PERGAMON

Transportation Research Part B 36 (2002) 275–290

TRANSPORTATION RESEARCH PART B

www.elsevier.com/locate/trb

A non-equilibrium traffic model devoid of gas-like behavior

H.M. Zhang *

Department of Civil and Environmental Engineering, 156 Everson Hall, University of California at Davis, Davis, CA 95616, USA

Received 20 December 1999; received in revised form 3 September 2000; accepted 12 September 2000

Abstract

The validity of existing higher-order continuum models has been seriously challenged in recent literature. A primary objection to these models is that they exhibit gas-like behavior, that is, they predict that vehicles from behind can influence the behavior of vehicles in front, as elucidated by Daganzo [Transp. Res. B 29 (1995) 277]. Considerable efforts devoted to remove this deficiency (del Castillo et al. [In: C.-F. Daganzo (Ed.), Transportation and Traffic Theory, Elsevier Science, Amsterdam, 1993, p. 387], Zhang [Transp. Res. B 32(7) (1998) 485; Transp. Res. B 33(6) (1999) 387; Transp. Res. B 34 (2000) 583]) have met with only partial success. In this paper we present a non-equilibrium traffic model shown to be devoid of the gas-like behavior that plagues other higher-order models. The system of partial differential equations that describes this model is hyperbolic and has two characteristic fields: one is genuinely nonlinear and the other is linearly degenerate. The first field gives rise to shock and rarefaction waves that are similar to those of the Lighthill–Whitham–Richards (LWR) model, while the second field produces contact discontinuities. All these waves travel no faster than traffic; thus the trajectory of a vehicle cannot be influenced by what happens behind it. The model is also shown to exhibit correct queue-end behavior and is able to explain some of the observed traffic phenomena that challenge old models.

1. Introduction

The development of higher-order, or more precisely, non-equilibrium continuum traffic models, as exemplified by the Payne–Whitham (PW for short) model (Payne, 1971; Whitham, 1974), was largely motivated by certain deficiencies of the first-order model, developed independently by Lighthill and Whitham (1955), and Richards (1956) (LWR for short). The LWR model, which is a scalar conservation law of the form

* Tel.: +1-530-754-9203; fax: +1-530-752-7872.
E-mail address: hmzhang@ucdavis.edu (H.M. Zhang).

0191-2615/01/$ - see front matter
PII: S0191-2615(00)00050-3

$$\rho_t + (\rho V_*(\rho))_x = 0, \tag{1}$$

is an elegant theory. It is simple yet sufficiently powerful to describe the most basic traffic flow phenomena – traffic congestion formation and dissipation in dense traffic. The essence of this model is its assumption that travel speed (v) and vehicle concentration (ρ) (both v and ρ are functions of space (x) and time (t)) observe a unique, strictly decreasing relation $v = V_*(\rho)$ when traffic is in *equilibrium*. Moreover, it is also customary to require that the equilibrium flux function $f_*(\rho) = \rho V_*(\rho)$ be strictly concave (i.e., $f_*''(\rho) < 0$), such that only compression waves exist in this model.

The nature of the solutions of the LWR model is well understood. Small disturbances are propagated at the characteristic speed $f_*'(\rho)$, which is always less than $V_*(\rho)$ due to the decreasing nature of $V_*(\rho)$. This implies that traffic information travels from leading vehicles to following vehicles, not vice versa. That is, the LWR model is *anisotropic*.

Moreover, when traffic concentration is heavy downstream and light upstream, a shock wave forms and travels at a speed of

$$\sigma_* = \frac{[f_*(\rho)]}{[\rho]} \equiv \frac{f_*(\rho_\mathrm{l}) - f_*(\rho_\mathrm{r})}{\rho_\mathrm{l} - \rho_\mathrm{r}}, \tag{2}$$

where ρ_l and ρ_r are traffic concentrations upstream and downstream of the shock, respectively. When traffic concentration is light downstream and heavy upstream, a rarefaction wave arise and the heavy traffic will disperse, just as one would observe in a situation where queued traffic in front of a traffic light disperses when the light turns green.

Although the LWR model makes sensible predictions of propagation and dissipation of traffic jams, it fails to model two important traffic phenomena: stop–start waves caused by instability in traffic flow and forward propagation of disturbances in heavy traffic (Edie and Bavarez, 1967). Various early observations (e.g., Treiterer and Myers, 1974; Maes, 1979) indicate that (ρ, v) does not observe the equilibrium curve $(\rho, V_*(\rho))$ when traffic is not in equilibrium. Rather, acceleration flows and deceleration flows follow distinctively different paths in the (ρ, v) phase plane, and these paths usually form one or more hysteresis loops. Newell (1965) exploited these features and extended the LWR model to a model that has two speed-spacing curves (one for acceleration and one for deceleration). Newell's model retained the basic conservation structure of the LWR model and was able to explain instabilities observed in tunnel traffic. Few other researchers, however, pick up from where Newell left and further refine and operationalize this model. [1] Rather, most of the researchers who sought to improve the LWR model took a different approach: developing a speed evolution equation that allows non-equilibrium states (ρ, v) to deviate from the equilibrium $(\rho, V_*(\rho))$ curve. The most well known results of these efforts is the PW equation

$$v_t + v v_x = \frac{V_*(\rho) - v}{\tau} - \frac{c_0^2}{\rho}\rho_x, \tag{3}$$

[1] I suspect that this may be partially due to the difficulty in obtaining precise formulas for the acceleration and deceleration phase curves. Experimental evidence appears to discount the uniqueness of such curves.

where τ is relaxation time and $c_0 < 0$ is the so-called traffic 'sound' speed. [2] The PW momentum equation is typical of higher-order models in that it has a relaxation term that contains the equilibrium speed–concentration relation and an anticipation term that contains the space gradient of vehicle concentration.

