Pergamon

Transpn. Res.-B. Vol. 29B, No. 5, pp. 373-389, 1995

0191-2615/95 $9.50+.00

0191-2615(95)00008-9

ON THE FUNCTIONAL FORM OF THE SPEED-DENSITY RELATIONSHIP—I: GENERAL THEORY

J. M. DEL CASTILLO and F. G. BENÍTEZ
Department of Mechanical Engineering, University of Seville, Avd. Reina Mercedes, 41012 Sevilla, Spain

(Received 19 January 1994; in revised form 4 August 1994)

Abstract—In this work a functional form for the speed-density relationship is presented. This functional form is made up of a nondimensional spacing, the *equivalent spacing* and of a function, the *generating function*, whose argument is the equivalent spacing. This functional form is derived by means of two different arguments. The first argument is based on the set of properties that the volume-speed-density relationships should satisfy. The second one arises when applying a dimensional analysis to a generic car-following model. Finally, several examples of generating functions are shown.

INTRODUCTION

The correlation of the speed of the vehicles with the spacing under time-stationary and space-homogeneous traffic flow is a fact which has been sufficiently contrasted experimentally. Under these traffic flow conditions one often speaks of equilibrium speed (V_e) although throughout this work for speed we will mean equilibrium speed. From the speed-spacing correlation it follows that all variables of traffic flow — mean speed, density and volume — are correlated. The early attempts to estimate the capacity of the roads led to the first speed-volume relationships. During the 1920s and 30s a series of works on this matter were published, mostly in Germany, England and the United States. It is difficult to determine which was the first work published. According to the bibliography consulted for the present paper, the work of Schaar (1925) might deserve this honor. The speed-volume relationships proposed by the pioneers of the Traffic Engineering obeyed the general expression

$$Q = \frac{V}{L + TV + V^2/2g\mu}$$

where Q is the volume and V the speed. The remaining parameters are the vehicle length L, the reaction time of drivers T, the friction coefficient μ and the gravity acceleration g. The above expression results from the safety spacing allowed by the drivers, which is the vehicle length plus the distance traveled before the driver's reaction, plus the braking length.

A completely different approach to speed-density relationships was initiated by Greenshields (1935) in his study of traffic flow on rural roads. No assumptions were made as to the behavior of drivers and speed-density relationships were obtained by fitting models to observed traffic data. The model proposed by Greenshields was a linear one

$$V = V_f\left(1 - \frac{K}{K_j}\right)$$

where K is the traffic density, K_j its maximum value and V_f the maximum speed. Other speed-density relationships are due to Greenberg (1959)

$$V = V_c \ln\left(\frac{K_j}{K}\right)$$

Underwood (1961)

$$V = V_f \exp\left(-\frac{K}{K_c}\right)$$

and Drake *et al.* (1967)

$$V = V_f \exp\left[-\frac{1}{2}\left(\frac{K}{K_j}\right)^2\right]$$

where K_c and V_c are the values of density and speed at which maximum flow or capacity is reached.

Another family of speed–density curves has been derived from *car-following* models

$$V = V_f\left[1 - \left(\frac{K}{K_j}\right)^m\right]^n \tag{1}$$

which is a generalisation of the relationship proposed by Pipes (1967)

$$V = V_f\left(1 - \frac{K}{K_j}\right)^n.$$

In addition to the former models, Prigogine and Herman (1971) have numerically calculated speed–density curves from the equations of the kinetic theory of traffic flow. The search of a general analytical expression for the speed–density relationship is therefore an open problem. The adequate selection of such a relation is of utmost importance when describing traffic flow by means of fluid dynamics models. An extensive review of the most common speed–density curves may be found in Drake *et al.* (1967).

Two approaches for stating speed–density relationships may be distinguished. The classical approach has been a purely mathematical one. Firstly, an analytical expression containing several parameters is proposed and then the values of these parameters are estimated by fitting the expressions to traffic data. Finally, an interpretation of the parameters in terms of properties of traffic flow is sought, in order to provide the analytical expression with a phenomenological meaning. The models of Greenshield, Greenberg, Underwood and Drake have been derived in such a manner.

A different approach, which may be called phenomenological or behavioral, is based on assumptions about the driver behavior with respect to some traffic variables. The early models for estimating the capacity and those derived from car-following models belong to this approach. The models of Greenshields and Greenberg may also be included in this family, since they can be derived from behavioral considerations, although they were not initially so. These two models arise by considering the acceleration of a vehicle to be proportional to the absolute and relative variations of the visual angle subtended by the preceding vehicle, see Gerlough and Huber (1975) for the derivation.

In principle none of these two approaches should give superior results, although the latter is more desirable since it is based on the behavior of drivers. Both approaches have been used in this work. The functional form of the speed–density relationship is initially derived from the mathematical properties of the speed–density curves. These properties impose such severe restrictions for the curves that one can only easily find a curve satisfying all of them if it has the above-mentioned functional form. Later on, the application of the dimensional analysis to a general car-following model yields the same functional form. The assumptions made for this analysis concern the expected behavior of the drivers and the fundamental parameters of traffic flow.

THE PROPERTIES OF THE TRAFFIC FLOW RELATIONSHIPS

This section reviews the mathematical properties of the speed–density and flow–density curves. A first group of properties are those that may be regarded as trivial or obvious:

(i) the values of the speed go from zero to a maximum referred to as *free flow speed*, V_f, the speed of a vehicle traveling alone;
(ii) the values of traffic density lie between zero (free flow) to the maximum called *jam density* K_j;
(iii) the free flow speed is the limit of the desired speed when the spacing tends to infinity, or inversely when the density approaches zero, thus $V(0) = V_f$;
(iv) vehicles stop at jam density, therefore $V(K_j) = 0$;
(v) speed decreases with density, that is, $V'(K) < 0$ for $0 < K \leqslant K_j$, where ($'$) stands for the derivative with respect to density.

