dependence with \( B \) properly chosen, without the requirement of bounded variation of \( v \). Indeed, in this paper we apply the second way to establish the optimal rate of convergence for the vanishing anisotropic viscosity method for (1.1).

While the vanishing anisotropic viscosity method has received almost no attention in the literature, the vanishing isotropic viscosity method for the purely hyperbolic case \( (A \equiv 0) \) is well-studied [3, 10, 19, 20, 27]. After our main results were finished, we noticed a preprint by Makridakis and Perthame [23], whose main result is the optimal rate of convergence for the vanishing anisotropic viscosity method for the hyperbolic problem, with the aid of the kinetic approach in Chen-Perthame [9] for (1.1) and an estimate technique via an auxiliary parabolic equation with constant diffusion. As we can see from our previous discussion, their result can also be obtained directly from our general \( L^1 \)-framework (see Section 5 for the details). One motivation for studying the vanishing anisotropic viscosity method is that anisotropic viscosity approximations are closely related to finite volume numerical schemes on unstructured grids, for which uniform \( BV \) bounds are not available for finite volume schemes, and the standard error estimate theory for hyperbolic problems provides only a sub-optimal rate of convergence.

Although the significant applications of our \( L^1 \)-framework are the estimate for continuous dependence on the nonlinearities and the error estimate for the vanishing anisotropic viscosity method, as an example of direct applications of this framework to numerical methods, we focus in Section 6 on a linear convection-diffusion model equation and derive an \( L^1 \) error estimate for a upwind-central difference scheme. We will present further applications of our \( L^1 \)-framework to numerical methods for nonlinear degenerate parabolic-hyperbolic equations elsewhere. Also we remark that the results in this paper can be extended to more general equations with \((t, x)\)-dependent coefficients; the details will be presented elsewhere.

This paper is organized as follows. We first establish the \( L^1 \)-framework for continuous dependence and error estimates in Sections 2 and 3. Then we apply our general \( L^1 \)-framework to obtain the following results: (i) an explicit estimate for continuous dependence on the nonlinearities in Section 4; (ii) an optimal error estimate for the anisotropic vanishing viscosity method in Section 5; (iii) an error estimate for an upwind-central finite difference scheme for a linear convection-diffusion equation in Section 6.

2. Entropy Solutions and Kinetic Formulation

For any entropy function \( \eta : \mathbb{R} \to \mathbb{R} \), the corresponding entropy fluxes
\[
q = (q_1, \ldots, q_d) : \mathbb{R} \to \mathbb{R}^d \quad \text{and} \quad R = (r_{ij}) : \mathbb{R} \to \mathbb{R}^{d \times d}
\]
are defined by
\[
q'(u) = \eta'(u)f'(u), \quad R'(u) = \eta'(u)A(u).
\]
We will refer to \((\eta, q, R)\) as an entropy-entropy flux triple.

For \( i = 1, \ldots, d \) and \( k = 1, \ldots, K \), we let
\[
\zeta_{ik}^A(u) = \int_0^u \sigma_{ik}^A(w) \, dw
\]
and

\[ \zeta_{ik}^{A,\psi}(u) = \int_0^u \sqrt{\psi(w)} \sigma_{ik}^A(w) \, dw, \quad \text{for } \psi \in C_0(\mathbb{R}). \]

According to Chen-Perthame [9], entropy solutions can now be defined as follows.

**Definition 2.1 (Entropy Solutions).** A function \( u \in L^\infty(\mathbb{R}_+; L^1(\mathbb{R}^d)) \cap L^\infty(\mathbb{R}_+ \times \mathbb{R}^d) \) is an entropy solution of the Cauchy problem (1.1) if the following conditions are satisfied:

**(D.1)** For any \( k = 1, \ldots, K \),

\[ \sum_{i=1}^d \partial_{x_i} \zeta_{ik}^{A}(u) \in L^2(\mathbb{R}_+ \times \mathbb{R}^d). \]

**(D.2)** For any \( k = 1, \ldots, K \) and \( \psi \in C_0(\mathbb{R}) \) with \( \psi \geq 0 \),

\[ \sum_{i=1}^d \partial_{x_i} \zeta_{ik}^{A,\psi}(u) = \sqrt{\psi(u)} \sum_{i=1}^d \partial_{x_i} \zeta_{ik}^{A}(u) \in L^2(\mathbb{R}_+ \times \mathbb{R}^d), \]

and the parabolic dissipation measure \( n^{u,\psi}(t, x) \), defined by

\[ n^{u,\psi}(t, x) = \psi(u(t, x)) \sum_{k=1}^K \left( \sum_{i=1}^d \partial_{x_i} \zeta_{ik}^{A}(u(t, x)) \right)^2, \]

satisfies

\[ n^{u,\psi}(t, x) = \sum_{k=1}^K \left( \sum_{i=1}^d \partial_{x_i} \zeta_{ik}^{A,\psi}(u(t, x)) \right)^2 \text{ a.e. in } \mathbb{R}_+ \times \mathbb{R}^d. \]

**(D.3)** There exists an entropy dissipation measure \( m^{u,\psi}(t, x) \) of the form

\[ m^{u,\psi}(t, x) = \int_{\mathbb{R}} m^\psi(\xi, t, x) \psi(\xi) \, d\xi, \quad \text{for any } \psi \in C_0(\mathbb{R}), \]

for some nonnegative entropy defect measure \( m^\psi(\xi, t, x) \) such that, for any \( C^2 \) entropy-entropy flux triple \( (\eta, q, R) \) with \( \eta'' \in C_0(\mathbb{R}) \), there holds

(2.1) \[ \partial_t \eta(u) + \text{div } q(u) - \nabla \cdot (R'(u) \nabla u) = - \left( m^{u,\eta''} + n^{u,\eta''} \right) \text{ in } \mathcal{D}'(\mathbb{R}_+ \times \mathbb{R}^d), \]

with initial data \( \eta(u)|_{t=0} = \eta(u_0) \). That is, for any test function \( \phi \in \mathcal{D}(\mathbb{R}_+ \times \mathbb{R}^d) \),

\[ \int_{\mathbb{R}_+ \times \mathbb{R}^d} \left( \eta(u) \partial_t \phi + \sum_{i=1}^d q_i(u) \partial_{x_i} \phi + \sum_{i,j=1}^d r_{ij}(u) \partial_{x_i}^2 \phi \right) \, dt \, dx \]

\[ + \int_{\mathbb{R}^d} \eta(u_0(x)) \phi(0, x) \, dx = \int_{\mathbb{R}_+ \times \mathbb{R}^d} \left( m^{u,\eta''} + n^{u,\eta''} \right) \phi \, dt \, dx. \]
Remark 2.1. The nonnegative parabolic defect measure \( n^u(\xi,t,x) \) can be defined as
\[
\tag{2.2} n^u(\xi,t,x) = \sum_{k=1}^{K} \left( \sum_{i=1}^{d} \zeta_{ik}^A(u(t,x)) \right)^2.
\]
Using the duality \((C_0(\mathbb{R}); M(\mathbb{R}))\), the parabolic dissipation measure \( n^{u,\psi}(t,x) \) then takes the form
\[
\tag{2.3} n^{u,\psi}(t,x) = \int_{\mathbb{R}} n^u(\xi,t,x) \psi(\xi) \, d\xi, \quad \psi \in C_0(\mathbb{R}).
\]
In the “diagonal case” \( a_{ij} \equiv 0 \) for all \( i \neq j \), the chain rule \((D.2)\) is automatically satisfied.

We also follow Chen-Perthame [9] to give the equivalent kinetic formulation of entropy solutions for \((1.1)\) which can be derived essentially from duality and the representation formula
\[
\tag{2.4} \partial_t \chi(\xi;u) + f'(\xi) \cdot \nabla_x \chi(\xi;u) = d \sum_{i,j=1}^{d} a_{ij}(\xi) \partial^2_{x_ix_j} \chi(\xi;u) + \partial_\xi \left( m^u + n^u \right)(\xi,t,x) \quad \text{in} \ D'_{\xi,t,x},
\]
where the indicator function \( \chi(\xi;u) \) is defined by
\[
\chi(\xi;u) = \begin{cases} 1_{0<\xi<u}, & \text{when } u > 0, \\ 0, & \text{when } u = 0, \\ -1_{u<\xi<0}, & \text{when } u < 0; \end{cases}
\]
see also Lions-Perthame-Tadmor [21].

For later use, we note that the following formulas are valid:
\[
\tag{2.5} \partial_\xi \chi(\xi;u) = \delta(\xi-u), \quad \partial_\xi \chi(\xi;u) = \delta(\xi) - \delta(\xi-u).
\]

**Definition 2.2** (Kinetic Formulation). Let \( u \) be an entropy solution of \((1.1)\) in the sense of Definition 2.1. Then the kinetic formulation of \((1.1)\) reads
\[
\partial_t \chi(\xi;u) + f'(\xi) \cdot \nabla_x \chi(\xi;u)
\]
\[
= \sum_{i,j=1}^{d} a_{ij}(\xi) \partial^2_{x_ix_j} \chi(\xi;u) + \partial_\xi \left( m^u + n^u \right)(\xi,t,x) \quad \text{in} \ D'_{\xi,t,x},
\]
\[
\chi(\xi;u)|_{t=0} = \chi(\xi;u_0),
\]
for some nonnegative entropy defect measure \( m^u \), which measures “hyperbolicity” in the solution, and some nonnegative parabolic defect measure \( n^u \) with the form \((2.2)\), which measures “parbolicity” in the solution.