The PW non-equilibrium model, consisting of the momentum equation and the conservation of flow, became a popular choice of traffic simulation in the seventies and eighties. However, the mathematical properties of this model were not fully understood at that time. As a result, early applications of the PW model and its variants generated quite mixed results – some researchers reported excellent fit with observed data using this model (e.g., Papageorgiou et al., 1989), while others claimed that the PW model failed to model traffic bottlenecks properly (e.g., Hauer and Hurdle, 1979). Some of the poor results might have been due to the inadequate numerical schemes that were used to convert the system of partial differential equations to a set of difference equations. Leo and Pretty (1992) used a correct numerical scheme to discretize the PW equations and were able to show that the PW model did produce correct predictions of bottleneck flow. But they also found that it did not perform better than the LWR model. [3] The misuse of numerical procedures, however, explains only part of the problem. A more serious problem, as pointed out by Daganzo (1995) and del Castillo et al. (1993), is that information in the PW model can travel faster than vehicles. This means that vehicles from behind can influence vehicles in front, which violates the anisotropic property of traffic flow. Moreover, the PW model is found to be diffusive. Because diffusion travels both ways, the PW model can produce negative travel speeds (i.e., 'wrong-way travel') under certain circumstances.

In an effort to eradicate the 'wrong-way travel' problem in the PW model, Zhang (1998) proposed an improved non-equilibrium model that differs from the PW model essentially by only one term. It replaces the traffic sound speed c_0 in the PW momentum equation by a concentration-dependent traffic sound speed: $c(\rho) = \rho V_*'(\rho)$. His analyses of the improved non-equilibrium model indicate that it behaves essentially as the LWR model near equilibrium conditions, does not produce 'wrong-way travel', and its waves are linearly and nonlinearly stable (Zhang, 1999, 2000a). This model does not, however, completely suppress the gas-like behavior, although it was shown that the effect of such behavior decays exponentially fast (at a rate of $\exp(-t/\tau)$) in the improved non-equilibrium model (Zhang, 2000a).

That Zhang's non-equilibrium model does not completely suppress the gas-like behavior should not surprise us, because the two models (PW and Zhang) have essentially the same mathematical structure. In fact, the momentum equations of both models can be expressed in a general form

$$v_t + v v_x + \frac{c^2(\rho)}{\rho}\rho_x = \frac{V_*(\rho) - v}{\tau} \tag{4}$$

[2] There are other versions of the PW model. We choose this one for its stronger resemblance to Zhang's model and for an easy interpretation of traffic sound speed.

[3] This is not surprising because, for sufficiently small relaxation times, the PW model behaves much like the LWR model. Leo and Pretty used a relaxation time of 0.75–1 s, which makes the relaxation effect dominate the anticipation effect in the simulation.

Table 1
Various traffic flow models

Model	U	$A(U)$	$S(U)$
LWR	(ρ)	$(f'_*(\rho))$	(0)
Payne–Whitham	(ρ, v)	$\begin{pmatrix} v & \rho \\ c_0^2/\rho & v \end{pmatrix}$	$(0, V_*(\rho) - v/\tau)$
Zhang	(ρ, v)	$\begin{pmatrix} v & \rho \\ \rho(V'_*(\rho))^2 & v \end{pmatrix}$	$(0, V_*(\rho) - v/\tau)$
PW-like	(ρ, v)	$\begin{pmatrix} v & \rho \\ c^2(\rho)/\rho & v \end{pmatrix}$	$(0, V_*(\rho) - v/\tau)$

if we introduce a concentration-dependent traffic sound speed $c(\rho) : c(\rho) = c_0$ for the PW model and $c(\rho) = \rho V'_*(\rho)$ for Zhang's model. We shall name models that consist of a momentum equation in the form of (4) and the conservation equation

$$\rho_t + (\rho v)_x = 0 \tag{5}$$

PW-like models. (Table 1 lists all the traffic flow models discussed in this section.)

To understand why PW-like models behave isotropically we first express the two equations in a vector form

$$U_t + A(U)U_x = s(U), \tag{6}$$

where

$$U = (\rho, v)^{\mathrm{T}}, \quad A(U) = \begin{pmatrix} v & \rho \\ c^2(\rho)/\rho & v \end{pmatrix}, \quad s(U) = (0, V_*(\rho) - v/\tau)^{\mathrm{T}}.$$

It turns out that the properties of this system are largely controlled by the eigenvalues of the flux Jacobian matrix A (it is called a flux Jacobian because when (6) is written in a conservative form $U_t + F(U)_x = s(U)$, A is the Jacobian matrix of the vector flux $F(U)$). These eigenvalues, also known as characteristic speeds, determine how traffic disturbances are propagated in a traffic stream. It turns out that a PW-like model has two real and distinct eigenvalues λ_1 and λ_2: $\lambda_1 = v + c(\rho) < v - c(\rho) = \lambda_2$.

When this happens, the model is said to be strictly hyperbolic. The consequences of hyperbolicity are that (1) information (such as slow-downs and speed-ups in traffic) in a hyperbolic system travels at finite speeds, and (2) discontinuities (or shocks) in the state variables U arise in the solution under certain conditions. Note that in a PW-like model the second characteristic, λ_2, always travels faster than traffic because $c(\rho) < 0$. As a result, waves associated with the second characteristic, be they expansion waves or shock waves, always reach vehicles from behind, either slowing traffic down or speeding traffic up (Zhang, 2000a). This is not what one would expect from real traffic, at least when it is on a single-lane highway. [4]

[4] Anisotropic property may not hold in multi-lane traffic. Zhang (2000c) showed how isotropic behavior might arise from multi-lane traffic wherein each lane's traffic behaves anisotropically.

It is clear that characteristic speeds, when exceeding vehicle speeds, lead to gas-like traffic behavior. A non-equilibrium model devoid of such behavior must therefore suppress its characteristic speeds so that they do not exceed vehicle speeds. The characteristics of both PW and the improved non-equilibrium models are the eigenvalues of the flux Jacobian $A(U)$. Upon observing this Jacobian matrix, it is not difficult to see that a possible way to eliminate characteristics that travel faster than vehicles is to replace $c^2(\rho)/\rho$ by zero in the flux Jacobian, that is, to remove the dependency of traffic acceleration on the spatial concentration gradient in the momentum equation. This is precisely what we will do in developing an anisotropic non-equilibrium traffic model, which is the focus of the remainder of this paper.

Our presentation of this new non-equilibrium model is organized as follows: Section 2 describes the new traffic model and discusses its qualitative features, Section 3 examines the physical consequences of its mathematical properties, and Section 4 concludes this paper with a summary.