Another property should be added to account for the fact that as traffic flow becomes lighter, the dependence of speed on density vanishes, since the interactions between drivers also vanish. Thus, the following equation must hold

$$V'(0) = 0. \tag{2}$$

The above properties were already stated by Greenshields and they can be deduced by considering the traffic flow as a stationary phenomenon. Hence, they will be referred to as "static" properties. There is still an important property concerning the flow–density relationship. This property is a "dynamic" one and arises from the equation of continuity of traffic flow

$$\frac{\partial K}{\partial t} + \frac{\partial Q}{\partial x} = 0. \tag{3}$$

The above equation expresses the balance between the growth rate of the number of vehicles in a road section with the net flow across the section. The flow–density curve must be concave so that eqn (3) only renders solutions with deceleration shock waves, Whitham (1974). The mathematical condition is

$$Q''(K) < 0. \tag{4}$$

A shock wave in real traffic would correspond to a sudden change of speed. Such changes actually describe violent brakings that appear when avoiding a collision with the preceding vehicle. The concavity condition is therefore in accordance with the expected behavior of the drivers. However, it should be remarked that there are some situations in which concavity is not observed, for instance, during saturation conditions.

None of the speed–density models mentioned in the previous section entirely fulfils the above restrictions. Some of them extend over an infinite range of speeds or densities, which is obviously not realistic. The property (2) is not taken into account either, except in the curve of Drake. However, the property more often violated is that of concavity. This property was first pointed out by Franklin (1961): "The relationship between flow and density is such that for densities greater than that corresponding to maximum flow the gradient of the curve increases in magnitude with density. In such a case, if the flow is decelerated each successive wave carrying a small increase in density travels faster than the previous waves and the waves coalesce to from a shock wave". Astonishingly, the concavity property has been completely forgotten in the traffic flow research done since then. Almost 30 yr had elapsed before Ansorge (1990) brought it back into the theory. He has proven that (4) is equivalent to the entropy condition which allows to select out, from the set of possible solutions of (3), the physically relevant solution. In addition, Ansorge has given a physical interpretation of the concavity condition in terms of what he names "driver's ride impulse". He concludes: "this interpretation of the entropy condition in the traffic flow model under consideration seems to coincide with real life behaviour of motorists". For all these reasons it seems convenient to devise a technique for formulating speed–density relationships that satisfy the above properties.

Apart from these problems, there is a controversy over the continuity of the speed–density function. Since Edie published his paper (1961), several authors have suggested that discontinuous functions are required to properly describe the speed–density dependence. More recently, authors such as Koshi *et al.* (1983) and Payne (1984) have based this opinion on a thorough analysis of freeway traffic data. Edie observed a discontinuity for the volume at 90 cars/mile — when it reached capacity — and deduced the existence of two different regimes of traffic flow through the Lincoln Tunnel in New York City, the site of the study. The data gathered by Payne were filtered to remove the non-equilibrium data. After this process, the appearance of continuity vanished since the resulting equilibrium data set did not contain speed values in the 30 mph to 40 mph range. Koshi *et al.* collected data on Tokyo Freeways through several methods and found a rather exotic volume–density curve exhibiting a shape like a mirror image of the greek letter λ.

However, one can argue against the findings of Edie and Koshi *et al.* that the data used do not necessarily correspond to equilibrium conditions. Thus, the results will be strongly influenced by the nature of the traffic operations at the particular location of the study and so will the speed–density curves obtained. We share the opinion of Hall *et al.* (1986) on this matter: the observed gap could depend on the existence of a bottleneck downstream from the location where the data were acquired. Furthermore, the data have probably been grouped together, representing conditions that never occured, because vehicles belonging to a certain speed class, could not have traveled together. Besides these critics, Hall *et al.* make a thorough analysis of the works of Ceder (1976) and Easa (1983) about discontinuous or two-regime flow–density relationships. Their analysis reveals some of the misleading techniques commonly used in the study of traffic flow relationships.

As for the findings of Payne, the procedure for rejecting non-equilibrium points might be the cause of the discontinuity. Unfortunately, he does not explain the data selection procedure. In conclusion, when developing speed–density models, the locations of the detectors should be carefully chosen and analytical procedures based on the statistical properties of traffic flow should be used to properly select the collected data.

There is a point on which we do not agree with Hall *et al.*: the differentiability of the speed–density curve. They found that a smooth curve did not fit the data well in the troublesome transition regime at capacity, whereas an inverted V shaped curve was ideal for explaining the transition between the congested and uncongested regimes. They gave an argument based on shock wave analysis to justify the inverted V shape. We think that if the speed–density curve has to account for equilibrium states of traffic flow, it is a misconception to interpret features of this curve in terms of dynamic phenomena. Such features, like gaps and peaks, should therefore be explained by a dynamic microscopic model. The reason for this is clear: the possible existence of a gap would be the manifestation of an instability phenomenon and should be explained by a traffic flow dynamic model. Further, if one assumes a discontinuous or a non-differentiable speed–density curve, one rules out, beforehand, the possibility of a smooth transition regime. The above considerations serve as arguments to support the assumption of continuity and differentiability of the speed–density curve.

THE PARAMETERS OF TRAFFIC FLOW

The subject of this section is to identify the parameters that characterize the equilibrium traffic flow relationships. We look for a set of parameters that can account for all the local properties of the traffic flow relationships. Thus, the free flow speed and the jam density should be included in that set since they are the extreme values of the speed–density curve. Most authors consider a speed–density relationship characterized by these two parameters and by a series of coefficients with no physical meaning. This is the case of the exponents m and n in the speed–density relationship (1) deduced from car-following models. We claim instead that the traffic flow relationships are strongly characterized by an essential parameter which seems to have been omitted in the existing models. This parameter is the kinematic wave speed at jam density C_j, which is given by

$$C_j = Q'(K_j). \tag{5}$$

Therefore, the speed–density relationship should be written in the form

$$V = F(K, K_j, V_f, C_j, n_1, \ldots, n_p)$$

where, $n_1, \ldots, n_p$ are a set of nondimensional parameters that takes into account all the properties we are not able to interpret in terms of physical parameters. This collection of nondimensional parameters somehow measures our ignorance about the nature of traffic flow equilibrium states.

Once the set of fundamental parameters has been established—K_j, C_j, and V_f—we are interested in classifying them according to the range of variation of their values. We claim that the jam density and the kinematic wave speed at jam density may be considered as traffic flow constants, since their values do not show great variations. On the contrary, the free flow speed varies over a much broader range and is strongly dependent on the road characteristics. The jam density is mainly determined by the vehicle length and the variation of the values found should be attributed to the randomness of the traffic flow composition. In general, values on the order of 150 vehicles per km and lane may be assumed, which correspond to a mean vehicle length plus bumper-to-bumper distance of 6.7 m. This value is the *jam spacing*, $H_j = 1/K_j$, which is logically another fundamental parameter. Thus, we will also refer to H_j, C_j, and V_f as the set of fundamental parameters.