3. General \( L^1 \)-Framework

Let \( u \) be an entropy solution of the original problem \((1.1)\). Let \( g \) be the flux function defined in \((1.7)\) and \( \bar{B} = (b_{ij}) \) be the \( d \times d \) symmetric matrix defined in \((1.8)\). We then let \( v \) solve the “approximate” kinetic problem
\[
\partial_t \chi(\xi;v) + g'(\xi) \cdot \nabla_x \chi(\xi;v)
\]
\[
= \sum_{i,j=1}^{d} b_{ij}(\xi) \partial^2_{x_ix_j} \chi(\xi;v) + \partial_\xi \left( m^v + n^v + E \right)(\xi,t,x) \quad \text{in} \ D'_{\xi,t,x},
\]
\[
\chi(\xi;v)|_{t=0} = \chi(\xi;v_0),
\]
for some nonnegative entropy defect measure $m^v$ and nonnegative parabolic defect measure $n^v$ taking the particular form

$$n^v(\xi, t, x) = \delta(\xi - v(t, x)) \left( \sum_{k=1}^K \left( \sum_{i=1}^d \zeta_{ik}^B(v(t, x)) \right)^2 \right),$$

with

$$\zeta_{ik}^B(v) = \int_0^v \sigma_{ik}^B(w) \, dw.$$ 

Correspondingly, for $\psi \in C_0(\mathbb{R})$, define the function

$$\zeta_{ik}^{B, \psi}(v) = \int_0^v \psi(w) \sigma_{ik}^B(w) \, dw.$$ 

Motivated by Bouchut-Perthame [3] and Perthame [27] in their treatment of the hyperbolic problem, we assume that the error term $E(\xi, t, x)$ takes the form of “partial derivatives”:

$$E(\xi, t, x) = \partial^J e_0(\xi, t, x) + \sum_{J=(J_1, \ldots, J_d)} D^J e_1(\xi, t, x)$$

for some error terms $e_0$ and $e_1$ with $J_0, J_\ast \geq 0$ integers and $J$ multi-indices. We assume that the error terms $e_0$ and $e_1$ satisfy

$$\left\| \left( \sup_{\xi} |e_0(\xi, \cdot, \cdot)| , \sup_{\xi} |e_1(\xi, \cdot, \cdot)| \right) \right\|_{L^1_{\text{loc}}(\mathbb{R}_+; L^1(\mathbb{R}^d))} < \infty, \quad 0 \leq |J| \leq J_\ast,$$

where sup is taken over all

$$\xi \in I(v) := [\inf v, \sup v].$$

Define the $d \times d$ symmetric matrix

$$S(\xi) = \left( \sqrt{A(\xi)} - \sqrt{B(\xi)} \right) \left( \sqrt{A(\xi)} - \sqrt{B(\xi)} \right)^\top = (\sigma^A(\xi) - \sigma^B(\xi)) (\sigma^A(\xi) - \sigma^B(\xi))^\top.$$ 

Then the entries of $S(\xi) = (s_{ij}(\xi))$ take the form:

$$s_{ij}(\xi) = \sum_{k=1}^K \left\{ \sigma_{ik}^A(\xi) \sigma_{jk}^A(\xi) - 2 \sigma_{ik}^A(\xi) \sigma_{jk}^B(\xi) + \sigma_{ik}^B(\xi) \sigma_{jk}^B(\xi) \right\}.$$ 

To state the following theorem, we use the notations:

$$S_\infty := \|S\|_\infty = \sup_{\xi \in I(v_0)} \|S(\xi)\|_\infty = \sup_{\xi \in I(v_0)} |s_{ij}(\xi)|,$$

and

$$\|f' - g'\|_\infty := \sup_{\xi \in I(v_0)} \|f'(\xi) - g'(\xi)\|_\infty = \sup_{\xi \in I(v_0)} |f'_i(\xi) - g'_i(\xi)|.$$ 

Hereafter, $C$ will denote positive constants, not necessarily the same at different occurrences, which are independent of the small parameters and time variable $t$.

The main result of this section is the following abstract $L^1$-framework for error estimates.
Theorem 3.1 (General $L^1$-Framework). Let $u \in C(R_+; L^1(R^d))$ be an entropy solution of (1.1), and suppose $v \in L^\infty(R_+; L^1(R^d)) \cap L^\infty(R_+ \times R^d) \cap C(R_+; L^1(R^d))$ solves the “approximate” kinetic problem (3.1). Then, for any $t > 0$ and any $\varepsilon_0, \tilde{\varepsilon}_0, \tilde{\varepsilon}_1 > 0$,

$$
\|u(t, \cdot) - v(t, \cdot)\|_{L^1(R^d)} \leq \|u_0 - v_0\|_{L^1(R^d)} + C\left(\mathcal{E}_{u_0,1}(\varepsilon_1) + \mathcal{E}_{v_0,1}(\tilde{\varepsilon}_0) + \mathcal{E}_{v_0,1}(\tilde{\varepsilon}_1) + \mathcal{E}_{L^f}^g(\varepsilon_1) + \mathcal{E}_{L^f}^{A-B}(\varepsilon_1) + \mathcal{E}_{v_0,1}(\tilde{\varepsilon}_0, \tilde{\varepsilon}_1)\right),
$$

(3.3)

where

$$
\mathcal{E}_{u_0,1}(\varepsilon_1) = \sup_{|y| < \varepsilon_1} \|u(t, \cdot + y) - u(t, \cdot)\|_{L^1(R^d)},
$$

$$
\mathcal{E}_{v_0,1}(\tilde{\varepsilon}_0) = \sup_{0 < s, \tau < \tilde{\varepsilon}_0} \|v(s, \cdot) - v(\tau, \cdot)\|_{L^1(R^d)},
$$

$$
\mathcal{E}_{v_0,1}(\tilde{\varepsilon}_1) = \sup_{|y| < \tilde{\varepsilon}_1} \|v(t, \cdot + y) - v(t, \cdot)\|_{L^1(R^d)},
$$

$$
\mathcal{E}_{L^f}^g(\varepsilon_1) = \begin{cases} 
\|u_0\|_{L^1(R^d)} \frac{\|f^g\|_{L^\infty}}{\varepsilon_1}, & u_0 \notin BV(R^d), \\
\|u_0\|_{BV(R^d)} t \|f^g\|_{L^\infty}, & u_0 \in BV(R^d), 
\end{cases}
$$

$$
\mathcal{E}_{L^f}^{A-B}(\varepsilon_1) = \begin{cases} 
\|u_0\|_{L^1(R^d)} \frac{\|s\|_{L^\infty}}{\varepsilon_1}, & u_0 \notin BV(R^d), \\
\|u_0\|_{BV(R^d)} \frac{\|s\|_{L^\infty}}{\varepsilon_1}, & u_0 \in BV(R^d), 
\end{cases}
$$

and

$$
\mathcal{E}_{v_0,1}(\tilde{\varepsilon}_0, \tilde{\varepsilon}_1) = \frac{1}{\tilde{\varepsilon}_0} \left( \sup_{\xi} \left\| e_0(\xi, \cdot, \cdot) \right\|_{L^1(0, t_0 + \tilde{\varepsilon}_0; L^1(R^d))} + \frac{1}{\tilde{\varepsilon}_1} \sum_{J = \{j_1, \ldots, j_\ell\} \subseteq J} \left( \sup_{\xi} \left\| e_0(\xi, \cdot, \cdot) \right\|_{L^1(0, t_0 + \tilde{\varepsilon}_0; L^1(R^d))} \right) \right).$$

If $g \equiv f$ and $B \equiv A$, then the terms $\mathcal{E}_{u_0,1}(\varepsilon_1)$, $\mathcal{E}_{L^f}^g(\varepsilon_1)$, and $\mathcal{E}_{L^f}^{A-B}(\varepsilon_1)$ in (3.3) can be dropped, that is, there holds

$$
\|u(t, \cdot) - v(t, \cdot)\|_{L^1(R^d)} \leq \|u_0 - v_0\|_{L^1(R^d)} + C\left(\mathcal{E}_{v_0,1}(\tilde{\varepsilon}_0) + \mathcal{E}_{v_0,1}(\tilde{\varepsilon}_1) + \mathcal{E}_{v_0,1}(\tilde{\varepsilon}_0, \tilde{\varepsilon}_1)\right),
$$

(3.4)

for any $t > 0$ and $\tilde{\varepsilon}_0, \tilde{\varepsilon}_1 > 0$.

Proof. Some arguments in this proof follow Chen-Perthame [9] closely, for which we are very concise here and refer instead to [9] for more details.

We set $\varepsilon = (\varepsilon_0, \varepsilon_1)$, $\varepsilon_0 > 0$ for the forward time regularization and $\varepsilon_1 > 0$ for the space regularization. We then define

$$
\omega_\varepsilon(t, x) := \omega_{v_0}(t)\omega_{\varepsilon_1}(x),
$$
where
\[
\omega_{\varepsilon_0}(t) := \frac{1}{\varepsilon_0} \omega_0 \left( \frac{t}{\varepsilon_0} \right), \quad \omega_{\varepsilon_1}(x) := \frac{1}{\varepsilon_1} \omega_1 \left( \frac{x_1}{\varepsilon_1} \right) \cdots \omega_1 \left( \frac{x_d}{\varepsilon_1} \right),
\]
and \( \omega_\ell \geq 0, \ell = 0, 1, \) denote the normalized regularization kernels with
\[
\int_\mathbb{R} \omega_\ell(\tau) \, d\tau = 1, \quad \text{supp}(\omega_0) \subset (-1, 0), \quad \text{supp}(\omega_1) \subset (-1, 1).
\]

We use the notations
\[
\chi := \chi(\xi, t, x) = \chi(\xi; u(t, x)), \quad \tilde{\chi} := \tilde{\chi}(\xi, t, x) = \chi(\xi; v(t, x)),
\]
\[
\chi_\varepsilon := \chi_\varepsilon(\xi, t, x) = \left( \chi * \omega_\varepsilon \right)(\xi, t, x), \quad \tilde{\chi}_\varepsilon := \tilde{\chi}_\varepsilon(\xi, t, x) = \left( \tilde{\chi} * \omega_\varepsilon \right)(\xi, t, x),
\]
where \( \varepsilon = (\varepsilon_0, \varepsilon_1) > 0 \) is another pair of time-space regularization parameters. Moreover, we use the notations
\[
m^u_\varepsilon := m^u_\varepsilon(\xi, t, x) = \left( m^u * \omega_\varepsilon \right)(\xi, t, x), \quad n^u_\varepsilon := n^u_\varepsilon(\xi, t, x) = \left( n^u * \omega_\varepsilon \right)(\xi, t, x),
\]
\[
m^v_\varepsilon := m^v_\varepsilon(\xi, t, x) = \left( m^v * \omega_\varepsilon \right)(\xi, t, x), \quad n^v_\varepsilon := n^v_\varepsilon(\xi, t, x) = \left( n^v * \omega_\varepsilon \right)(\xi, t, x),
\]
and \( E_\varepsilon = E_\varepsilon(\xi, t, x) \), which is similarly defined.