2. A new model and its mathematical properties [5]

As in nearly all non-equilibrium traffic models, the new model comprises the conservation equation (of vehicles) (5) and a momentum equation that describes speed dynamics. This momentum equation is given by

$$v_t + vv_x = -c(\rho)v_x \tag{7}$$

and is derived from a car-following model

$$\tau(s_n(t))\ddot{x}_n(t) = \dot{x}_{n-1}(t) - \dot{x}_n(t), \quad s_n(t) = x_{n-1}(t) - x_n(t). \tag{8}$$

Here $c(\rho) = \rho V_*'(\rho)$ is the traffic sound speed at which small traffic disturbances are propagated relative to a moving traffic stream. Other variables involved in (7) and (8) are: $\ddot{x}_n(t)$ and $\dot{x}_n(t)$ are the acceleration and speed of the nth car, respectively, and $\tau(s_n(t))$ is the driver response time, [6] which is a function of local spacing $s_n(t)$.

The derivation of Eq. (7) starts from the introduction of a velocity field $v(x,t) : \dot{x}_n(t) = v(x_n(t),t)$ and a vehicle spacing function $s(x,t) : s_n(t) = s(x_n(t),t)$; both are assumed sufficiently smooth. Next Eq. (8) is expressed in the new field variables

$$\tau(s(x(t),t))\frac{\mathrm{d}v(x(t),t)}{\mathrm{d}t} = \frac{\mathrm{d}s(x(t),t)}{\mathrm{d}t}, \tag{9}$$

and further expanded to obtain

$$\tau(s)(v_t + vv_x) = s_t + vs_x, \tag{10}$$

[5] Hoping to reach a wider audience, we keep the mathematical details to a bare minimum in our subsequent presentation. In most of the places we simply state the end results without presenting the proofs. Readers who are interested in the mathematical details are referred to Zhang (2000a). For a general reading on hyperbolic conservation laws, we recommend Le Veque (1992).

[6] *Driver response time* differs from *driver reaction time* in that the latter is more or less a constant (of the order of a second) while the former varies with spacing. Driver response time can approach driver reaction time in near-jam traffic, and approach infinity in very dilute traffic (i.e., when vehicle spacing approaches infinity).

where (x,t) is dropped to improve readability. Moreover, the conservation of vehicles implies that

$$s_t + vs_x = sv_x. \tag{11}$$

Substituting (11) in (10) one obtains

$$\tau(s)(v_t + vv_x) = sv_x, \tag{12}$$

which becomes (7) through the introduction of a traffic sound speed

$$\frac{s}{\tau(s)} = -\rho V_*'(\rho) \equiv -c(\rho) \geqslant 0,$$

as in Zhang (1998). [7]

Alternatively, one can obtain Eq. (7) by first expressing the right-hand side of Eq. (8) as $v(x(t)+s(x(t),t),t) - v(x(t),t)$ and then expand $v(x(t)+s(x(t),t),t)$ around (x,t) up to the first order. This derivation, however, is not exact because one neglected higher-order terms in the expansion. The benefit of taking this approach is that it can lead to a viscosity term v_{xx} if one considers the second-order approximation, hence a viscous momentum equation (the properties of a newly derived viscous model are discussed elsewhere (Zhang, 2000b)). Why one derivation leads to (7) exactly and the other approximately is at present not fully understood. Although expressions $\dot{x}_{n-1}(t) - \dot{x}_n(t)$, $\dot{s}_n(t)$, $v_{n-1}(t) - v_n(t)$ are fully equivalent in the microscopic description, their counterparts in the macroscopic description, $\dot{s}(x(t),t)$ and $v(x(t)+s(x(t),t),t) - v(x(t),t)$ appear not to be so. It seems that the definition and physical interpretation of the macroscopic field variables plays a subtle role in this discrepancy. This discrepancy warrants further investigation and we shall deal with it in future work.

Again, we express the new non-equilibrium traffic model in a vector form

$$\begin{pmatrix} \rho \\ v \end{pmatrix}_t + \begin{pmatrix} v & \rho \\ 0 & v + c(\rho) \end{pmatrix} \begin{pmatrix} \rho \\ v \end{pmatrix}_x = 0 \tag{13}$$

and find the eigenvalues of the flux Jacobian matrix to be

$$\lambda_1 = v + \rho V_*'(\rho) < v = \lambda_2, \tag{14}$$

with corresponding eigenvectors

$$r_1 = \begin{pmatrix} 1 \\ V_*'(\rho) \end{pmatrix}, \qquad r_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The new model is therefore strictly hyperbolic. Note that the momentum equation of the new model does not depend on the concentration gradient. Consequently, both of its characteristics travel no faster than traffic. In fact, the fastest characteristic travels at precisely vehicle speed.

[7] The quantity $S/\tau(S)$, given the notation $-\varpi(S)$ (or $-c(\rho)$ because $\rho = 1/S$), measures how fast a disturbance generated by a leading car propagates along a line of cars relative to moving traffic. It is often called traffic sound speed for its similarity to the speed of sound, which characterizes the speed of propagation of a small disturbance in a stationary or moving medium, such as gas.

2.1. Disturbance propagation in the new model

To examine how a small disturbance propagates in the new model, we perturb an equilibrium solution (ρ_0, v_0) $(v_0 = V_*(\rho_0))$ with a small disturbance $(\xi(x,t), w(x,t))$, where $(\xi(x,t), w(x,t))$ are sufficiently smooth functions of x and t.

Substituting the perturbed solution into (13), performing Taylor series expansions and neglecting higher order terms of $\xi(x,t), w(x,t)$, we obtain

$$\begin{pmatrix} \xi \\ w \end{pmatrix}_t + \begin{pmatrix} v_0 & \rho_0 \\ 0 & v_0 + c_0 \end{pmatrix} \begin{pmatrix} \xi \\ w \end{pmatrix}_x = 0, \tag{15}$$

where $c_0 = \rho_0 V_*'(\rho_0)$.