The kinematic wave speed is also determined by the vehicle length and the response time of drivers. Thus, relative variations of C_j should be of the same order as those of the jam density, since the response time of drivers at jam density does not exhibit a great dispersion. We have found a mean value of −20 km/h. This finding will be justified next in this section. The important point is to recognize that these two parameters are almost constant in the sense that they are weakly conditioned by the characteristics of the road. At the worst, they may depend on the lane due to the different vehicle composition in each lane. On the contrary the free flow speed is entirely determined by the characteristics of the road.

In the literature about traffic flow numerous testimonies confirming that the kinematic wave speed at jam density is almost constant can be found. Most of them arise when studying speed–density relationships in congested flow or platoons. The value of C_j is then given by

$$C_j = K_j \left.\frac{\mathrm{d}V}{\mathrm{d}K}\right|_{K_j} = -H_j \left.\frac{\mathrm{d}V}{\mathrm{d}H}\right|_{H_j}.$$

It can also be estimated from the time for a car to start moving as a function of its position in a stopped platoon.

Pipes (1953) proposed a car following model based on a rule stated in the California Vehicle Code: "a good rule for following another vehicle at a safe distance is to allow yourself the length of a car (about fifteen feet) for every ten miles per hour you are traveling". If we assume a jam spacing slightly greater than the vehicle length, say 20 ft, the value of C_j deduced from the above following rule is

$$C_j = -20\ \text{ft} \times \frac{10\ \text{mile/h}}{15\ \text{ft}} = -21\ \text{km/h}.$$

Lighthill and Whitham (1955) reported on the linearity of the speed–spacing relationship at high densities: "as V increases, the mean spacing increases almost linearly (by about 1.2 ft for each 1 mile/h increase in speed)". Since "spacing takes a value (around 17 ft in Great Britain) only just greater than the average vehicle length", the above considerations yield the following value of C_j

$$C_j = -17 \text{ ft} \times \frac{1 \text{ mile/h}}{1.2 \text{ ft}} = -23 \text{ km/h}.$$

Foote (1963) refers to a work of Daou on the formation of vehicle platoons in the Holland and Lincoln Tunnels under the Hudson River in New York City: "for constant speed platoons, Daou hypothesized a mean time headway versus mean speed relationship of $t_h = 1.6 + 25/V$". Taking into account that the time headway t_h is equal to H/V, one obtains

$$C_j = -\frac{25 \text{ ft}}{1.6 \text{ sec}} = -17 \text{ km/h}.$$

Edie and Baverez (1965) collected data in the Holland Tunnel to study the propagation of stop–start waves. The tunnel was divided into four sections and slightly different wave speeds were found for each one, −16 km/h being the mean value.

Dörfler (1965) studied the dynamics of a stopped platoon at a traffic light. He found that the elapsed time between the start of the green phase t_s, and the start of a vehicle was a linear function of the vehicle position in the platoon n_s. By applying a regression analysis to numerous observations he obtained

$$t_s \text{ (sec)} = -0.68 + 1.08 n_s,$$

with a correlation coefficient of 0.994. The linearity of this expression is undoubtedly due to the uniformity of the starting wave speed which is in turn a reasonable estimator of the kinematic wave speed at jam density. Taking a value of 5.6 m for the mean vehicle length plus bumper-to-bumper distance, as suggested by Dörfler, the above expression provides

$$C_j = -\frac{5.6 \text{ m}}{1.08 \text{ sec}} = -19 \text{ km/h}.$$

Lam and Rothery (1970) studied the propagation of speed fluctuations on a freeway. They calculated the time lag that maximized the cross-correlation function of the speed measured at various locations. The wave speed was estimated from the distance between the locations and the maximization time lag. This procedure yielded a value of 17.8 ft/sec, that is, $C_j = -19.4$ km/h.

Very recently Ozaki (1993) has reported on a series of speed–spacing observations on Tokyo freeways. These observations were performed under steady state conditions defined as the situation in which the acceleration or decceleration rate was within a range of ± 0.05 g. The speed–spacing plots show a linear portion up to 50 km/h. The wave speed estimated from these plots is about −17 km/h.

Hence numerous empirical evidences support our claim that the kinematic wave speed at jam density ranges between −17 and −23 km/h, although values slightly below this range have been found by Franklin (1965) and Duckstein *et al.* (1970). Franklin carried out some experiments on platoons to measure the propagation speed of starting and stopping waves. In these experiments the leader of the platoon was first instructed to maintain various constant speeds and then to perform certain speed changes. The wave speed and C_j have similar values because he found a linear volume–density relationship for vehicles in a platoon. He found, namely, an average value of 8.2 mile/h for the starting waves and 9 mile/h for the stopping waves. Thus $C_j = -13.1$ km/h and $C_j = -14.4$ km/h respectively. Duckstein *et al.* found as a first approximation to the equilibrium spacing $H = 18 + 1.3V$ that gives $C_j = -18$ ft/1.3 sec $= 13$ km/h.

On the other hand, Greenberg (1959) and Herman *et al.* (1971) found values of $|C_j|$ above 23 km/h. Greenberg obtained, from data collected in the Lincoln Tunnel in New York, the relation $V = -27.5 \ln K/142$ where the speed is in km/h and the density in vehicles per km and lane. From this relation, $C_j = -27.5$ km/h. Likewise, Herman and his coworkers performed a series of car-following experiments obtaining a similar value for the propagation speed of the starting wave in a platoon.

The overestimation of C_j is much more serious for the speed–density relationships given by Drake *et al.* (1967). Several models were fitted to a traffic data set, yielding values of $|C_j|$ always greater than 43 km/h. In the scatter diagram included in this paper, there is no data taken beyond 120 vehicles per mile and lane, which is likely to be the cause for the overestimation of C_j.