We intend to study the microscopic functional
\[
0 \leq Q_{\varepsilon, \tilde{\varepsilon}}(\xi, t, x) = |\chi_\varepsilon| + |\tilde{\chi}_\varepsilon| - 2\chi_\varepsilon\tilde{\chi}_\varepsilon.
\]

More precisely, we will calculate
\[
\frac{d}{dt} \int_{\mathbb{R}^d} Q_{\varepsilon, \tilde{\varepsilon}}(\xi, t, x) \, dx \, d\xi.
\]

Note that \( \chi_\varepsilon(\xi, t, x) \) satisfies
\[
\partial_t \chi_\varepsilon + f'(\xi) \cdot \nabla_x \chi_\varepsilon = \sum_{i,j=1}^d a_{ij}(\xi) \partial^2_{x_i x_j} \chi_\varepsilon + \partial_\xi (m^u_\varepsilon + n^u_\varepsilon),
\]
and that \( \tilde{\chi}_\varepsilon(\xi, t, x) \) satisfies
\[
\partial_t \tilde{\chi}_\varepsilon + g'(\xi) \cdot \nabla_x \tilde{\chi}_\varepsilon = \sum_{i,j=1}^d b_{ij}(\xi) \partial^2_{x_i x_j} \tilde{\chi}_\varepsilon + \partial_\xi (m^v_\varepsilon + n^v_\varepsilon + E_\varepsilon).
\]

Multiplying (3.6) by \( \text{sign}(\xi) \), using \( \text{sign}(\xi) \chi_\varepsilon = |\chi_\varepsilon| \), and then integrating in \( (\xi, x) \in \mathbb{R}^d \times \mathbb{R}^d \) yield
\[
\frac{d}{dt} \int_{\mathbb{R}^d} |\chi_\varepsilon| \, dx \, d\xi = -2 \int_{\mathbb{R}^d} (m^u_\varepsilon + n^u_\varepsilon)(0, t, x) \, dx.
\]

Similarly,
\[
\frac{d}{dt} \int_{\mathbb{R}^d} |\tilde{\chi}_\varepsilon| \, dx \, d\xi = -2 \int_{\mathbb{R}^d} (m^v_\varepsilon + n^v_\varepsilon + E_\varepsilon)(0, t, x) \, dx.
\]
We now consider the quadratic term. To this end, we need an additional regularization in the kinetic/velocity variable $\xi$:

$$
\chi_{\varepsilon,\delta}(\xi, t, x) := \left( \chi_{\varepsilon} * \psi_{\delta} \right)(\xi, t, x), \quad \tilde{\chi}_{\varepsilon,\delta}(\xi, t, x) := \left( \tilde{\chi}_{\varepsilon} * \psi_{\delta} \right)(\xi, t, x),
$$

for a standard regularization kernel $\psi_{\delta}$. We also need a $\xi$-truncation $T_L(\xi)$, which is a smooth nonnegative function with bounded support. That is, $T_L(\xi) = T(\xi/L) \to 1$ as $L \to \infty$ with

$$
0 \leq T(\xi) \leq 1, \quad \text{for } \xi \in (-\infty, \infty),
$$

$$
T(\xi) = 1, \quad \text{for } |\xi| \leq 1/2,
$$

$$
T(\xi) = 0, \quad \text{for } |\xi| \geq 1.
$$

The destiny of these additional parameters is that $\delta \downarrow 0$ first and $L \uparrow \infty$ second. Then $\chi_{\varepsilon,\delta}$ satisfies

$$
\partial_t \chi_{\varepsilon,\delta} + f'(\xi) \cdot \nabla_x \chi_{\varepsilon,\delta} = \sum_{i,j=1}^{d} \partial_{\xi_i}^2 \left( (a_{ij} \chi_{\varepsilon}) * \psi_{\delta} \right) + \partial_{\xi} \left( (m^u_{\varepsilon} + n^u_{\varepsilon}) * \psi_{\delta} \right) + R_{\varepsilon,\delta}^u,
$$

(3.10)

and $\tilde{\chi}_{\varepsilon,\delta}$ satisfies

$$
\partial_t \tilde{\chi}_{\varepsilon,\delta} + g'(\xi) \cdot \nabla_x \tilde{\chi}_{\varepsilon,\delta} = \sum_{i,j=1}^{d} \partial_{x_i}^2 \left( (b_{ij} \tilde{\chi}_{\varepsilon}) * \psi_{\delta} \right) + \partial_{\xi} \left( (m^v_{\varepsilon} + n^v_{\varepsilon} + E_{\varepsilon}) * \psi_{\delta} \right) + R_{\varepsilon,\delta}^v.
$$

(3.11)

In (3.10) and (3.11),

$$
R_{\varepsilon,\delta}^u = \text{div}_x \left( f'(\xi) \chi_{\varepsilon,\delta} - (f' \chi_{\varepsilon}) * \psi_{\delta} \right), \quad R_{\varepsilon,\delta}^v = \text{div}_x \left( g'(\xi) \tilde{\chi}_{\varepsilon,\delta} - (g' \tilde{\chi}_{\varepsilon}) * \psi_{\delta} \right).
$$
A simple calculation reveals

\[
\frac{d}{dt} \int_{\mathbb{R}^d} T_L(\xi) \chi_{\varepsilon,\delta} \tilde{\chi}_{\tilde{\varepsilon},\tilde{\delta}} \, dx \, d\xi \\
= - \int_{\mathbb{R}^d} T_L(\xi) \tilde{\chi}_{\tilde{\varepsilon},\tilde{\delta}} \left( f'(\xi) - g'(\xi) \right) \cdot \nabla_x \chi_{\varepsilon} \, dx \, d\xi \\
+ \int_{\mathbb{R}^d} T_L(\xi) \tilde{\chi}_{\tilde{\varepsilon},\tilde{\delta}} \sum_{i,j=1}^d \partial^2 x_{ix_j} \left( (a_{ij} \chi_{\varepsilon}) \star \psi_{\delta} \right) \, dx \, d\xi \\
+ \int_{\mathbb{R}^d} T_L(\xi) \chi_{\varepsilon,\delta} \sum_{i,j=1}^d \partial^2 x_{ix_j} \left( (b_{ij} \tilde{\chi}_{\tilde{\delta}}) \star \psi_{\delta} \right) \, dx \, d\xi \\
+ \int_{\mathbb{R}^d} T_L(\xi) \chi_{\varepsilon,\delta} \partial_\xi \left( (m_{\tilde{\varepsilon}}^u + n_{\tilde{\varepsilon}}^u) \star \psi_{\delta} \right) \, dx \, d\xi \\
+ \int_{\mathbb{R}^d} T_L(\xi) \chi_{\varepsilon,\delta} \partial_\xi \left( (m_{\tilde{\delta}}^u + n_{\tilde{\delta}}^u + E_{\tilde{\varepsilon}}) \star \psi_{\delta} \right) \, dx \, d\xi \\
+ \int_{\mathbb{R}^d} T_L(\xi) \tilde{\chi}_{\tilde{\varepsilon},\tilde{\delta}} R_u \chi_{\varepsilon,\delta} \, dx \, d\xi \\
=: \sum_{\ell=1}^6 I_\ell(t; \varepsilon, \tilde{\varepsilon}, \delta, L).
\]

As in Chen-Perthame [9], we have

\[
\lim_{\delta \downarrow 0} I_6(t; \varepsilon, \tilde{\varepsilon}, \delta, L) = 0 \quad \text{in } L^p(0,T) \text{ for any } 1 \leq p < \infty.
\]

Writing out the convolution products explicitly, we have

\[
I_1(t; \varepsilon, \tilde{\varepsilon}, \delta, L) \\
= - \int T_L(\xi) (f'(\xi) - g'(\xi)) \cdot \nabla_x \omega_\varepsilon(t - s, x - y) \omega_{\tilde{\varepsilon}}(t - s', x - y') \\
\times \psi_\delta(\xi - \eta) \psi_\delta(\xi - \eta') \chi(\eta; u(s, y)) \chi(\eta'; v(s', y')) \, ds \, dy \, d\eta \, ds' \, dy' \, d\eta' \, dx \, d\xi.
\]

Sending first \( \delta \downarrow 0 \) and second \( L \uparrow \infty \), we get

\[
\lim_{L \uparrow \infty} \lim_{\delta \downarrow 0} I_1(t; \varepsilon, \tilde{\varepsilon}, \delta, L) \\
= - \int (f'(\xi) - g'(\xi)) \cdot \nabla_x \omega_\varepsilon(t - s, x - y) \omega_{\tilde{\varepsilon}}(t - s', x - y') \\
\times \chi(\xi; u(s, y)) \chi(\xi; v(s', y')) \, ds \, dy \, ds' \, dy' \, dx \, d\xi \\
=: - \mathcal{E}^{f-g}(t; \varepsilon, \tilde{\varepsilon}).
\]
Integrating by parts yields

\begin{equation}
I_3(t; \varepsilon, \tilde{\varepsilon}, \delta, L) = -\int_{R_\varepsilon \times R_\delta^2} T^t_\varepsilon(\xi) \chi_{\varepsilon, \delta} \left( (m^u_\varepsilon + n^u_\varepsilon) \ast \psi_\delta \right) dx \, d\xi
\end{equation}

\begin{align*}
&- \int_{R_\varepsilon \times R_\delta^2} T_\varepsilon(\xi) \psi_\delta(\xi) \left( (m^u_\varepsilon + n^u_\varepsilon) \ast \psi_\delta \right) dx \, d\xi \\
&+ \int_{R_\varepsilon \times R_\delta^2} T_\varepsilon(\xi) \left( \delta(\xi - u) \ast (\omega_\varepsilon \psi_\delta) \right) \left( m^u_\varepsilon \ast \psi_\delta \right) dx \, d\xi \\
&+ \int_{R_\varepsilon \times R_\delta^2} T_\varepsilon(\xi) \left( \delta(\xi - u) \ast (\omega_\varepsilon \psi_\delta) \right) \left( n^u_\varepsilon \ast \psi_\delta \right) dx \, d\xi \\
=: & \sum_{\ell=1}^4 I_{3,\ell}(t; \varepsilon, \tilde{\varepsilon}, \delta, L),
\end{align*}