By eliminating w from (15) we get

$$\xi_{tt} + 2v_0\xi_{tx} + v_0^2\xi_{xx} + c_0(\xi_{tx} + v_0\xi_{xx}) = 0.$$

The latter equation can be written in a compact form

$$(\partial_t + (v_0 + c_0)\partial_x)(\partial_t + v_0\partial_x)\xi = 0. \tag{16}$$

Here we introduce the wave operator notation $(\partial_t + v_0\partial_x)\xi = \partial_t\xi + v_0\partial_x\xi$, where ∂_t and ∂_x denotes partial time and space derivatives. Noting that $v_0 + c_0 = f_*'(\rho_0)$, one can further simplify (16) to

$$(\partial_t + f_*'(\rho_0)\partial_x)(\partial_t + v_0\partial_x)\xi = 0. \tag{17}$$

This last equation implies that any small disturbance in the new model is propagated by two waves: the slower wave travels at a speed of $f_*'(\rho_0)$, identical to the first-order wave in the LWR model, and the faster wave travels at the speed of the traffic stream that carries it. The particular speeds of the two waves mean that the disturbances, on one hand, are carried downstream by the vehicles that generated them and, on the other hand, are propagated upstream through a line of vehicles that are behind the disturbance-generating vehicles. Both waves never travel faster than the traffic that carries them. The new model is therefore anisotropic, and we shall refer to it as the anisotropic non-equilibrium traffic flow model.

2.2. Consistency with the LWR model

The fact that the slower wave in the anisotropic non-equilibrium model travels at the same speed as the first-order wave in the LWR model is not accidental. This is because the anisotropic non-equilibrium model includes the LWR model as a special case. That is, when $v = V_*(\rho)$, the anisotropic non-equilibrium model reduces to the LWR model. To see this, we substitute $v = V_*(\rho)$ into (6) to obtain

$$V_*'(\rho)\rho_t + (V_*(\rho) + \rho V_*'(\rho))V_*'(\rho)\rho_x = 0, \tag{18}$$

or

$$V_*'(\rho)(\rho_t + (\rho V_*(\rho))_x) = 0. \tag{19}$$

Considering that $V_*'(\rho) < 0$, (19) leads to the LWR equation.

Note that the current anisotropic model does not have a relaxation term $(V_*(\rho) - v)/\tau(\rho)$ to force asymptotic convergence to the equilibrium equation (i.e., LWR) $\rho_t + (\rho V_*(\rho))_x = 0$ as all other non-equilibrium models (in Table 1) do. Any non-equilibrium state $(\rho, v) \neq (\rho, V_*(\rho))$ in this model will therefore remain in a non-equilibrium state over time. This allows for sustained presence of certain wave phenomena that do not exist in the equilibrium model (see Section 2.3 for details).

2.3. Propagation of discontinuities

Because the anisotropic non-equilibrium model is strictly hyperbolic, jump discontinuities can develop spontaneously in this model from even smooth initial data. Unlike their counterparts in the LWR model, the jump discontinuities in the anisotropic non-equilibrium model are not all shocks. In fact, there are two families of jump discontinuities: one family, which is associated with the first characteristic, comprises of shocks, and the other family, which is associated with the second characteristic, comprises of contact discontinuities. [8] This is because the first characteristic field, $r_1 = \binom{1}{V_*'(\rho)}$, is genuinely nonlinear

$$\nabla\lambda_1 r_1 = (c'(\rho), 1)\begin{pmatrix} 1 \\ V_*'(\rho) \end{pmatrix} = f_*'(\rho) < 0, \tag{20}$$

while the second characteristic field, $r_2 = \binom{1}{0}$, is linearly degenerate

$$\nabla\lambda_2 r_2 = (0, 1)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0. \tag{21}$$

Suppose that $U_l = \binom{\rho_l}{v_l}$ is the traffic state upstream of a shock and $U_r = \binom{\rho_r}{v_r}$ is the traffic state downstream of a shock, and the vector V of the conserved variables is given by a mapping $V_{l,r} = G(U_{l,r})$, then the speed of the shock is again determined by the Rankine–Hugoniot jump condition (Le Veque, 1992)

$$\sigma[V] = [F(V)]; \quad [V] = V_r - V_l, \ \ [F(V)] = F(V_r) - F(V_l), \tag{22}$$

where $F(V)$ is the flux vector of the conserved variables V. Eliminating σ from (22) one finds that $U_l = \binom{\rho_l}{v_l}$ and $U_r = \binom{\rho_r}{v_r}$ are related by the following equation

$$v_r = V_*(\rho_r) + v_l - V_*(\rho_l), \quad \rho_l < \rho_r. \tag{23}$$

Substituting (23) back into (22), one obtains the shock velocity

$$\sigma = \frac{\rho_l v_l - \rho_r v_r}{\rho_l - \rho_r} = (v_l - V_*(\rho_l)) + \frac{\rho_l V_*(\rho_l) - \rho_r V_*(\rho_r)}{\rho_l - \rho_r} = (v_l - V_*(\rho_l)) + \sigma_*. \tag{24}$$

Clearly, shocks in the anisotropic non-equilibrium model generally travel at different speeds from the corresponding shocks in the LWR model – their speeds differ by the amount of the difference between equilibrium and non-equilibrium speeds of the upstream traffic. [9] The only exception is

[8] A contact discontinuity is an interface that separates two regions of different traffic concentration/speed, across which no flow passes.

[9] Note that v_l is the actual traffic speed of the traffic upstream of a shock, whose concentration is ρ_l. The equilibrium speed $V_*(\rho_l)$ here is a fictitious speed that should be interpreted this way: if the upstream traffic of concentration ρ_l was in equilibrium then its travel speed would be $V_*(\rho_l)$.

when traffic states on both sides of a shock are equilibrium states, i.e., $v_{\mathrm{l,r}} = V_*(\rho_{\mathrm{l,r}})$, then $\sigma = \sigma_*$. In light of the consistency result of Section 2.2, the latter revelation should not surprise us.

In contrast, contact discontinuities are propagated simply at the characteristic speed $\lambda_2(U_{\mathrm{l}}) = \lambda_2(U_{\mathrm{r}}) = v$, that is, they are stationary relative to the moving traffic on both sides of the contacts. In other words, they travel at the speed of traffic. These waves cannot occur in the congested regime of LWR traffic.

2.4. Dissipation of traffic jams

As in the LWR model, initial jump discontinuities in the anisotropic non-equilibrium model can also expand and smooth themselves out. These expansion waves are called rarefaction waves, a term originating from the study of gas dynamics. Not all jumps in the anisotropic non-equilibrium model expand as a single rarefaction wave. In fact, for a jump $U_{\mathrm{l,r}}$ to expand as a single rarefaction wave, it must satisfy the following condition [10]

$$v_{\mathrm{r}} = V_*(\rho_{\mathrm{r}}) + v_{\mathrm{l}} - V_*(\rho_{\mathrm{l}}), \quad \rho_{\mathrm{l}} > \rho_{\mathrm{r}}, \tag{25}$$

that is, traffic must be heavier upstream than downstream of the initial jump, and for any fixed U_{l}, U_{r} must fall on the shifted equilibrium curve.