In the last 15 yr, several authors have proposed macroscopic traffic flow models that include speed–density relationships with unrealistic values of the kinematic wave speed at jam density. Payne (1979) took the speed–density curve given by the following expression

$$V_e = \min\left\{88.5,\ 88.5\left[1.94 - 6\left(\frac{K}{143}\right) + 8\left(\frac{K}{143}\right)^2 - 3.93\left(\frac{K}{143}\right)^3\right]\right\}$$

with K in veh/km and V in km/h, for the simulation program FREFLO. The above expression yields $C_j = -158$ km/h! Michalopoulos and Pisharody (1980) considered Greenshields speed–density curves with $V_f = 48$ km/h, therefore $C_j = -48$ km/h. In Michalopoulos *et al.* (1984) the same curve is used but this time with $V_f = 96.5$ km/h, therefore $C_j = -96.5$ km/h. Cremer and Papageorgiou (1981) assumed a curve of the form (1) with $n = 4$, $m = 1.4$. For such a curve, C_j is always zero provided that $n > 1$.

In this section, the existence of three fundamental parameters of traffic flow relationships has been established. These parameters are the jam density (K_j), the kinematic wave speed at jam density (C_j) and the free flow speed (V_f). This fact was already pointed out by Franklin (1965): "the flow–concentration relationship is determined by three main factors — the jam concentration, the slope of the steady-state part of the curve, and the maximum safe speed". It is surprising that the important work of Franklin has gone unnoticed for almost 30 yr. It is even more surprising that most of the work on macroscopic traffic flow modelling done during the last 10 yr has not taken into account the nature of the kinematic wave speed at jam density. In the next section, the functional form of the speed–density relationship is proposed. The trio of fundamental parameters of traffic flow plays a crucial role in the derivation of the functional form.

THE FUNCTIONAL FORM

In the last section, it has been shown that the jam density and the kinematic wave speed at jam density may be considered almost constant parameters for a given road and traffic flow composition. Therefore, it becomes obvious to nondimensionalize density and speed by taking these parameters as reference values. We then define

$$u = \frac{V}{|C_j|}, \qquad \rho = \frac{K}{K_j}, \qquad q = \frac{Q}{K_j\,|C_j|}.$$

The properties of the speed–density and the flow–density relationships become

$$u(0) = u_f \tag{6}$$

$$u(1) = 0 \tag{7}$$

$$u'(\rho) < 0, \qquad 0 < \rho < 1 \tag{8}$$

$$u'(0) = 0 \tag{9}$$

$$q''(\rho) < 0, \qquad 0 < \rho < 1 \tag{10}$$

where $u_f = V_f / |C_j|$ and the derivatives are with respect to the nondimensional density ρ. Moreover, due to the nondimensionalization the following property should be added

$$u'(1) = q'(1) = -1. \tag{11}$$

The problem is then to find a continuous and differentiable function $u(\rho)$ that satisfies the above properties. The naive solution to this problem would be to try a polynomial function of the form

$$u = u_f + a_2\rho^2 + a_3\rho^3 + \cdots + a_n\rho^n$$

and adjust the coefficients a_i so that all properties hold. For the typical values of u_f, 4–6, this task has turned out to be impossible, whereas for smaller values of u_f it leads to rather cumbersome conditions for the coefficients. We therefore believe that it is not possible to find such a polynomial function for relevant values of the free flow speed. Instead, the problem is simplified considerably if we restrict the search to functions having the following functional form

$$u = u_f \left\{ 1 - f\left[\frac{1}{u_f}\left(\frac{1}{\rho} - 1 \right) \right] \right\}. \tag{12}$$

Writing

$$\lambda = \frac{1}{u_f}\left(\frac{1}{\rho} - 1 \right) \tag{13}$$

we will have

$$u = u_f \left[1 - f(\lambda) \right].$$

The function $f(\lambda)$ will be called *generating function*. Its argument λ is a nondimensional rescaled spacing and so it will be named *equivalent spacing*.

Hence, the generating function must satisfy the following properties:

$$\lim_{\lambda \to \infty} f(\lambda) = 0 \tag{14}$$

$$f(0) = 1 \tag{15}$$

$$\dot{f}(\lambda) < 0, \qquad \lambda > 0 \tag{16}$$

$$\lim_{\lambda \to \infty} \lambda^2 \dot{f}(\lambda) = 0 \tag{17}$$

$$\ddot{f}(\lambda) > 0, \qquad \lambda > 0 \tag{18}$$

$$\dot{f}(0) = -1 \tag{19}$$

where ($\dot{}$) means derivative with respect to λ. In addition

$$0 < f(\lambda) < 1, \qquad \lambda > 0. \tag{20}$$

Taking into account that the generating function is continuous and differentiable, it is easy to prove that (14), (18) and (20) imply (16). Furthermore, (17) implies (14). But on the contrary, (14), (16) and (20) do not imply (18), since the function may start from $\lambda = 0$ with negative second derivative. Therefore, the above list of properties may be reduced to (14), (15), (17), (18), (19) and (20).

The conditions for the generating function are less restrictive than those for the original traffic flow relationships. Furthermore, it is easy, at least in theory, to set up a procedure for obtaining generating functions. Such a procedure would consist of the following steps:

(i) Choose a function $\ddot{f}(\lambda)$ with two adjustable parameters satisfying (18)

(ii) Integrate $\ddot{f}(\lambda)$ twice with the conditions

$$\int_0^{\infty} \ddot{f}(\lambda)\mathrm{d}\lambda = 1,$$

$$\int_0^{\infty} \dot{f}(\lambda)\mathrm{d}\lambda = -1,$$

$$\dot{f}(0) = -1, \qquad f(0) = 1$$

(iii) Check condition (17).

In practice, such a procedure is useless since the double integration of $\ddot{f}(\lambda)$ is a serious obstacle for its application. And even if the double integration could be done analytically, the resulting generating function would have, in general, a rather complicated dependence on λ.

The only applicable procedure we have found is merely "trial and error". The lack of a constructive method is not a good result. To overcome this inconvenience, we offer in the next section a "smorgasbord" of simple generating functions that provide the traffic engineer with a broad selection of speed–density curves.

EXAMPLES OF GENERATING FUNCTIONS

In this section, several expressions for generating functions are proposed. We have found up to four types of generating functions that we have called: exponential, double exponential, rational and reciprocal–exponential families. In these models, the generating functions are expressed in terms of elementary functions containing λ as a single argument, which is an important requirement for simplicity.