and

\begin{equation}
I_5(t; \varepsilon, \tilde{\varepsilon}, \delta, L) = -\int_{R_\varepsilon \times R_\delta^2} T^t_\varepsilon(\xi) \chi_{\varepsilon, \delta} \left( (m^u_\varepsilon + n^u_\varepsilon + E_\varepsilon) \ast \psi_\delta \right) dx \, d\xi
\end{equation}

\begin{align*}
&- \int_{R_\varepsilon \times R_\delta^2} T_\varepsilon(\xi) \psi_\delta(\xi) \left( (m^u_\varepsilon + n^u_\varepsilon + E_\varepsilon) \ast \psi_\delta \right) dx \, d\xi \\
&+ \int_{R_\varepsilon \times R_\delta^2} T_\varepsilon(\xi) \left( \delta(\xi - v) \ast (\omega_\varepsilon \psi_\delta) \right) \left( m^u_\varepsilon \ast \psi_\delta \right) dx \, d\xi \\
&+ \int_{R_\varepsilon \times R_\delta^2} T_\varepsilon(\xi) \left( \delta(\xi - v) \ast (\omega_\varepsilon \psi_\delta) \right) \left( n^u_\varepsilon \ast \psi_\delta \right) dx \, d\xi \\
&+ \int_{R_\varepsilon \times R_\delta^2} T_\varepsilon(\xi) \left( \delta(\xi - v) \ast (\omega_\varepsilon \psi_\delta) \right) \left( E_\varepsilon \ast \psi_\delta \right) dx \, d\xi \\
=: & \sum_{\ell=1}^5 I_{5,\ell}(t; \varepsilon, \tilde{\varepsilon}, \delta, L).
\end{align*}

As in [9], we have

\begin{equation}
\lim_{L \to \infty} \lim_{\delta \to 0} I_{3,1}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) = \lim_{L \to \infty} \lim_{\delta \to 0} I_{5,1}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) = 0,
\end{equation}

in $L^p(0, T)$ for any $1 \leq p < \infty$.

Moreover,

\begin{equation}
\lim_{\delta \to 0} \left( I_{3,2}(t; \varepsilon, \tilde{\varepsilon}, \delta, R) + I_{5,2}(t; \varepsilon, \tilde{\varepsilon}, \delta, R) \right)
= -\int_{R_\delta^2} (m^u_\varepsilon + n^u_\varepsilon + m^u_\tilde{\varepsilon} + n^u_\tilde{\varepsilon} + E_\varepsilon) (0, t, x) \, dx,
\end{equation}

in $L^p(0, T)$ for any $1 \leq p < \infty$.

Clearly,

$$I_{3,3}(t; \varepsilon, \tilde{\varepsilon}, \delta, L), I_{5,3}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) \geq 0 \quad \text{for any} \quad t > 0.$$
Integrating by parts also yields

\[
I_2(\varepsilon, \delta, L) = - \int_{\mathbb{R}^d \times \mathbb{R}^d} T_L(\xi) \sum_{k=1}^K \sum_{i,j=1}^d \partial_{x_i} \left( (\sigma_{ik}^A \tilde{\chi}_\varepsilon) \frac{\partial}{\xi} \right) \partial_{x_j} \left( (\sigma_{jk}^A \chi_\delta) \frac{\partial}{\xi} \right) dx \, d\xi
\]

\begin{align*}
&+ \int_{\mathbb{R}^d \times \mathbb{R}^d} T_L(\xi) \sum_{k=1}^K \sum_{i,j=1}^d \left\{ \partial_{x_i} \left( (\sigma_{ik}^A \tilde{\chi}_\varepsilon) \frac{\partial}{\xi} \right) \partial_{x_j} \left( (\sigma_{jk}^A \chi_\delta) \frac{\partial}{\xi} \right) \\
&\quad - \partial_{x_i} (a_{ij} \chi_\delta) \frac{\partial}{\xi} \right\} dx \, d\xi \\
&=: I_{2,1}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) + I_{2,2}(t; \varepsilon, \tilde{\varepsilon}, \delta, L).
\end{align*}

Similarly,

\[
I_4(t; \varepsilon, \tilde{\varepsilon}, \delta, L) = I_{4,1}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) + I_{4,2}(t; \varepsilon, \tilde{\varepsilon}, \delta, L),
\]

where

\[
I_{4,1}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) = - \int_{\mathbb{R}^d \times \mathbb{R}^d} T_L(\xi) \sum_{k=1}^K \sum_{i,j=1}^d \partial_{x_i} \left( (\sigma_{ik}^B \tilde{\chi}_\varepsilon) \frac{\partial}{\xi} \right) \partial_{x_j} \left( (\sigma_{jk}^B \chi_\delta) \frac{\partial}{\xi} \right) dx \, d\xi
\]

and

\[
I_{4,2}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) = \int_{\mathbb{R}^d \times \mathbb{R}^d} T_L(\xi) \sum_{k=1}^K \sum_{i,j=1}^d \left\{ \partial_{x_i} \left( (\sigma_{ik}^B \tilde{\chi}_\varepsilon) \frac{\partial}{\xi} \right) \partial_{x_j} \left( (\sigma_{jk}^B \chi_\delta) \frac{\partial}{\xi} \right) \\
&\quad - \partial_{x_i} (b_{ij} \tilde{\chi}_\varepsilon) \frac{\partial}{\xi} \right\} dx \, d\xi.
\]

As in Chen-Perthame [9], we have

\[
\lim_{\delta \downarrow 0} I_{2,2}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) = \lim_{\delta \downarrow 0} I_{4,2}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) = 0
\]

in $L^p(0, T)$ for any $1 \leq p < \infty$.

We now study the new term

\[
\mathcal{E}^{A-B}(t; \varepsilon, \tilde{\varepsilon}, \delta, L)
\]

\[
:= I_{2,1}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) + I_{3,4}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) + I_{4,1}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) + I_{5,4}(t; \varepsilon, \tilde{\varepsilon}, \delta, L).
\]

From now on in this proof, for notational simplicity, we drop writing the domains of integration. Writing out explicitly the convolution products, we have

\[
I_{2,1}(t; \varepsilon, \tilde{\varepsilon}, \delta, L)
\]

\[
= - \sum_{k=1}^K \sum_{i,j=1}^d \int T_L(\xi) \partial_{x_i} \omega_\varepsilon(t-s, x-y) \partial_{x_j} \omega_\delta(t-s', x-y') \psi_\theta(\xi-\eta) \psi_\delta(\xi-\eta') \\
\times \sigma_{ik}^A(\eta \chi(\eta; u(s, y))) \sigma_{jk}^A(\eta') \chi(\eta'; v(s', y')) ds \, dy \, d\eta \, ds' \, dy' \, d\eta' \, d\xi \, d\xi.
\]
Similarly,

\[ I_{4,1}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) = - \sum_{k=1}^{K} \sum_{i,j=1}^{d} \int T_L(\xi) \partial_{x_i} \omega_\varepsilon(t - s, x - y) \partial_{x_j} \omega_\tilde{\varepsilon}(t - s', x - y') \psi_\delta(\xi - \eta) \psi_\delta(\xi - \eta') \times \sigma_{ik}^B(\eta) \chi(\eta; u(s, y)) \sigma_{jk}^B(\eta') \chi(\eta'; v(s', y')) \, ds \, dy \, d\eta \, ds' \, dy' \, d\eta' \, dx \, d\xi. \]

Note that

\[ I_{3,4}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) + I_{5,4}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) \]

\[ \geq 2 \sum_{k=1}^{K} \int T_L(\xi) \omega_\varepsilon(t - s, x - y) \omega_\tilde{\varepsilon}(t - s', x - y') \psi_\delta(\xi - u(s, y)) \psi_\delta(\xi - v(s', y')) \times \sum_{i=1}^{d} \partial_{y_i} \zeta_{ik}^A(u(s, y)) \sum_{j=1}^{d} \partial_{y_j} \zeta_{jk}^B(v(s', y')) \, ds \, dy \, ds' \, dy' \, dx \, d\xi, \]

\[ = 2 \sum_{k=1}^{K} \sum_{i,j=1}^{d} \int T_L(\xi) \partial_{x_i} \omega_\varepsilon(t - s, x - y) \partial_{x_j} \omega_\tilde{\varepsilon}(t - s', x - y') \psi_\delta(\xi - \eta) \psi_\delta(\xi - \eta') \times \sigma_{ik}^A(\eta) \chi(\eta; u(s, y)) \sigma_{jk}^B(\eta') \chi(\eta'; v(s', y')) \, ds \, dy \, d\eta \, ds' \, dy' \, d\eta' \, dx \, d\xi, \]

where we have used the chain rule (D.2), integration by parts, and (2.3). From this and the previous calculations, we find

\[ E^{A-B}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) \]

\[ \geq - \sum_{k=1}^{K} \sum_{i,j=1}^{d} \int T_L(\xi) \partial_{x_i} \omega_\varepsilon(t - s, x - y) \partial_{x_j} \omega_\tilde{\varepsilon}(t - s', x - y') \psi_\delta(\xi - \eta) \psi_\delta(\xi - \eta') \times \left\{ \sigma_{ik}^A(\eta) \sigma_{jk}^B(\eta') - 2\sigma_{ik}^A(\eta) \sigma_{jk}^B(\eta') + \sigma_{ik}^A(\eta) \sigma_{jk}^B(\eta') \right\} \times \chi(\eta; u(s, y)) \chi(\eta'; v(s', y')) \, ds \, dy \, d\eta \, ds' \, dy' \, d\eta' \, dx \, d\xi. \]