It should be noted that when $v_{\mathrm{l}} = V_*(\rho_{\mathrm{l}})$, that is, the upstream state is an equilibrium state, $v_{\mathrm{r}} = V_*(\rho_{\mathrm{r}})$ and the anisotropic non-equilibrium model gives rise to the same rarefaction waves as the LWR model. This again attests to the consistency between the anisotropic non-equilibrium model and the LWR model. It is also worth mentioning that the anisotropic non-equilibrium model has only one family of rarefaction waves, which is associated with the first characteristic. This is in contrast with other higher-order models that have two families of rarefaction waves, one for each characteristic.

3. Physical properties of the anisotropic non-equilibrium model

In this section we examine the physical consequences of the mathematical properties of the anisotropic non-equilibrium model discussed in Section 2. We are particularly concerned with four things: the model's ability to explain vehicle clustering, its ability to model forward propagation of disturbances in congested traffic, its ability to correctly predict queue-end behavior (Daganzo, 1995), and its ability to model phase transitions.

3.1. Vehicle clustering and forward propagation of disturbances in heavy traffic

When traffic concentration is inhomogeneous on a highway, such as the traffic hump in Fig. 1, the LWR model predicts two things: the front jump will expand to form a rarefaction wave, and the rear jump will propagate as a shock (Fig. 2(a)). The end result is that at a certain time the

[10] This condition is the result of the following property: rarefaction waves are smooth, scaling-invariant (i.e., a function of x/t) solutions of the non-equilibrium model.

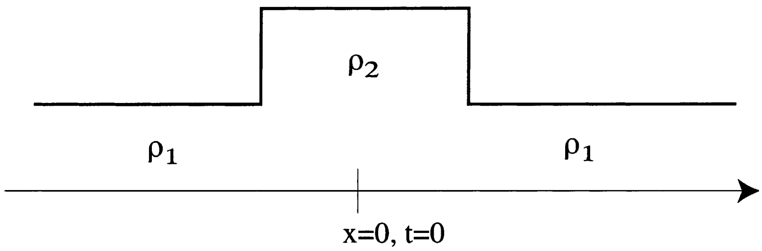


Fig. 1. A traffic hump.

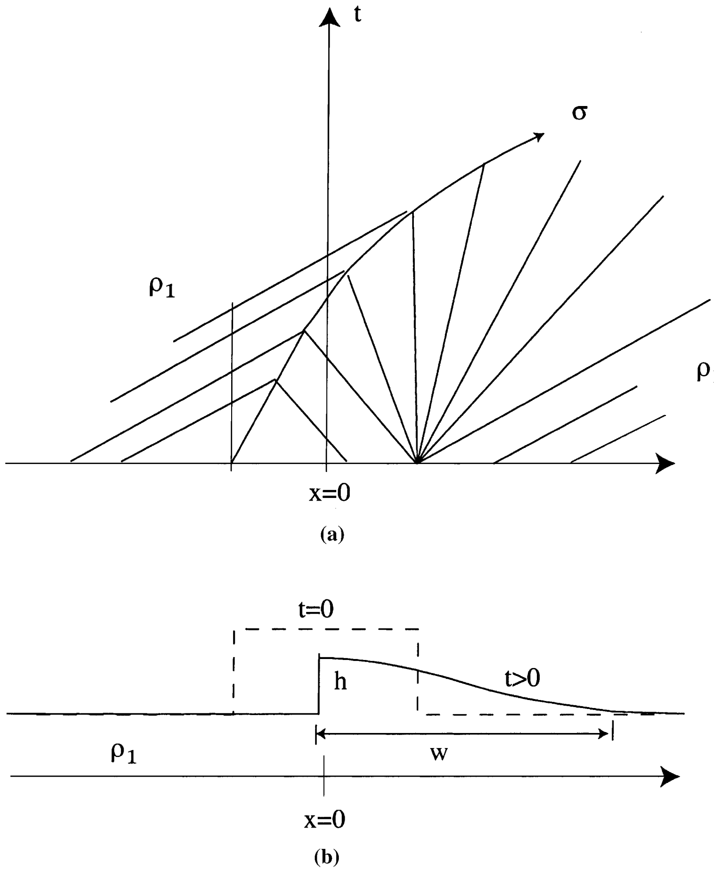
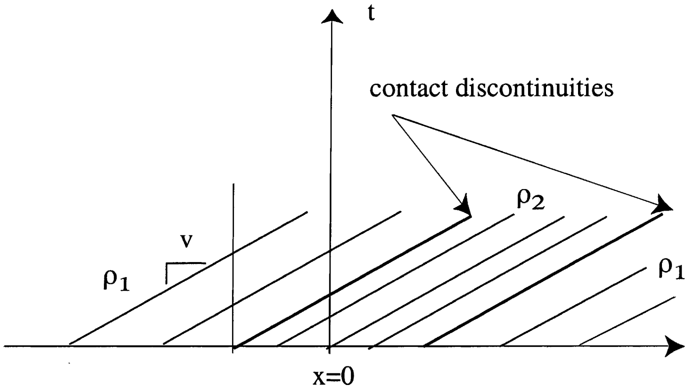


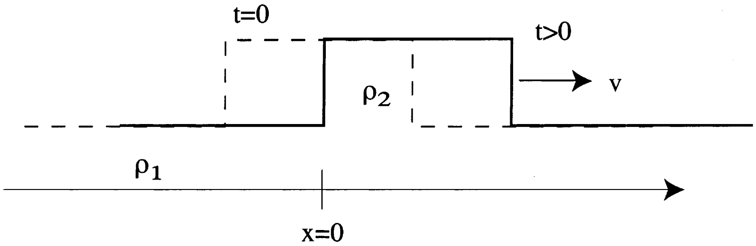
Fig. 2. The evolution of a traffic hump predicted by the LWR model.

shock and rarefaction waves will interact to form a wedge with height h and width w (Fig. 2(b)). As time increases, h decreases and w increases, but the total area under the wedge will be constant, equal to the area under the initial square hump. This behavior implies that any cluster of vehicles that is surrounded by traffic of lighter concentration will disperse itself. Recent observations of freeway traffic, however, indicate that disturbances of higher flow/concentration tend to move with traffic without diffusing (Daganzo, 1995; Cassidy and Windover, 1995), a phenomenon that contradicts the LWR predictions. A somewhat related phenomenon, observed much earlier by Edie and Bavarez (1967) also contradicts the predictions of the LWR model. In their study of tunnel traffic, Edie and Bavarez noted that "...small changes in flow may not propagate at a speed equal to the slope of the tangent to a steady-state *q–k* curve as suggested by the hydrodynamic wave theories of traffic flow. Instead, they are carried along at about stream speed or only slightly less than stream speed right up to saturation flows, at which level they suddenly reverse directions".