The family of exponential generating functions is given by

$$f(\lambda) = \exp\left[1 - \left(1 + \frac{\lambda}{n}\right)^n\right] \tag{21}$$

with $n > 0$. When $n = 1$, the above expression reduces to $f(\lambda) = \exp(-\lambda)$. The resulting speed–density curve is

$$V = V_f\left\{1 - \exp\left[\frac{|C_j|}{V_f}\left(1 - \frac{K_j}{K}\right)\right]\right\}. \tag{22}$$

As $n \to \infty$ the exponential generating function tends to the function

$$f(\lambda) = \exp\left[1 - \exp(\lambda)\right] \tag{23}$$

which is itself a generating function.

The double exponential family may be obtained by introducing a parameter in the above generating function

$$f(\lambda) = \exp\left\{n\left[1 - \exp\left(\frac{\lambda}{n}\right)\right]\right\}$$

with $n > 1$. The limit of these curves as $n \to \infty$ is $f(\lambda) = \exp(-\lambda)$.

The family of rational generating functions corresponds to the expression

$$f(\lambda) = \frac{1}{\left(\dfrac{\lambda}{n} + 1\right)^n}$$

with $n > 1$. The speed–density curve is then given by

$$u = u_f\left[1 - \left(\frac{nu_f\rho}{1 + (nu_f - 1)\rho}\right)^n\right]. \tag{24}$$

When $n \to \infty$, the above relationship approaches the one generated by the exponential function (22).

Finally, the reciprocal–exponential family has the expression

$$f(\lambda) = \frac{n}{\exp(n\lambda) + n - 1}$$

where $0 < n \leqslant 2$. For $n = 1$, we recover the exponential curve (22), and for $n = 2$, the function takes the more familiar form

$$f(\lambda) = 1 - \tan\text{h}\ (\lambda).$$

In Fig. 1, the speed–spacing curves generated by these families are depicted for several

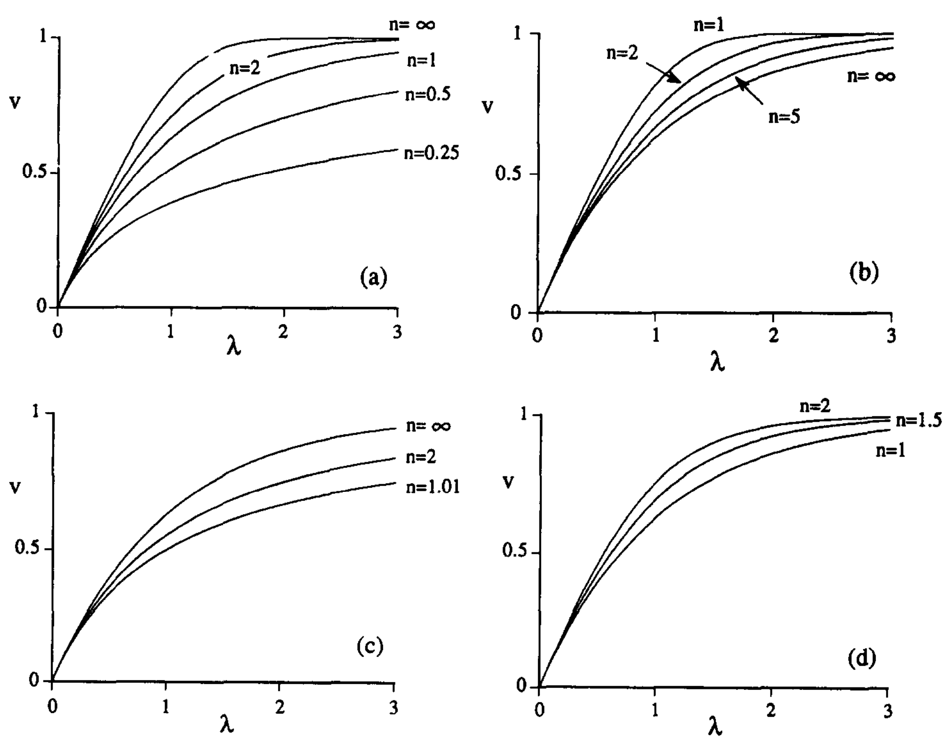

Fig. 1. Nondimensional speed (v) vs equivalent spacing (λ) for the exponential (a), double exponential (b), rational (c) and reciprocal–exponential (d) generating functions.

values of the parameters. The nondimensional speed v defined as

$$v = \frac{V}{V_f} = \frac{u}{u_f} = 1 - f(\lambda)$$

has been plotted against the equivalent spacing (λ). The rational curves always lie below the curve corresponding to $f(\lambda) = \exp(-\lambda)$, whereas the double exponential and reciprocal–exponential curves lie above it. The shape of these three families of curves can be very well reproduced by the exponential generating functions with $n < 1$ and $n > 1$ respectively. Thus, our interest will be focused on the exponential generating functions, since they exhibit a broader diversity of shape. Moreover, it is quite remarkable that all the $v - \lambda$ curves are bounded by the curve

$$v = 1 - \exp\left[1 - \exp(\lambda)\right],$$

which approaches to the idealized two-linear regime

$$\begin{aligned} v &= \lambda, & \lambda &< 1 \\ v &= 1, & \lambda &\geqslant 1. \end{aligned}$$

Only the speed–density curve given by (22) has been proposed before by several authors. No similar models to the rest of the speed–density curves have been found in the extensive literature reviewed for the present work. The exponential speed–density curve (22) has an interesting history. It was first formulated independently by Newell (1961) and Franklin (1961) as the steady state speed–density curve of the car-following models

$$V_i = V_f\left[1 - \exp\left(\frac{|C_j|}{V_f}\left(1 - K_j(X_{i-1} - X_i)\right)\right)\right] \qquad \text{(Newell)}$$

and

$$\frac{\mathrm{d}V_i}{\mathrm{d}t} = |C_j|\,K_j\left(1 - \frac{V_i}{V_f}\right)(V_{i-1} - V_i) \qquad \text{(Franklin)}$$

where the subscript i stands for the follower vehicle and $i - 1$ for the lead one. Newell takes (22) *a priori* simply because "it has approximately the correct shape and it is reasonably simple", whereas Franklin departs from the above car-following model since he finds that "a reasonable agreement with the measured acceleration characteristics of automobiles is obtained if it is assumed that the available acceleration decreases in proportion to the speed". The curve (22) is later mentioned in the work of Leutzbach and Bexelius (1966). This work was later mentioned in the book of Valdés (1971), from which we learnt about (22). Since then, the exponential speed–density curve disappears mysteriously from the traffic flow theory literature. That makes one wonder why something so simple and appealing has gone unnoticed for more than 25 yr, or if it was known, why it was not exploited.