After performing the changes of variables:

\[ z = \xi - \eta, \quad z' = \xi - \eta', \quad dz \, dz' = d\eta \, d\eta', \]
we get
\[ E^{A-B}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) \]
\[ \geq - \sum_{k=1}^{K} \sum_{i,j=1}^{d} \int T_L(\xi) \partial_{x_i} \omega_\varepsilon(t-s-x-y) \partial_{x_j} \omega_\varepsilon(t-s'-x'-y') \psi_\delta(z) \psi_\delta(z') \]
\[ \times \left\{ \sigma_{ik}^A(\xi-z)\sigma_{jk}^A(\xi-z') - 2\sigma_{ik}^B(\xi-z)\sigma_{jk}^B(\xi-z') + \sigma_{ik}^B(\xi-z)\sigma_{jk}^B(\xi-z') \right\} \]
\[ \times \chi(\xi-z; u(s,y)) \chi(\xi-z'; v(s',y')) \, ds \, dy \, dz \, dy' \, dz' \, dx \, d\xi. \]
From this, it easily follows that
\[ \lim_{L \uparrow \infty} \left( \lim_{\delta \downarrow 0} E^{A-B}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) \right) \]
\[ \geq - \sum_{i,j=1}^{d} \int \partial_{x_i}^2 \omega_\varepsilon(t-s-x-y) \omega_\varepsilon(t-s',x-y') \]
\[ \times s_{ij}(\xi) \chi(\xi; u(s,y)) \chi(\xi; v(s',y')) \, ds \, dy \, ds' \, dy' \, dx \, d\xi \]
\[ =: -E^{A-B}(t; \varepsilon, \tilde{\varepsilon}), \]
where we have also performed integration by parts and used the definition of \( s_{ij}(\xi) \) in (3.2).

Writing out the convolution products, we have
\[ I_{5,5}(t; \varepsilon, \tilde{\varepsilon}, \delta, L) \]
\[ = \int T_L(\xi) \psi_\delta(\xi-v(s,y)) \psi_\delta(\xi-\eta') \omega_\varepsilon(t-s-x-y) \]
\[ \times \left( \partial_{\eta}^2 \omega_\varepsilon(t-s',x-y') e_0(\eta',s',y') \right) \, ds \, dy \, ds' \, dy' \, d\eta \, dx \, d\xi \]
\[ + \sum_{\substack{J=(J_1,\ldots,J_d) \geq 0 \\mid J \leq J_0}} D_x^J \omega_\varepsilon(t-s',x-y') e_1^J(\eta',s',y') \, ds \, dy \, ds' \, dy' \, dx \, d\xi \]
\[ =: -E_v(t; \tilde{\varepsilon}), \quad \text{when } L \uparrow \infty \text{ and } \delta \downarrow 0. \]
Summarizing our calculations from (3.12) to (3.26), we obtain that, for any \( \varepsilon, \tilde{\varepsilon} > 0 \),
\[ \frac{d}{dt} \int_{\mathbb{R}_x \times \mathbb{R}_y} \chi \tilde{\chi}_\varepsilon \, dx \, d\xi \geq - \int_{\mathbb{R}_x} \left( m^u_x + n^u_x + m^v_y + n^v_y + E_\varepsilon \right) (0, t, x) \, dx \]
\[ + E^{f-g}(t; \varepsilon, \tilde{\varepsilon}) + E^{A-B}(t; \varepsilon, \tilde{\varepsilon}) + E_v(t; \tilde{\varepsilon}). \]
Then the estimates (3.5), (3.8), (3.9), and (3.27) yield that, for any \( \varepsilon, \tilde{\varepsilon} > 0 \),
\[
\int_{\mathbb{R}^d} \mathcal{Q}_{\varepsilon, \tilde{\varepsilon}}(\xi, t, x) \, dx \, d\xi
\]
\[
\leq \int_{\mathbb{R}^d} \mathcal{Q}_{\varepsilon, \tilde{\varepsilon}}(\xi, 0, x) \, dx \, d\xi + \int_0^t \mathcal{E}^{f-g}(\tau; \varepsilon, \tilde{\varepsilon}) \, d\tau + \int_0^t \mathcal{E}^{A-B}(t; \varepsilon, \tilde{\varepsilon}) \, d\tau + \int_0^t \mathcal{E}_v(t; \varepsilon, \tilde{\varepsilon}) \, d\tau.
\]

Similarly, we have
\[
\int_{\mathbb{R}^d} \mathcal{Q}_{\varepsilon, \tilde{\varepsilon}}(\xi, 0, x) \, dx \, d\xi
\]
\[
= \int \left( |u(s, y) - v(s', y')| - 2 \min \left( |u(s, y)|, |v(s', y)| \right) 1_{\{|\text{sign}(u(s, y)v(s', y))| > 0\}} \right) \times \omega_\varepsilon(-s, x-y) \omega_\tilde{\varepsilon}(-s', x-y') \, ds \, dy \, ds' \, dy' \, dx
\]
\[
= \int |u(s, y) - v(s', y')| \omega_\varepsilon(-s, x-y) \omega_\tilde{\varepsilon}(-s', x-y') \, ds \, dy \, ds' \, dy' \, dx.
\]

A standard calculation reveals
\[
\int_{\mathbb{R}^d} \mathcal{Q}_{\varepsilon, \tilde{\varepsilon}}(\xi, 0, x) \, dx \, d\xi
\]
\[
\leq \|u_0 - v_0\|_{L^1(\mathbb{R}^d)} + \sup_{0 < s < \varepsilon_0} \|u(s, \cdot) - u_0(\cdot)\|_{L^1(\mathbb{R}^d)} + \sup_{0 < s < \tilde{\varepsilon}_0} \|v(s, \cdot) - v_0(\cdot)\|_{L^1(\mathbb{R}^d)}
\]
\[
+ \sup_{|y| < \varepsilon_1} \|u_0(\cdot + y) - u_0(\cdot)\|_{L^1(\mathbb{R}^d)} + \sup_{|y| < \tilde{\varepsilon}_1} \|v_0(\cdot + y) - v_0(\cdot)\|_{L^1(\mathbb{R}^d)}.
\]

Similarly, we find
\[
\int_{\mathbb{R}^d} \mathcal{Q}_{\varepsilon, \tilde{\varepsilon}}(\xi, t, x) \, dx \, d\xi
\]
\[
\geq \|u(t) - v(t)\|_{L^1(\mathbb{R}^d)} - \sup_{0 < s - t < \varepsilon_0} \|u(s, \cdot) - u(t, \cdot)\|_{L^1(\mathbb{R}^d)} - \sup_{0 < s - t < \tilde{\varepsilon}_0} \|v(s, \cdot) - v(t, \cdot)\|_{L^1(\mathbb{R}^d)}
\]
\[
- \sup_{|y| \leq \varepsilon_1} \|u(t, \cdot + y) - u(t, \cdot)\|_{L^1(\mathbb{R}^d)} - \sup_{|y| \leq \tilde{\varepsilon}_1} \|v(t, \cdot + y) - v(t, \cdot)\|_{L^1(\mathbb{R}^d)}.
\]

Sending \( \varepsilon_0 \downarrow 0 \), we conclude
\[
\|u(t, \cdot) - v(t, \cdot)\|_{L^1(\mathbb{R}^d)}
\]
\[
\leq \|u_0 - v_0\|_{L^1(\mathbb{R}^d)} + 2\mathcal{E}^F_{\varepsilon_1}(\varepsilon_1) + 2\mathcal{E}^F_{\tilde{\varepsilon}_1}(\tilde{\varepsilon}_1) + 2\mathcal{E}^F_{\varepsilon_0}(\varepsilon_0) + \lim_{\varepsilon_0 \downarrow 0} \int_0^t \mathcal{E}^{f-g}(\tau; \varepsilon, \tilde{\varepsilon}) \, d\tau + \lim_{\varepsilon_0 \downarrow 0} \int_0^t \mathcal{E}^{A-B}(\tau; \varepsilon, \tilde{\varepsilon}) \, d\tau, \quad 0 \leq t \leq T.
\]

It remains to estimate the three terms on the second line.
Recall that \( u(t,x) \in I(u_0) = [\inf u_0, \sup u_0] \) for a.e. \((t,x) \in \mathbb{R}_+ \times \mathbb{R}^d\). It is easy to see that

\[
\lim_{\varepsilon_0 \to 0} \int_0^t \mathcal{E}^{f-g}(\tau; \varepsilon, \tilde{\varepsilon}) \, d\tau \\
\leq \int \|f' - g'\|_\infty \sum_{i=1}^d |\partial_{x_i} \omega_\varepsilon_1(x-y)| \omega_\varepsilon_2(\tau - s', x - y') \times \chi(\xi; u(\tau, y)) \, dy \, ds' \, dy' \, d\tau \, dx \\
\leq C \|u\|_{L^{\infty}(0,T; L^1(\mathbb{R}^d))} \frac{t \|f' - g'\|_\infty}{\varepsilon_1}.
\]

Note that, if \( u_0 \in L^1(\mathbb{R}^d) \), then \( \|u\|_{L^{\infty}(0,T; L^1(\mathbb{R}^d))} \leq \|u\|_{L^1(\mathbb{R}^d)} < \infty \) (see [31]).

On the other hand, if \( u \in L^\infty(\mathbb{R}_+; BV(\mathbb{R}^d)) \), then we can first integrate by parts in (3.14) and utilize (2.4) to obtain

\[
\lim_{\varepsilon_0 \to 0} \int_0^t \mathcal{E}^{f-g}(\tau; \varepsilon, \tilde{\varepsilon}) \, d\tau = \int \sum_{i=1}^d (f_i'(u(s,y)) - g_i'(u(s,y))) \omega_\varepsilon_1(x-y) \omega_\varepsilon_2(\tau - s', x - y') \times \partial_y u(\tau,y) \chi(u(\tau,y); v(s', y')) \, dy \, ds' \, dy' \, d\tau \, dx \\
\leq C \|u\|_{L^\infty(0,T; BV(\mathbb{R}^d))} t \|f' - g'\|_\infty.
\]

Note that, if \( u_0 \in BV(\mathbb{R}^d) \), then \( \|u\|_{L^\infty(\mathbb{R}_+; BV(\mathbb{R}^d))} \leq \|u_0\|_{BV(\mathbb{R}^d)} < \infty \) (see [31]).