Both observations suggest that, apart from the family of waves that travels against a moving traffic stream, there is another family of waves that travels with the traffic stream. Because the LWR model produces only waves of the first family, one would not expect it to be a proper model for explaining the aforementioned phenomena (one could with a suitably modified LWR model). One suspects, however, that the anisotropic non-equilibrium model might be able to model these phenomena because it has two families of waves and the second family travels at precisely traffic speed. We shall use the traffic hump to demonstrate that this is indeed the case. Without loss of generality, suppose that in the hump (ρ_2, v_2) is an equilibrium state and (ρ_1, v_1) is a non-equilibrium state. When drivers in different concentration regions prefer to drive at the same speed (i.e., $v_1 = v_2 = v$, which is admissible in the anisotropic non-equilibrium model but not in the LWR model), the anisotropic non-equilibrium model predicts that both front and rear jumps of the hump are propagated as contact discontinuities (Figs. 3(a) and (b)). That is,



(a)



(b)

Fig. 3. The motion of a traffic hump predicted by the anisotropic non-equilibrium model.

the hump is moving downstream at traffic speed without diffusing (Fig. 3(b)). As long as a hump is moving as a solid block, it does not matter whether ρ_2 is in the congested or uncongested region of the flow-concentration diagram. Thus, disturbances in the anisotropic non-equilibrium model can also propagate at traffic speed even in heavy traffic conditions. It should be noted, however, that humps as in Fig. 3(b) are likely to be short-lived in heavy traffic because opportunities to interact with other types of waves abound in such traffic. Moreover, the anisotropic non-equilibrium model appears to be inadequate to explain the sudden reversal of wave directions near saturation flow observed in (Edie and Bavarez, 1967). This difficulty will be examined in future work.

3.2. Queue-end behavior

The gas-like behavior of higher-order models manifests itself most prominently in the formation of stationary queues. An example is this. Consider a road with moderate traffic bounded by traffic lights at both ends, and suppose that the downstream light is red while the upstream light is green for some time. A queue of a certain size then forms in front of the downstream light. When the upstream light turns red, it starves the queue and prevents it from growing further. If both lights remain red indefinitely, there will be an empty road ($\rho = 0$) followed by jammed traffic ($\rho = \rho_j$), with a concentration jump separating these two regions. That is, the traffic concentration can be described by a Heavyside function (Daganzo, 1995)

$$\rho(x) = \rho_j H(x), \quad H(x) = \begin{cases} 0 & -L \leqslant x < 0 \\ 1 & L > x \geqslant 0 \end{cases}. \tag{26}$$

Here we assume, without loss of generality, that the jump is located at $x = 0$ and the queue occupies half of the road.

A system of strictly hyperbolic partial differential equations, together with an initial condition given by (26) with $L = +\infty$, forms a well-posed initial value problem called the Riemann problem. One would expect a correct model to predict $\rho(x, t) = \rho(x)$ for all time t if it started with this initial condition (i.e., $\rho(x, 0) = \rho(x)$). This is indeed the case with the LWR model, because the two constant states $(0, v_f)$ and $(\rho_j, 0)$ can be connected by a single shock with speed $\sigma_* = (0 - 0)/(\rho_j - 0) = 0$, that is, the shock is stationary. Existing higher-order theories, however, generally cannot predict the correct time evolution of the stationary queue – they predict a smooth concentration profile rather than a shock (Daganzo, 1995; Zhang, 2001). Because the total number of vehicles in the queue is constant, a smooth concentration profile implies that vehicles near the end of the queue move backward. This phenomenon is dubbed 'wrong-way' travel in Daganzo (1995). The reason for such behavior, as explained in Daganzo (1995), is that higher-order theories usually are diffusive, and diffusion waves travel both ways.

The anisotropic non-equilibrium model contains no diffusive waves. One would therefore expect it to also predict correctly the time evolution of the stationary queue, as did the LWR model. As it turns out, this is indeed the case. Note that the two states $(0, v_f)$ and $(\rho_j, 0)$ satisfy relation (23), the condition for connecting two constant states with a single shock. Therefore, the anisotropic non-equilibrium model also predicts that the jump will move as a shock, whose speed can be computed from the Rankine–Hugoniot jump condition. It turns out that this speed is zero. That is, the jump remains in its original location at all times according to the anisotropic non-equilibrium model, which is the correct prediction. The anisotropic non-equilibrium model therefore passed perhaps the most stringent validity check faced by every continuum model. To the best of our knowledge, only one other non-equilibrium continuum model (the model developed in Zhang (1998) with an affine $V_*(\rho)$) passes this validity check (Zhang, 2000a).

3.3. Phase transitions

A phase in traffic flow is simply a traffic state given by the triplet (ρ, v, q). Phase transitions refer to the event of traffic change from one state to another. It is customary to study phase transitions in the phase space, that is, the three-dimensional space of (ρ, v, q) in traffic flow. Because flow rate (q) is the product of speed and concentration, traffic phase transitions can be studied in one of the three phase planes (ρ, v), (ρ, q) or (q, v), all well-known to traffic engineers. We choose the (ρ, q), phase plane in the presentation below because of its popular use in current traffic literature.