THE DERIVATION OF THE FUNCTIONAL FORM

In this section, we will derive the functional form (12) of the speed–density curve through two different arguments: one of a mathematical type and one based in the behavior of drivers.

Mathematical argument

The mathematical argument justifies why the proposed functional form (12) simplifies the problem of finding an analytical expression for $u(\rho)$. The more restrictive conditions of the traffic flow relationships are the global ones, that is (8) and (10). Therefore, we would be interested in finding a transformation $\lambda = \lambda(\rho)$ that reduced these two conditions to only one, or in the worst case to a combination of them. Then we may write without loss of generality

$$u = u_f\Big\{1 - f\big[\lambda(\rho)\big]\Big\}, \qquad q = \rho u(\rho).$$

Differentiation with respect to ρ yields

$$u'(\rho) = -u_f \dot{f}(\lambda)\lambda'$$

$$q''(\rho) = -u_f\big[\rho\ddot{f}(\lambda)\lambda'^2 + (\rho\lambda'' + 2\lambda')\dot{f}(\lambda)\big],$$

where

$$\lambda' = \frac{d\lambda}{d\rho}.$$

The choice of a function $\lambda(\rho)$ satisfying

$$\rho\lambda'' + 2\lambda' = 0 \tag{25}$$

eliminates the dependence of $q''(\rho)$ on $\dot{f}(\lambda)$ and, in principle, may simplify the problem. The integration of the above differential equation with the conditions

$$\lambda(\rho_0) = \lambda_0, \qquad \lambda'(\rho_0) = \lambda'_0$$

yields

$$\lambda - \lambda_0 = \lambda'_0\rho_0^2\left(\frac{1}{\rho_0} - \frac{1}{\rho}\right).$$

Hence

$$\lambda' = \lambda'_0\frac{\rho_0^2}{\rho^2}$$

and the condition (11) forces

$$-u_f\dot{f}\big[\lambda(1)\big]\lambda'_0\rho_0^2 = -1. \tag{26}$$

Thus the transformation is given by

$$\lambda - \lambda_0 = -\frac{1}{u_f\dot{f}\big[\lambda(1)\big]}\left(\frac{1}{\rho} - \frac{1}{\rho_0}\right).$$

The choice of ρ_0, λ_0 and λ'_0 that leads to the equivalent spacing (13) is obviously: $\rho_0 = 1$, $\lambda_0 = 0$ and $\lambda'_0 = -1/u_f$. Then, the generating function $f(\lambda)$ should satisfy the conditions (14) to (20). Another choice of ρ_0, λ_0 and λ'_0 would lead to a generating function satisfying a different set of conditions. The choice of ρ_0, λ_0 and λ'_0 is irrelevant since one can always define a new equivalent spacing as $\bar{\lambda}$ as

$$\bar{\lambda} = \Big\{-\rho_0\dot{f}\big[\lambda(1)\big]\Big\}(\lambda - \lambda_0).$$

Thus

$$\bar{\lambda} = \frac{1}{u_f}\left(\frac{1}{\bar{\rho}} - 1\right)$$

with

$$\bar{\rho} = \rho/\rho_0$$

and with the generating function $\bar{f}(\bar{\lambda}) = f(\lambda)$ we recover the conditions (14) to (20). The transformed equivalent spacing $\bar{\lambda}$ satisfies the eqn (25) rewritten in terms of the transformed variables

$$\bar{\rho}\bar{\lambda}'' + 2\bar{\lambda}' = 0$$

and the simplification of $q''(\rho)$ is carried over. In other words, for a given choice of ρ_0, λ_0 and λ_0' it is always possible to obtain a "standard" generating function, $\bar{f}(\bar{\lambda})$, by a shift and a rescaling of the equivalent spacing. The setting of $\rho_0 = 1$ and $\lambda_0' = -1/u_f$ is a smart decision because it eliminates the parameter u_f in (26). The setting of $\lambda_0 = 0$ is completely arbitrary, but it is nicer when things start from zero rather than from a fancy number.

In conclusion, the functional form (12) is the only one that eases the search of an expression for the speed–density curve since it simplifies the nature of the required global conditions for the generating function. For the speed–density curve, the global conditions are (8) and (10), whereas for the generating function they have the simpler forms (18) and (20), since (16) may be deduced from the rest of conditions.

Drivers' behavior based argument

In what follows, we give a phenomenological interpretation of the functional form in terms of the behavior of drivers. The speed–density relationship may be regarded as the equilibrium solution of a car following model. A generic following law may be in turn written in nondimensional form as follows

$$\frac{du_v}{d\tau} = s^*\ (u_v,\ h_v,\ w;\ u_f)$$

where

$$w = \frac{dh_v}{d\tau}$$

and $\tau = t\,|\,C_j\,|\,K_j$ is the nondimensional time. The subscript v refers to individual or vehicle variables. In particular, $h_v = HK_j$ is the nondimensional spacing. In the above model, it has been assumed that a driver responds only to the vehicle immediately ahead. This assumption is widely used in car-following theory and it is commonly admitted as a reasonable approximation to the actual behavior. Therefore, this assumption does not represent a serious loss of generality.

In situations close to equilibrium, the traffic flow may be thought to be characterized by mean variables and small fluctuations about the mean variables. Further, the mean variables, h and u, will be related through $u = u(h)$. Then, the spacing and speed may be written in the following form

$$h_v = h + \epsilon h_1 \tag{27}$$

$$u_v = u + \epsilon u_1 \tag{28}$$

with $\epsilon \ll 1$, and h_1, u_1 being the fluctuations. If the function s^* is expanded in its Taylor series about the equilibrium point $(u, h, 0; u_f)$, one gets

$$\frac{du_v}{d\tau} \approx s^*(u,\ h,\ 0;\ u_f) + \epsilon \left.\frac{\partial s^*}{\partial h}\right|_0 h_1 + \epsilon \left.\frac{\partial s^*}{\partial u}\right|_0 u_1 + \left.\frac{\partial s^*}{\partial w}\right|_0 w + \cdots$$

The above approximation is valid for conditions close to equilibrium. The subscript '0' indicates that the derivatives are evaluated at the equilibrium point. If the model is to make sense one should have $s^*(u,\ h,\ 0;\ u_f) = 0$. Finally, introducing expressions (27) and (28) in the above car-following model and equating terms of the order of magnitude ϵ^0, one obtains

$$\frac{du}{d\tau} = \left.\frac{\partial s^*}{\partial w}\right|_0 \frac{dh}{d\tau}.$$

By taking into account that h is a function of u, and writing

$$s(u;\, u_f) = \left.\frac{\partial s^*}{\partial w}\right|_0$$

one arrives at the equation for the equilibrium speed–spacing relationship

$$\frac{\mathrm{d}u}{\mathrm{d}h} = s(u;\, u_f).$$

The function s is the sensitivity of the drivers with respect to the relative speed. The mean spacing and speed h and u become then the equilibrium spacing and speed.