Using again the standard properties of regularization kernels, we find

\[
\lim_{\varepsilon_0 \to 0} \int_0^t \mathcal{E}^{A-B}(\tau; \varepsilon, \tilde{\varepsilon}) \, d\tau \\
\leq S_\infty \sum_{i,j=1}^d \int \left| \partial_{x_i x_j} \omega_\varepsilon_1(x-y) \right| \omega_\varepsilon_2(\tau - s', x - y') \left| \chi(\xi; u(\tau,y)) \right| \, dy \, ds' \, dy' \, d\tau \, dx \, d\xi \\
= S_\infty \sum_{i,j=1}^d \int \left| \partial_{x_i x_j} \omega_\varepsilon_1(x-y) \right| |u(\tau,y)| \, dy \, d\tau \, dx \\
\leq C t S_\infty \frac{\varepsilon_1}{\varepsilon_1^2} \|u\|_{L^{\infty}(0,T; L^1(\mathbb{R}^d))}.
\]
For $u \in L^\infty(\mathbb{R}_+; BV(\mathbb{R}^d))$, then we can improve this estimate. In this case, we may integrate by parts in (3.25) and employ (2.4) to obtain

$$
\lim_{\varepsilon \to 0} \int_0^t \mathcal{E}_{A-B}(\tau; \varepsilon, \tilde{\varepsilon}) d\tau \\
\leq \sum_{i,j=1}^d \int \partial_{x_i} \omega_{x_j}(x-y) \omega_z(\tau - s', x-y') \\
\times s_{ij}(\xi) \delta(\xi - u(\tau,y)) \partial_{y_i} u(\tau,y) \chi(\xi; v(s', y')) dy ds' dy' d\tau dx d\xi \\
= -\sum_{i,j=1}^d \int \partial_{x_i} \omega_{x_j}(x-y) \omega_z(\tau - s', x-y') \\
\times s_{ij}(u(\tau,y)) \partial_{y_i} u(\tau,y) \chi(u(\tau,y); v(s', y')) dy ds' dy' d\tau dx d\xi \\
\leq S_\infty \sum_{i,j=1}^d \int |\partial_{x_i} \omega_{x_j}(x-y)| \omega_z(\tau - s', x-y') \\
\times \sup_{\tau} |\partial_{y_j} u(\tau,y)| \chi(u(\tau,y); v(s', y')) dy ds' dy' d\tau dx \\
\leq S_\infty \sum_{i,j=1}^d \int |\partial_{x_i} \omega_{x_j}(x-y)| |\partial_{y_j} u(\tau,y)| dy d\tau dx \\
\leq C \|u\|_{L^\infty(0,T; BV(\mathbb{R}^d))} \frac{tS_\infty}{\varepsilon_1}.
$$

Finally, we estimate

$$
\int_0^t \mathcal{E}_v(\tau; \tilde{\varepsilon}) d\tau \leq \int \omega_z(\tau - s, x-y) \\
\times \left( |\partial_{x_0} \omega_z(\tau - s', x-y')| \sup_{\xi} |e_0(\xi, s', y')| + \sum_{J=(J_1, \ldots, J_d) \geq 0} |D_{\xi} \omega_z(\tau - s', x-y')| \sup_{\xi} |e_{1,J}(\xi, s', y')| \right) ds dy ds' dy' d\tau dx \\
\leq \frac{C}{\varepsilon_0} \left\| \sup_{\xi} |e_0(\xi, \cdot, \cdot)| \right\|_{L^1(0,t+\tilde{\varepsilon}_0; L^1(\mathbb{R}^d))} \\
+ C \sum_{J=(J_1, \ldots, J_d) \geq 0} \frac{1}{\varepsilon_1^{|J|}} \left\| \sup_{\xi} |e_{1,J}(\xi, \cdot, \cdot)| \right\|_{L^1(0,t+\tilde{\varepsilon}_0; L^1(\mathbb{R}^d))}.
$$

This concludes the proof of (3.3). If $g \equiv f$ and $B \equiv A$, then $\mathcal{E}_{f-g}^{A-B} \equiv 0$. Consequently, we can let $\varepsilon_1 \downarrow 0$, and hence $\mathcal{E}_{u,t}^A \to 0$ in (3.3). This completes the proof of (3.4).
4. Estimates for Continuous Dependence on the Nonlinearities

We now apply the $L^1$-framework developed in Section 3 to derive an explicit estimate for continuous dependence on the nonlinearities in quasilinear degenerate parabolic equations with anisotropic diffusion. Consider the problem

\begin{equation}
\partial_t v + \text{div} g(v) = \nabla \cdot (B(v) \nabla v), \quad v(0,x) = v_0(x),
\end{equation}

where $g, B$, and $v_0$ satisfy the conditions stated in Section 1. The kinetic formulation of (4.1) is (3.1) with $E \equiv 0$. For simplicity of presentation, we assume that $u_0 \in BV(\mathbb{R}^d)$.

We can now apply Theorem 3.1 with $E = \mathbb{R}$, $v_0, \xi, \tilde{\xi} \downarrow 0$ (so that $E_{\xi,\tilde{\xi}}, E_{\xi,\tilde{\xi}} \to 0$), and $E_{\xi,\tilde{\xi}}(\varepsilon_1) \leq \varepsilon_1 \|u_0\|_{BV(\mathbb{R}^d)}$. Hence, for any $t > 0$ and any $\varepsilon_1 > 0$,

$$
\|u(t,\cdot) - v(t,\cdot)\|_{L^1(\mathbb{R}^d)} \leq \|u_0 - v_0\|_{L^1(\mathbb{R}^d)} + C\|u_0\|_{BV(\mathbb{R}^d)} \left( \varepsilon_1 + t \|f' - g'\|_{\infty} + \frac{tS_{\infty}}{\varepsilon_1} \right).
$$

Choosing the optimal $\varepsilon_1$, we end up with the following theorem.

**Theorem 4.1** (Continuous Dependence Estimate). Suppose $u_0 \in BV(\mathbb{R}^d)$. Let $u$ be an entropy solution of (1.1) with (1.2)–(1.4). Let $v \in C(\mathbb{R}^+; L^1(\mathbb{R}^d))$ be an entropy solution of (4.1) with (1.6)–(1.8). Then, for any $t > 0$,

\begin{equation}
\|u(t,\cdot) - v(t,\cdot)\|_{L^1(\mathbb{R}^d)} \leq \|u_0 - v_0\|_{L^1(\mathbb{R}^d)} + C\|u_0\|_{BV(\mathbb{R}^d)} \left( t \|f' - g'\|_{\infty} + \sqrt{t} \left\| \left( \sqrt{A} - \sqrt{B} \right) \left( \sqrt{A} - \sqrt{B} \right)^T \right\|_{\infty} \right).
\end{equation}

**Remark 4.1.** Note that Theorem 4.1 holds without assumption $v \in L^\infty(\mathbb{R}^+; BV(\mathbb{R}^d))$. Also, observe that, in the isotropic case: $A = \text{diag}(a_1, \ldots, a_d)$ and $B = \text{diag}(b_1, \ldots, b_d)$,

$$
\left\| \left( \sqrt{A} - \sqrt{B} \right) \left( \sqrt{A} - \sqrt{B} \right)^T \right\|_{\infty} = \sup_{\xi \in I(u_0)} \left| \sqrt{a_i(\xi)} - \sqrt{b_i(\xi)} \right|,
$$

We remark that, in the estimate (4.2) for continuous dependence, the strong norms for $f - g$ can be replaced by weaker norms in the spirit of [3], although we do not pursue this here.

5. Error Estimates for the Vanishing Anisotropic Viscosity Approximation

We now consider the anisotropic viscous problem (1.10) with $f, A, B$, and $v_0$ satisfying (1.3), (1.4), (1.6), and (1.8), respectively. Suppose $B(v) > 0$ (i.e, (1.10) is uniformly parabolic which admits a unique classical solution) and

$$
B_\infty := \|B\|_{\infty} = \sup_{\xi \in I(u_0)} b_{ij}(\xi) < \infty.
$$

We are interested in applying the $L^1$-framework established in Section 3 to derive an explicit error estimate for $u - v$ as $\mu \downarrow 0$, where $u$ is an entropy solution of the original problem (1.1). As in the previous section, we assume that $u_0 \in BV(\mathbb{R}^d)$. 
Let \((\eta, q, R)\) be an entropy-entropy flux triple. Multiplying (1.10) by \(\eta'(v)\), we recover the usual dissipation structure

\[
\partial_t \eta(v) + \text{div}(v) = \nabla \cdot (R'(v)\nabla v) - \eta''(v) \sum_{i,j=1}^d a_{ij}(v) \partial_{x_i} v \partial_{x_j} v
\]  
(5.1)

\[+ \mu \nabla \cdot (\eta'(v)B(v)\nabla v) - \mu \eta''(v) \sum_{i,j=1}^d b_{ij}(v) \partial_{x_i} v \partial_{x_j} v.\]

We identify the entropy defect measure \(m^v(\xi, t, x)\) as

\[m^v(\xi, t, x) = \delta(\xi - v(t, x)) \sum_{k=1}^K \left( \sum_{i=1}^d \partial_{x_i} \zeta^R_k(v) \right)^2,
\]

and also the entropy dissipation measure \(m^{v, \psi}(t, x)\) as

\[m^{v, \psi}(t, x) = \int_{\mathbb{R}} m^v(\xi, t, x) \psi(\xi) \, d\xi, \quad \psi \in C_0(\mathbb{R}),\]

via the duality \((C_0(\mathbb{R}); \mathcal{M}(\mathbb{R}))\).