The simplest and most basic phase transition in traffic flow is from one constant state to another, such as the one depicted in Fig. 4(d). In this figure the solid lines depict traffic concentration and dotted lines depict traffic flow rate. Traffic state U is defined differently than before; here it is

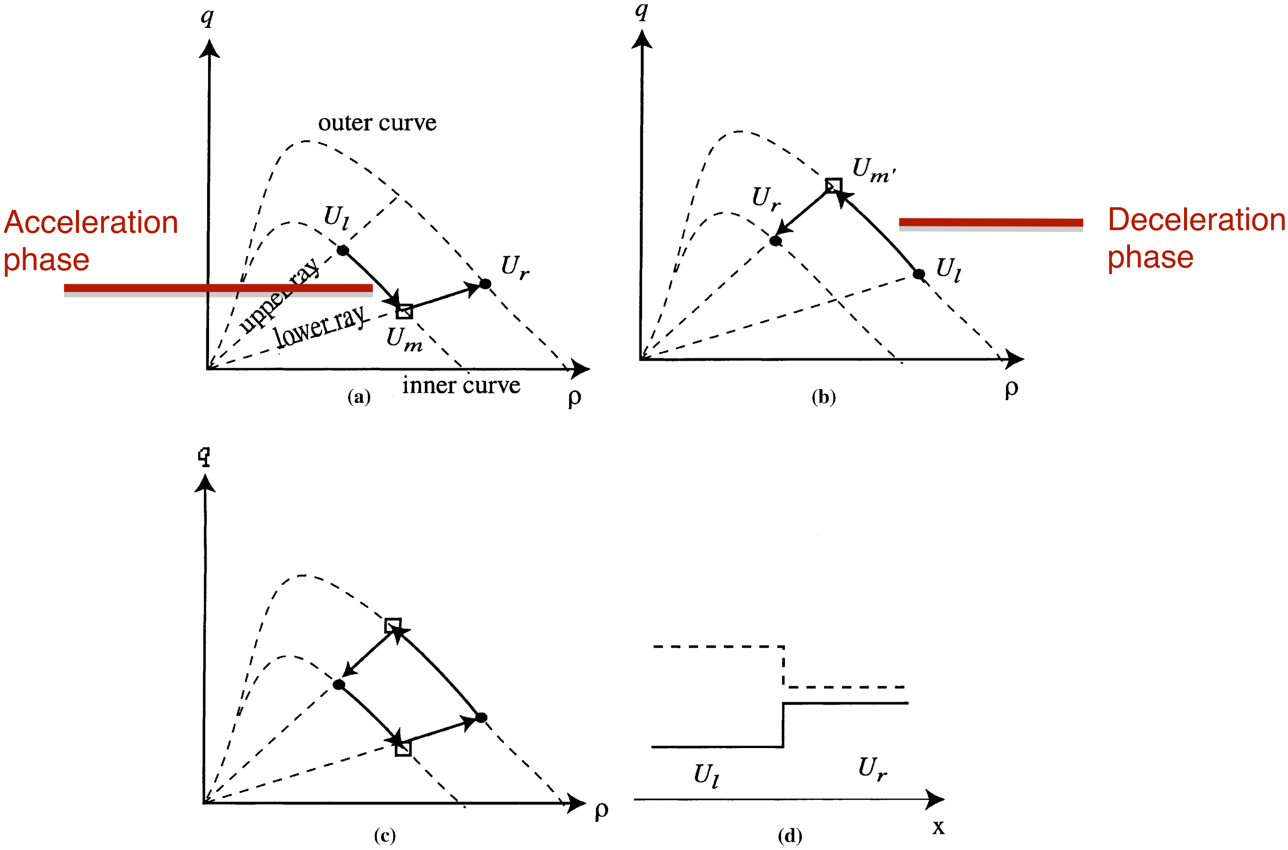


Fig. 4. Phase transitions predicted by the anisotropic non-equilibrium model.

$U = (\rho, q)$. Recall that, in the LWR model, both the left/upstream and right/downstream states U_l and U_r must fall on the equilibrium phase curve $(\rho, f_*(\rho))$, and the phase transition is a shock if traffic is lighter upstream and a smooth expansion wave if traffic is heavier upstream. In contrast, both states U_l and U_r do not have to be on the equilibrium phase curve in the anisotropic non-equilibrium model. Consequently, phase transitions in the anisotropic non-equilibrium model are more complicated.

It is noted that given any state U_l in the first quadrant of the phase plane a particular phase curve

$$q = f_*(\rho) + \left(\frac{q_\mathrm{l}}{\rho_\mathrm{l}} - V_*(\rho_\mathrm{l})\right)\rho \tag{27}$$

can be drawn to pass through the origin and U_l (e.g., Fig. 4(a)). A similar phase curve can also be drawn to pass through the origin and U_r:

$$q = f_*(\rho) + \left(\frac{q_\mathrm{r}}{\rho_\mathrm{r}} - V_*(\rho_\mathrm{r})\right)\rho. \tag{28}$$

We shall remark that these special phase curves serve precisely the same function in the anisotropic non-equilibrium model as the equilibrium phase curve does in the LWR model, that is, they are the loci of traffic states that form either shock or smooth expansion waves. Moreover, these two curves are connected through rays

$$q = \frac{q_\mathrm{r}}{\rho_\mathrm{r}}\rho \tag{29}$$

and

$$q = \frac{q_\mathrm{l}}{\rho_\mathrm{l}}\rho \tag{30}$$

that pass through U_r and U_l, respectively (Figs. 4(a)–(c)). These rays represent transitions that do not involve changes in speed.

With the help of the above phase curves and rays, we can describe how the anisotropic model handles transitions from U_l to U_r. First, we examine the case of $\rho_\mathrm{l} < \rho_\mathrm{r}$ (Fig. 4(a)). To reach U_r, the model prescribes that traffic first decelerates through a shock to state U_m along the inner phase curve, then transfers from U_m to U_r along the lower ray. Note that traffic maintains its speed when it is going through the second phase transition. Physically, such a transition implies the following driver behavior – when drivers encounter a congestion, they slow down to travel at the congested speed (shock effect) but leave a bigger gap between them than that of U_r traffic (the initially congested traffic). Once they slow down to the speed of the congested traffic, they adopt a wait-and-see attitude and are in no hurry to close the wider gaps. Therefore we have three evolving traffic states U_l (lighter and faster traffic upstream of the shock), U_m (slowed-down traffic downstream of the shock), and U_r (slower traffic downstream of slowed-down traffic).