In principle the sensitivity may depend separately on the speed and the nondimensional free flow speed u_f. But this assumption would imply that the sensitivity depends on the kinematic wave speed at jam density. It is not reasonable to suppose that the sensitivity of an individual driver depends on a macroscopic property such as the kinematic wave speed at jam density. A macroscopic property is the manifestation of the behavior of a group of vehicles, and not of a single one. Furthermore, it is absurd to admit that the drivers are aware of the waves they generate when accelerating or decelerating: not even traffic engineers are! Instead, the free flow speed should be an essential component of the driver's sensitivity since each driver is aware of the speed at which he would like to travel if he were not obstructed by other vehicles. Hence, the dependence of the sensitivity on C_j must be eliminated and the only manner to do it is by supposing that

$$s(u;\, u_f) = s(u/u_f) = s(V/V_f).$$

Then the integration of the equation

$$\frac{\mathrm{d}u}{\mathrm{d}h} = s(u/u_f), \tag{29}$$

with the condition $u(h = 1) = 0$ yields the solution

$$R(u/u_f) - R(0) = \frac{1}{u_f}\,(h - 1)$$

which written in terms of the nondimensional density ($h = 1/\rho$) leads to the proposed functional form (12), since the constant $R(0)$ is irrelevant.

The function s is actually the nondimensional sensitivity and its expression as a function of v is obtained from (29)

$$s(v) = \frac{\mathrm{d}u}{\mathrm{d}h} = \frac{\mathrm{d}v}{\mathrm{d}\lambda} = -\dot{f}(\lambda) = -\dot{f}\left[f^{-1}(1 - v)\right].$$

The physical sensitivity S is then given by

$$S(v) = \frac{\mathrm{d}V}{\mathrm{d}H} = \frac{|C_j|}{H_j}\, s(v). \tag{30}$$

It may seem that we have cheated when deriving the expression of the functional form: we eliminated the dependence of the nondimensional sensitivity on C_j, but now we get this parameter in the actual sensitivity S. However, $|C_j|$ does not appear alone but in the expression $|C_j|/H_j$. This expression is, in fact, a microscopic quantity since it is a measure of the reaction time of drivers. Thus, expression (30) does make sense because all the involved parameters are drivers' behavior characteristics: the desired speed in absence of vehicles and the reaction time.

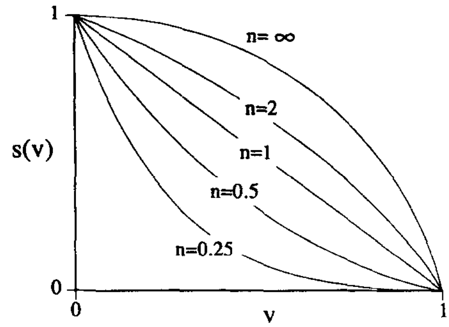

Fig. 2. Sensitivity curves given by the exponential generating functions.

The sensitivity is a decreasing function of the speed since

$$\frac{\mathrm{d}S}{\mathrm{d}V}=\frac{|C_j|}{H_j}\frac{\mathrm{d}^2u}{\mathrm{d}h^2}\frac{\mathrm{d}h}{\mathrm{d}V},$$

where the first factor is negative due to the concavity of the flow–density curve (18), and the second factor is positive. For the same reasons, the sensitivity is an increasing function of the free flow speed

$$\frac{\mathrm{d}S}{\mathrm{d}V_f}=-\frac{V}{V_f}\frac{\mathrm{d}S}{\mathrm{d}V}.$$

For the family of exponential generating functions we get:

$$s(v)=(1-v)\left[1-\ln(1-v)\right]^{(1-\frac{1}{n})}.$$

For the special case $n = 1$, corresponding to the exponential speed–density curve (22), the sensitivity becomes a linear function of the speed: $s(v) = 1 - v$. Values of n greater than 1 render sensitivity curves lying above the linear sensitivity curve, whereas the sensitivity curves for $n < 1$ lie below it, as depicted in Fig. 2. The upper limit of all the sensitivity curves is

$$s(v)=(1-v)\left[1-\ln(1-v)\right], \tag{31}$$

corresponding to the speed–density curve generated by (23). For this reason, we will call it the *maximum sensitivity curve*.

The sensitivity curves corresponding to the double exponential, rational and reciprocal–exponential generating functions are respectively

$$s(v)=(1-v)\left[1-\frac{\ln(1-v)}{n}\right]$$

$$s(v)=(1-v)^{(1+\frac{1}{n})};$$

$$s(v)=(1-v)\left[n+(1-n)(1-v)\right].$$

For all the values of the parameters n, the rational sensitivity curves lie below the linear one. The double exponential reciprocal–exponential sensitivity curves are confined between the linear one and (31).

CONCLUSIONS

The main contribution of this work is the formulation of a functional form for the speed–denstity relationship. This functional form is embedded in a nondimensional spacing, that we have called *equivalent spacing*. The specific dependence on the equivalent spacing is given by the *generating function*.

The derivation of the functional form has been based on two different arguments. The first one is simply a mathematical argument. The rationale for this argument arises from the set of properties that the volume–speed–density relationship should satisfy. It has been demonstrated that the proposed functional form is the only one that simplifies the problem of finding a continuous and differentiable function satisfying this set of properties.

The second argument is based on the drivers' behavior. The functional form has been derived by means of a dimensional analysis of a generic car-following model. Prior to this analysis, the set of fundamental or dimensional parameters of traffic flow has been identified. These parameters are the free flow speed, V_f, the jam density, K_j, and the kinematic wave speed at jam density, C_j. The key of the dimensional analysis is the rejection of any arbitrary functional form for the sensitivity of drivers. The sensitivity can only depend on a specific combination of these parameters, and this combination leads to the functional form for the speed–density relationship.