The parabolic defect measure \(n^v(\xi, t, x)\) is identified as

\[n^v(\xi, t, x) = \delta(\xi - v(t, x)) \sum_{k=1}^K \left( \sum_{i=1}^d \partial_{x_i} \zeta^A_k(v) \right)^2 \geq 0,
\]

and again, via the duality, the parabolic dissipation measure \(n^{v, \psi}(t, x)\) as

\[n^{v, \psi}(t, x) = \int_{\mathbb{R}} n^v(\xi, t, x) \psi(\xi) \, d\xi, \quad \psi \in C_0(\mathbb{R}).\]

Hence we can write (5.1) as

\[
\partial_t \eta(v) + \text{div}(v) - \nabla \cdot (R'(u)\nabla v) = - \left( \mu m^{v, \eta'''} + n^{v, \eta''} \right) + \mu \nabla \cdot (\eta'(v)B(v)\nabla v).
\]  
(5.2)

We can now transform the dissipation structure (5.2) via the duality into the kinetic structure [9]:

\[
\partial_t \chi(\xi; v) + f'(\xi) \cdot \nabla_x \chi(\xi; v)
\]  
(5.3)

\[= \sum_{i,j=1}^d a_{ij}(\xi) \partial_{x_i x_j}^2 \chi(\xi; v) + \partial_\xi (m^{v} + n^{v}) + \mu \sum_{i,j=1}^d b_{ij}(\xi) \partial_{x_i x_j}^2 \chi(\xi; v).
\]

We first assume that \(v_0 \in BV(\mathbb{R}^d)\). Then we can write

\[
\mu \sum_{i,j=1}^d b_{ij}(\xi) \partial_{x_i x_j}^2 \chi(\xi; v) = \partial_\xi \left( \mu \sum_{j=1}^d \partial_{x_j} \left( \sum_{i=1}^d \partial_{x_i} \left( \int_{\xi}^v b_{ij}(\eta) \chi(\eta; v) \, d\eta \right) \right) \right) =: \partial_\xi E(\xi, t, x),
\]

where

\[E(\xi, t, x) = \sum_{j=1}^d \partial_{x_j} e_1^v(\xi, t, x), \quad e_1(\xi, t, x) = \mu \sum_{i=1}^d \partial_{x_i} \left( \int_{\xi}^v b_{ij}(\eta) \chi(\eta; v) \, d\eta \right).
\]
From (2.4), it is clear that, for any $\xi$,$$
abla \left| e_1^i(\xi, t, x) \right| \leq \mu \sum_{i,j=1}^d b_{ij}(v) \left| \partial_{x_i} v \right| \leq \mu\beta_\infty \sum_{i=1}^d \left| \partial_{x_i} v \right| \leq \mu\beta_\infty \| v_0 \|_{BV(\mathbb{R}^d)}.$$Hence$$\left\| \sup_{\xi} \left| e_1^i(\xi, t, x) \right| \right\|_{L^1(0, t; L^1(\mathbb{R}^d))} \leq \mu\beta_\infty \| v_0 \|_{BV(\mathbb{R}^d)}.$$Applying Theorem 3.1 with $\varepsilon_0, \varepsilon_1, \varepsilon_0 \downarrow 0$ (so that $\mathcal{E}_{u,t}, \mathcal{E}_{u,t}, \mathcal{E}_{v,t} \to 0$), and$$\mathcal{E}_{\varepsilon_4}(\varepsilon_1) \leq \varepsilon_1 \| v_0 \|_{BV(\mathbb{R}^d)},$$we obtain that, for any $t > 0$ and $\varepsilon_1 > 0$,$$\| u(t, \cdot) - v(t, \cdot) \|_{L^1(\mathbb{R}^d)} \leq \| u_0 - v_0 \|_{L^1(\mathbb{R}^d)} + C \| v_0 \|_{BV(\mathbb{R}^d)} \left( \varepsilon_1 + \frac{t\mu}{\varepsilon_1} \right).$$Choosing the optimal $\varepsilon_1$, we get a rate of convergence in $\mu \downarrow 0$ that is of order $\sqrt{t\mu}$.

If $v_0 \notin BV(\mathbb{R}^d)$, then we must write the error term as$$\mu \sum_{i,j=1}^d b_{ij}(\xi) \partial_{x_i} x_j \chi(\xi; v) = \partial_{x} E(\xi, t, x),$$where$$E(\xi, t, x) = \sum_{i,j=1}^d \partial_{x_i} x_j \xi e_1^i(\xi, t, x), \quad e_1^i(\xi, t, x) = \mu \int_{\xi} b_{ij}(\eta) \chi(\eta; v) d\eta.$$Clearly, for any $\xi$,$$
abla \left| e_1^i(\xi, t, x) \right| \leq \beta_\infty \mu \| v(t, x) \|, \quad i, j = 1, \ldots, d.$$Hence$$\left\| \sup_{\xi} \left| e_1^i(\xi, t, x) \right| \right\|_{L^1(0, t; L^1(\mathbb{R}^d))} \leq t\beta_\infty \mu \| v_0 \|_{L^1(\mathbb{R}^d)},$$so that, for any $t > 0$ and $\varepsilon_1 > 0$,$$\| u(t, \cdot) - v(t, \cdot) \|_{L^1(\mathbb{R}^d)} \leq \| u_0 - v_0 \|_{L^1(\mathbb{R}^d)} + C \| v_0 \|_{L^1(\mathbb{R}^d)} \left( \varepsilon_1 + \frac{t\mu}{\varepsilon_1} \right).$$Choosing the optimal $\varepsilon_1$, we get an rate of convergence in $\mu \downarrow 0$ that is of order $(t\mu)^{\frac{1}{2}}$.

Observe that we do not get an optimal convergence rate when $v_0 \notin BV(\mathbb{R}^d)$ in this way. However, as already mentioned in Section 1, one of our observations is that, by interpreting the desired error estimate as a continuous dependence estimate, we can obtain the optimal result. Indeed, in the present context, Theorem 4.1 gives$$\| u(t, \cdot) - v(t, \cdot) \|_{L^1(\mathbb{R}^d)} \leq \| u_0 - v_0 \|_{L^1(\mathbb{R}^d)} + C \| v_0 \|_{BV(\mathbb{R}^d)} \sqrt{t} \left\| \left( \sqrt{A} - \sqrt{\mu B} \right) \left( \sqrt{A} - \sqrt{\mu B} \right) \right\|_{\infty}, \quad t > 0.$$Now a simple calculation reveals$$\left\| \left( \sqrt{A} - \sqrt{\mu B} \right) \left( \sqrt{A} - \sqrt{\mu B} \right) \right\|_{\infty} \leq C\mu.$
We summarize the discussion in this section in the following theorem.

**Theorem 5.1** (Vanishing Anisotropic Viscosity). Suppose \( u_0 \in BV(\mathbb{R}^d) \), and let \( u \) be an entropy solution of (1.1) with (1.2)–(1.4). Let \( v \in C(\mathbb{R}_+; L^1(\mathbb{R}^d)) \) be an entropy solution of (1.10) with (1.3), (1.6), and (1.8). Then, for any \( t > 0 \),
\[
\|u(t, \cdot) - v(t, \cdot)\|_{L^1(\mathbb{R}^d)} \leq \|u_0 - v_0\|_{L^1(\mathbb{R}^d)} + C\|u_0\|_{BV(\mathbb{R}^d)} \sqrt{t} \mu.
\]

**Remark 5.1.** As indicated in Section 1, after our main results were finished, we noticed a preprint by Makridakis and Perthame [23]. From Theorem 5.1, we can recover an “x independent version” of their result (the \( x \) dependent version can be proved along the same lines). Let \( u \) be an entropy solution to the scalar conservation law
\[
\frac{\partial t}{\partial t} u + \text{div}(u) = 0, \quad u(0, x) = u_0(x),
\]
with \( BV \) initial data \( u_0 \). Let \( v \) be an entropy solution to the anisotropic viscous problem
\[
\frac{\partial v}{\partial t} + \text{div}(v) = \mu \nabla \cdot (B(v) \nabla v), \quad v(0, x) = v_0(x),
\]
where \( v_0 \) is only in \( L^1 \), that is, there is no \( BV \) bound available on the approximate solution \( v \). As above, we suppose that \( B(v) \) satisfies (1.8) and also \( B(v) > 0 \) (i.e, (5.5) is uniformly parabolic which admits a unique classical solution). Setting \( \Lambda \equiv 0 \) in Theorem 5.1, we deduce
\[
\|u(t, \cdot) - v(t, \cdot)\|_{L^1(\mathbb{R}^d)} \leq \|u_0 - v_0\|_{L^1(\mathbb{R}^d)} + C\|u_0\|_{BV(\mathbb{R}^d)} \sqrt{t} \mu, \quad t > 0.
\]

### 6. Error Estimates for a Finite Difference Approximation

In this section, as an example of direct applications of the \( L^1 \)–framework to make error estimates for numerical methods, we focus on a linear convection-diffusion model equation:
\[
\frac{\partial_t}{\partial t} u + \text{div}(Vu) = \nabla \cdot (A \nabla u), \quad u(0, x) = u_0(x),
\]
for some constant velocity vector \( V = [V_1, \ldots, V_d] > 0 \) and some small constant diffusion matrix \( A = \text{diag}(a_1, \ldots, a_d) \geq 0 \). We assume that \( u_0 \in BV(\mathbb{R}^d) \).