Now if we switch the positions of the two states on the road, that is, $\rho_\mathrm{l} > \rho_\mathrm{r}$ (Fig. 4(b)), the anisotropic non-equilibrium model does not allow traffic to return to U_l via the same path (i.e., U_r–U_m–U_l). Rather, it predicts that traffic first accelerates through an expansion wave to a new state $U_{m'}$ that is on both the outer phase curve and the upper ray, then slides down to U_l along the upper ray. Traffic that switches between two states U_l and U_r will therefore completes a counter-clockwise phase loop U_l–U_m–U_r–$U_{m'}$–U_l, resulting in a periodic wave motion in the x–t plane.

A few remarks are in order here. First, the anisotropic non-equilibrium model does not preclude clockwise phase loops such as $U_{m'}$–U_r–U_m–U_l–$U_{m'}$. In fact, they are as logical as the counter-clockwise phase loops. Second, one cannot fail to notice the striking similarities of these phase transitions to the ones described in Newell (1965). The inner/outer phase curves can be viewed as the deceleration/acceleration phase curves when traffic traverses phase loops counter-clockwise, and acceleration/deceleration phase curves when traffic traverses phase loops clockwise. These phase curves, however, are not fixed a priori as in Newell (1965). They are determined endogenously through the equilibrium phase curve and transient traffic states. Finally, more complicated phase transitions can be dealt with in the same manner, with special attention being paid to wave interactions.

4. Concluding remarks

Despite its simplicity, the LWR model is a remarkably robust and powerful theory. With it we could model, correctly, the most prominent features of traffic flow: formation, propagation and dissolution of traffic jams. Yet it is not a perfect theory. Two important traffic phenomena – stop–start waves and vehicle clustering, for example, elude the LWR model. The inability of the LWR theory to model such phenomena derives from its suppression of non-equilibrium phase transitions to a single phase curve: the equilibrium curve $V_*(\rho)$. Subsequent efforts to improve the LWR model were therefore focused on bringing non-equilibrium phase transitions into continuum theories, largely through introducing an evolution equation to model speed dynamics. These so-called higher-order models indeed can model non-equilibrium phase transitions, but they also destroy the conservation structure of the LWR model; worse than that, the majority of them violates the anisotropic property of traffic flow and produces negative travel speeds under certain circumstances.

In this paper we presented a non-equilibrium model that not only retains all the capabilities of the LWR model, but also completely eradicates the gas-like behavior that plagues most other non-equilibrium models. Moreover, this model captures certain traffic phenomena that elude the LWR model, such as vehicle clustering and forward propagation of disturbances in dense traffic. It is not clear at present if it can model stop-and-go traffic. We believe that such a possibility is quite real because the anisotropic non-equilibrium model allows for non-equilibrium phase transitions that are similar to those described in Newell (1965). We plan to explore this possibility in our future work, together with numerical approximations, experimental validation and extension of the model to describe traffic on inhomogeneous roads.

References

Cassidy, M.J., Windover, J., 1995. A methodology for assessing the dynamics of freeway traffic flow. Paper presented at the 1995 meeting of the Transportation Research Board, Washington, DC.

Daganzo, C.F., 1995. Requiem for second-order fluid approximations of traffic flow. Transp. Res. B 29, 277–286.

del Castillo, J.-M., Pintado, P., Benitez, F.-G., 1993. A formulation of reaction time of traffic flow models. In: Daganzo, C.-F. (Ed.), Transportation and Traffic Theory. Elsevier Science, Amsterdam, pp. 387–405.

Edie, L.C., Bavarez, E., 1967. Generation and propagation of stop–start traffic waves. In: Edie, L.C., Herman, R., Roghery, R. (Eds.), Vehicular Traffic Science. Elsevier, Amsterdam, pp. 26–37.

Hauer, E., Hurdle, V.F., 1979. Discussions in Payne, H.J., "FREFLO: a macroscopic simulation model of freeway traffic", Transportation Research Record 722, TRB.

Leo, C.H., Pretty, R.L., 1992. Numerical simulation of macroscopic continuum traffic models. Transp. Res. B 26 (3), 207–220.

Le Veque, R.J., 1992. Numerical Methods for Conservation Laws. 2nd ed. In: Lectures in Mathematics ETH Zurich. Birkhauser, Basel, Switzerland.

Lighthill, M.J., Whitham, G.B., 1955. On kinematic waves: II. A theory of traffic flow on long crowded roads. Proc. R. Soc. London Ser. A 229 (1178), 317–345.

Maes, W., 1979. Traffic data collection system for the Belgian motorway network – measures of effectiveness aspects. In: Proceedings of the International Symposium on Traffic Control Systems, vol. 2D. pp. 45–73.

Newell, G.F., 1965. Instability in dense highway traffic, a review. In: Almond, J. (Ed.), Proceedings of the Second International Symposium on the Theory of Traffic Flow. pp. 73–83.

Payne, H.J., 1971. Models of freeway traffic and control. Simulation Councils Proc. Ser. Math. Models Public Syst. 28 (1), 51–61.

Papageorgiou, M., Blosseville, J., Hadj-Salem, H., 1989. Macroscopic modelling of traffic flow on the Boulevard Peripherique in Paris. Transp. Res. B 23, 29–47.

Richards, P.I., 1956. Shock waves on the highway. Oper. Res. 4, 42–51.

Treiterer, J., Myers, J.A., 1974. The hysteresis phenomena in traffic flow. In: Buckley, D.J. (Ed.), Proceedings of the Sixth Symposium on Transportation and Traffic Theory. pp. 13–38.

Whitham, G.-B., 1974. Linear and Nonlinear Waves. Wiley, New York.

Zhang, H.M., 1998. A theory of non-equilibrium traffic flow. Transp. Res. B 32 (7), 485–498.

Zhang, H.M., 1999. Analyses of the wave and stability properties of a new continuum traffic theory. Transp. Res. B 33 (6), 387–398.

Zhang, H.M., 2000a. Structural properties of solutions arising from a non-equilibrium traffic flow theory. Transp. Res. B. 34, 583–603.

Zhang, H.M., 2000b. Driver memory, traffic viscosity and a general viscous traffic model. ITS Working Paper. University of California at Davis.

Zhang, H.M., 2000c. Anisotropic property revisited – does it hold in multi-lane traffic? Transportation Research (accepted for publication).

Zhang, H.M., 2001. New perspectives on continuum traffic flow models (to appear in Zhang, H.M. (Guest Ed.). Traffic Flow Theory (special issue). J. Networks Spatial Econ. 1 (1)).