Finally we give several examples of generating functions. There are two generating functions which are noteworthy. They originate the exponential and maximum sensitivity speed–density curves. The exponential curve has the charm and simplicity of the exponential function. The maximum sensitivity curve yields the maximum value of the sensitivity of drivers.

The extensive literature reviewed for this work has lead to a surprising result: of all the speed–density curves proposed in the last three decades, only one has the functional form presented here. It is, namely, the exponential curve, which was mentioned in a few works 25 yr ago.

Acknowledgements—J. M. Del Castillo thanks the Dirección General de Universidades of the Andalusian regional government for the research grant that has made this work possible. The authors would also like to thank two anonymous referees for the helpful suggestions and corrections on an earlier draft of the paper.

REFERENCES

Ansorge R. (1990) What does the Entropy Condition mean in Traffic Flow Theory? *Transpn. Res.* **24B**, 133–143.

Ceder A. (1976) A Deterministic Traffic Flow Model for the Two-regime Approach. *Transpn. Res. Rec.* **567**, 16–30.

Cremer M. and Papageorgiou M. (1981) Parameter Identification for a Traffic Flow Model. *Automatica* **17**, 837–843.

Dörfler W. (1965) Fahrdynamik an signalgesteuerten Kreuzungen. *Straßenbau und Straßenverkehrstechnik* **38**.

Drake J. S., Schofer J. L. and May A. D. (1967) A Statistical Analysis of Speed-Density Hypotheses. *Highw. Res. Rec.* **156**, 53–87.

Duckstein L., Unwin E. A. and Boyd E. T. (1970) Variable Perception Time in Car Following and its Effect on Model Stability. *IEEE Trans. on Man-Machine Sys.* **3**, 149–156.

Easa S. M. (1983) Selecting Two-regime Traffic Flow Models. *Transpn. Res. Rec.* **869**, 25–36

Edie L. C. (1961) Car-Following and Steady State Theory for Non-congested Traffic. *Ops. Res.* **9**, 66–76.

Edie L. C. and Baverez E. (1965). Generation and Propagation of Stop-Start Waves. *Proc. 3rd Int. Symp. on the Theory of Traffic Flow*, Elsevier, New York, 26–37.

Foote R. S. (1963) Single Lane Traffic Flow Control. *Proc. 2nd Int. Symp. on Transportation and Traffic Theory* 84–103.

Forbes T. W. and Simpson M. E. (1968) Driver and Vehicle Response in Freeway Decceleration Waves. *Transpn. Sci.* **2**, 77–104.

Franklin R. E. (1961) The Structure of a Traffic Shock Wave. *Civ. Engng. publ. Wks. Rev.* **56**, 1186–1188.

Franklin R. E. (1965) Further Analysis on some Experiments on Single-Lane Traffic Flow. *Road Res. Lab.* Note No. 829.

Gerlough D. L. and Huber M. J. (1975) Traffic Flow Theory — A Monograph. *Transpn. Res. Board*, Special Report **165**, Washington D. C.

Greenberg H. (1959) An Analysis of Traffic Flow. *Ops. Res.* **7**, 79–85.

Greenshields B. D. (1935) A Study in Highway Capacity. *Highw. Res. Board Proc.* **14**, 448–477.

Hall F. L., Allen B. L. and Gunter M. A. (1986) Empirical Analysis of Freeway Flow–Density Relationships. *Transpn. Res.* **20A,** 197–210.

Herman R., Lam T. and Rothery R. W. (1971) The Starting Characteristics of Automobile Platoons. *Proc. 5th Int. Symp. on Transportation and Traffic Theory*, Elsevier, 1–17.

Koshi M., Iwasaki M. and Ohkura I. (1983) Some Findings and an Overview on Vehicular Flow Characteristics. *Proc. 8th Int. Symp. on Transportation and Traffic Theory*, University of Toronto Press, Toronto, Ontario, 403–426.

Lam T. and Rothery R. (1970) The Spectral Analysis of Speed Fluctuations on a Freeway. *Transpn. Sci.* **4,** 293–310.

Leutzbach W. and Bexelius S. (1966) Probleme der Kolonnenfahrt. *Straßenbau und Straßenverkehrstechnik*, **44**.

Lighthill M. J. and Whitham G. B., (1955) On Kinematic Waves: a Theory of Traffic Flow on Long Crowded Roads. *Proc. of the Royal Society*, Series A, **229**, 317–345.

Michalopoulos P. G. and Pisharody V. B., (1980) Platoon Dynamics on Signal Controlled Intersections. *Transpn. Sci.* **14**, 365–96.

Michalopoulos P. G., Beskos D. E., and Yamauchi Y. (1984) Multilane Traffic Flow Dynamics: some Macroscopic Considerations. *Transpn. Res.* **18B**, 377–395.

Newell G. F. (1961) Nonlinear Effects in the Dynamics of Car-Following. *Ops. Res.* **9**, 209–229.

Ozaki H. (1993) Reaction and Anticipation in the Car-Following Behavior. *Proc. 12th Int. Symp. on Transportation and Traffic Theory*, Elsevier, 349–366.

Payne H. J. (1979) FREFLO: a Macroscopic Simulation Model of Freeway Traffic. *Transpn. Res. Rec.* **772,** 68–75.

Payne H. J. (1984). Discontinuity in Equilibirum Freeway Traffic Flow. *Transpn. Res. Rec.* **971**, 140–146.

Pipes L. A. (1953) An Operational Analysis of Traffic Dynamics. *J. of appl. Phys.* **24**, 274–287.

Pipes L. A. (1967) Car-Following Models and the Fundamental Diagram of Road Traffic. *Transpn. Res.* **1**, 21–29.

Prigogine I. and Herman R. (1971) *Kinetic Theory of Vehicular Traffic*. Elsevier, New York.

Schaar (1925). Die Leistungsfähigkeit einer Straße für den Kraftverkehr. *Verkehrstechnik*, **23**.

Underwood R. T. (1961) Speed, Volume and Density Relationships, Quality and Theory of Traffic Flow. Yale Bureau of Highway Traffic, New Haven, Connecticut, 141–188.

Valdés A. (1971) *Ingeniería de Tráfico*. Dossat, Madrid.

Whitham G. B. (1974) *Linear and Nonlinear Waves*. John Wiley & Sons, New York.