Fix a time step size \( \Delta t > 0 \) and a spatial step size \( \Delta x > 0 \). We use \( \partial_{t, \Delta t} \) for the temporal difference operator:
\[
\partial_{t, \Delta t} v(t, x) = \frac{v(t + \Delta t, x) - v(t, x)}{\Delta t},
\]
\( \partial_{x_i, \Delta x} \) for the first order spatial difference operator in the direction \( x_i \):
\[
\partial_{x_i, \Delta x} v(t, x) = \frac{v(t, x) - v(t, x - \Delta x e_i)}{\Delta x}, \quad i = 1, \ldots, d,
\]
where \( e_i \) denotes the \( i \)th unit vector in \( \mathbb{R}^d \); and \( \partial_{x_i, \Delta x}^2 \) for the second order spatial central difference operator in the direction \( x_i \):
\[
\partial_{x_i, \Delta x}^2 v(t, x) = \frac{v(t, x - \Delta x e_i) - 2v(t, x) + v(t, x + \Delta x e_i)}{(\Delta x)^2}, \quad i = 1, \ldots, d.
\]
We consider the explicit upwind-central finite difference scheme:
\[
\partial_{t, \Delta t} v(t, x) + \sum_{i=1}^d V_i \partial_{x_i, \Delta x} v(t, x) = \sum_{i=1}^d a_i \partial_{x_i, \Delta x}^2 v(t, x), \quad v(0, x) = u_0(x).
\]
As usual, to ensure the stability, it is necessary to require the CFL condition:
\[
V_\infty d \frac{\Delta t}{\Delta x} + 2A_\infty d \frac{\Delta t}{(\Delta x)^2} \leq 1, \quad V_\infty := \max_{i=1,\ldots,d} V_i, \quad A_\infty := \max_{i=1,\ldots,d} a_i.
\]
From the CFL condition (6.3) and $u_0 \in BV$, it follows in a standard fashion (see, e.g., [13, 17]) that $\|v(t, \cdot)\|_{L^\infty(\mathbb{R}^d)} \leq \|u_0\|_{L^\infty(\mathbb{R}^d)}$ and $\|v(t, \cdot)\|_{BV(\mathbb{R}^d)} \leq \|u_0\|_{BV(\mathbb{R}^d)}$ for all $t > 0$, and $\|v(t_2, \cdot) - v(t_1, \cdot)\|_{L^1(\mathbb{R}^d)} \leq C\sqrt{|t_2 - t_1|}$ for all $t_1, t_2 > 0$.

Our goal is to derive an $L^1$ error estimate that is uniform with respect to small diffusion matrix $A$. For technical reasons, we are not going to work directly with $v$, but instead with a regularized version $v_\rho$ defined by

$$v_\rho(t, x) = \left( \omega_\rho * v \right)(t, x), \quad \rho = (\rho_0, \rho_1) > 0,$$

where $\omega_\rho = \omega_\rho(t, x)$ is a standard regularization kernel of the type used in the proof of Theorem 3.1 with the smoothing radius $\rho_0$ in $t$ and $\rho_1$ in $x$.

Evaluating (6.2) at $(t - s, x - y)$, then multiplying by $\omega_\rho(s, y)$, and finally integrating the result over $(s, y)$, we obtain that the “smooth” function $v_\rho$ satisfies the finite difference equation:

$$\partial_{t, \Delta t} v_\rho(t, x) + \sum_{i=1}^d V_i \partial_{x_i, \Delta x} v_\rho(t, x) = \sum_{i=1}^d a_i \partial_{x_i, \Delta x}^2 v_\rho(t, x),$$

with initial data

$$v_\rho(0, x) = \left( \omega_\rho * u_0 \right)(x).$$

Clearly, the approximate solution $v_\rho$ satisfies the following a priori estimates, uniform in $\Delta t, \Delta x, \rho$:

$$\begin{cases}
\|v_\rho(t, \cdot)\|_{L^\infty(\mathbb{R}^d)} \leq \|u_0\|_{L^\infty(\mathbb{R}^d)}, \\
\|v_\rho(t, \cdot)\|_{BV(\mathbb{R}^d)} \leq \|u_0\|_{BV(\mathbb{R}^d)}, \\
\|v_\rho(t_2, \cdot) - v_\rho(t_1, \cdot)\|_{L^1(\mathbb{R}^d)} \leq C\sqrt{|t_2 - t_1|}, \text{ for any } t_1, t_2 > 0.
\end{cases}$$

(6.5)

A Taylor expansion yields

$$\partial_{t, \Delta t} v_\rho(t, x) = \partial_t v_\rho(t, x) + \mathcal{E}_0(t, x),$$

with

$$\mathcal{E}_0(t, x) = \frac{1}{2} \int_0^{\Delta t} \partial^2_{t, 2} v_\rho(t + z, x) \, dz.$$ 

Observe that, for any $t > 0$, $\|\mathcal{E}_0\|_{L^1(0, t; L^1(\mathbb{R}^d))} \leq C\frac{\Delta t}{\rho_0^2}$, where the third part of (6.5) was used.

Similarly,

$$\sum_{i=1}^d V_i \partial_{x_i, \Delta x} v_\rho(t, x) = \sum_{i=1}^d V_i \partial_{x_i} v_\rho(t, x) + \mathcal{E}_1(t, x),$$

with

$$\mathcal{E}_1(t, x) = \frac{1}{2} \int_{-\Delta x}^{0} \partial^2_{x, 2} v_\rho(t, x + ze_i) \, dz,$$

and, for any $t > 0$, $\|\mathcal{E}_1\|_{L^1(0, t; L^1(\mathbb{R}^d))} \leq C\frac{\Delta x}{\rho_1}$, in which the second part of (6.5) has been used.
Finally, 
\[ \sum_{i=1}^{d} a_i \partial_{x_i}^2 v_\rho(t, x) = \sum_{i=1}^{d} a_i \partial_{x_i}^2 v_\rho(t, x) + E_2(t, x), \]
with 
\[ E_2(t, x) = \sum_{i=1}^{d} \frac{a_i \Delta x}{24} \int_{\Delta x}^{\partial_{x_i}^4 v_\rho(t, x + z e_i) \, dz}, \]
and, for any \( t > 0 \), \( \| E_2 \|_{L^1(0, t; L^1(\mathbb{R}^d))} \leq C_4 \Delta x^2 \rho^2 \), in which we have used (6.5) again.

Hence, from (6.4), it follows that \( v_\rho \) satisfies the “approximate” convection-diffusion equation
\[ \partial_t v_\rho + \text{div}(V v_\rho) = \nabla \cdot (A \nabla v_\rho) + E(t, x), \]
where \( E(t, x) := E_0(t, x) + E_1(t, x) + E_2(t, x) \), which suggests that we may apply Theorem 3.1 with \( J_0 = J_1 = 0 \) to estimate \( u - v_\rho \).

Let \( \eta : \mathbb{R} \to \mathbb{R} \) be an entropy function. Multiplying (6.6) by \( \eta'(v_\rho) \), we obtain the usual dissipation structure
\[ \partial_t \eta(v_\rho) + \text{div}(V \eta(v_\rho)) - \nabla \cdot (A \nabla \eta(v_\rho)) = -\eta''(v_\rho) \left( \sum_{i=1}^{d} \sqrt{a_i} \partial_{x_i} v_\rho \right)^2 + \eta'(v_\rho) E(t, x). \]
As usual, we can transform this dissipation structure via the duality into the kinetic structure:
\[ \partial_t \chi(\xi; v_\rho) + V \cdot \nabla \chi(\xi; v_\rho) = \sum_{i=1}^{d} a_i \partial_{x_i}^2 \chi(\xi; v_\rho) + \partial_\xi (m^v + n^v + E)(\xi, t, x), \]
where \( m^v \equiv 0 \) and
\[ n^v = \delta(\xi - v_\rho) \left( \sum_{i=1}^{d} \sqrt{a_i} \partial_{x_i} v_\rho \right)^2, \quad E(\xi, t, x) = 1_{\xi > v_\rho} E(t, x), \]
so that \( \partial_\xi E = \delta(\xi - v_\rho) E. \)

Observe that, for any \( t > 0 \),
\[ \left\| \sup_{\xi} \left| E(\xi, \cdot) \right| \right\|_{L^1(0, t; L^1(\mathbb{R}^d))} \leq C t \left( \frac{\Delta t}{\rho^{3/2}} + \frac{\Delta x}{\rho} + \frac{\Delta x^2}{\rho^3} \right). \]

Hence, using (3.4) in Theorem 3.1 with \( J_0 = J_1 = 0 \) (after having sent \( \tilde{\xi}_0, \tilde{\xi}_1 \downarrow 0 \)), we get
\[ \| u(t, \cdot) - v_\rho(t, \cdot) \|_{L^1(\mathbb{R}^d)} \leq \| u_0 - v_\rho(0, \cdot) \|_{L^1(\mathbb{R}^d)} + C t \left( \frac{\Delta t}{\rho^{3/2}} + \frac{\Delta x}{\rho_1} + \frac{\Delta x^2}{\rho^3} \right), \]
for any \( t > 0 \) and \( \rho > 0 \). From (6.5), we have
\[ \| u_0 - v_\rho(0, \cdot) \|_{L^1(\mathbb{R}^d)}, \| v(t, \cdot) - v_\rho(t, \cdot) \|_{L^1(\mathbb{R}^d)} \leq C (\sqrt{\rho_0} + \rho_1). \]
Hence
\[ \| u(t, \cdot) - v(t, \cdot) \|_{L^1(\mathbb{R}^d)} \leq C \left( \sqrt{\rho_0} + \rho_1 + \frac{t \Delta t}{\rho^{3/2}} + \frac{t \Delta x}{\rho_1} + \frac{t \Delta x^2}{\rho^3} \right), \quad t > 0. \]
Choosing the optimal $\rho$, we get the following theorem regarding the convergence rate for the upwind-central finite difference scheme.

**Theorem 6.1.** Let $u$ be an entropy solution of (6.1) with (1.2) and $u_0 \in BV(\mathbb{R}^d)$. Let $v = v(t,x)$ be the upwind-central finite difference solution generated by (6.2) with (6.3). Then, for any $t > 0$,

$$\|u(t, \cdot) - v(t, \cdot)\|_{L^1(\mathbb{R}^d)} \leq C \left( (t\Delta t)^{\frac{1}{2}} + \sqrt{t\Delta x} \right) \leq C\sqrt{t\Delta x}.$$

**Remark 6.1.** Note that the $L^1$ error estimate in Theorem 6.1 is robust with respect to sending the diffusion matrix $A$ to zero. We emphasize that, although the problem (6.1) under consideration is linear, our $L^1$ method of analysis is still very much nonlinear! We will develop further our approach to analyze and derive $L^1$ error estimates for monotone finite difference schemes for nonlinear degenerate parabolic equations elsewhere.

**Acknowledgments**

The research of Gui-Qiang Chen was supported in part by the National Science Foundation under Grants DMS-0204225, DMS-0204455, and INT-9987378. The research of Kenneth H. Karlsen was supported in part by the BeMatA program of the Research Council of Norway and the European network HYKE, funded by the EC as contract HPRN-CT-2002-00282.

**References**


