

June 10, 1999

The Theory of Caustics and Wavefront Singularities with Physical Applications

Juergen Ehlers¹ and Ezra T. Newman²

¹Max-Planck-Institut für Gravitationsphysik,
Albert Einstein Institut, Schlaatzweg 1,
D-14473, Potsdam, Germany

²Dept. of Physics and Astronomy,
Univ. of Pittsburgh, Pgh. PA. 15260 USA

Abstract

This is intended as an introduction to and review of the work of V. I. Arnold and his collaborators on the theory of Lagrangian and Legendrian submanifolds and their associated maps. The theory is illustrated by applications to Hamilton-Jacobi theory and the eikonal equation, with an emphasis on null surfaces and wavefronts and their associated caustics and singularities.

I. Introduction

The following paper is intended to be an introduction to the theory of smooth Lagrangian and Legendrian maps from manifolds into manifolds, with a wide range of examples from physics; Hamilton-Jacobi theory, the theory of the eikonal equation, wave-fronts, their singularities, caustics, etc. Our interest in this subject arose out of efforts to understand the beautiful ideas of V.I. Arnold^{1,2,3,4,5,6} and his collaborators concerning the theory of singularities of maps. This effort, in turn, was motivated originally by a recent reformulation of General Relativity^{7,8} in terms of families of null hypersurfaces which naturally necessitated a study of the pertinent singularities associated with null hypersurfaces. Another reason for studying this theory was our interest in gravitational lens theory^{9,10} and Ya. B. Zeldovich's theory of structure formation in the early universe.¹¹

As Arnold's treatment is much more general than most physicists need or use and his approach is often quite abstract, we and many colleagues found it initially difficult to get the essential overall picture of this remarkable theory. Eventually, to a large extent, the picture did get clarified and we thought that an elementary presentation, from a physicist's view, of these ideas might be of use to others who do not have the patience to struggle through Arnold's beautiful work.^{1,2,3,4}

This work is organized as follows. In Sec.II, we discuss the basic geometric ideas behind the local theory of caustics and wave front singularities^{3,12,6,13} based on the construction of Lagrangian and Legendrian submanifolds in phase space via the use of *generating functions* - along with some simple illustrative examples. In Sec.III we give a particularly instructive example from the theory of the Hamilton-Jacobi equation. Sec.IV serves to establish various relations of the eikonal equation and its solutions to (small and large) wave fronts in arbitrary space-times. After a brief discussion, in Sec.V, of some technical difficulties which emerge from the use of generating functions in the detailed implementation of the theory, we proceed, in Sec.VI, to the core of the method, showing how the concept of *generating families* (for the construction of Lagrangian and Legendrian submanifolds) naturally arises and is used to overcome some of the aforementioned difficulties. Finally, in Sec.VII, we apply this method to wavefronts in $(3 + 1)$ - dimensional space-time, to ensembles of non-interacting particles in phase space and to gravitational lensing.

The main message we learned from this powerful theory that we wish to convey to the reader is: in order to describe wave propagation phenomena in space or space-time - (processes which generically lead to intersections of rays, focal points or focal lines, sharp edges of wavefronts, infinite densities or similar “catastrophes”) - it is advisable to treat the evolution of the requisite structures not directly in space (space-time), but lift them to a suitable bundle over space (space-time), evolve them there, and at the end project the result down into space (space-time). In this way one can describe the singularities which occur “downstairs” in terms of smooth regular structures “upstairs”. What might appear at first sight as a complication turns out to be, in fact, a simplification. While the general definitions and constructions in Section II are based exclusively on the differential topology of a manifold, the physical examples employ, as an additional structure, a (generalized) Hamiltonian which defines a hypersurface in the bundle space. This leads to the restricted class of those Lagrangian/Legendrian submanifolds which are contained in those hypersurfaces and which are ruled by the phase space trajectories. These particular submanifolds are distinguished from general ones by a certain rigidity: pieces of them can be continued uniquely by those trajectories. In accordance with this, in the applications, besides having to satisfy a certain rank condition, the generating families have to satisfy a first order partial differential equation, e.g. the Hamilton-Jacobi or eikonal equation.

Except for the manner of presentation, we do not claim any originality here and any errors are ours. Though of course there is an overall unity to the subject, in this elementary treatment we have tried, especially in Sec. II, to keep separate ideas apart. We have denoted a new idea or topic by a \blacklozenge . Examples are denoted by a \spadesuit .

II. Lagrangian and Legendrian Submanifolds of Symplectic and Contact Bundles

We begin with an arbitrary, n -dimensional manifold M to be considered as a configuration space, with local coordinates q^a .

\blacklozenge 1. Consider the cotangent bundle T^*M (or phase space) over M with fiber coordinates p_b , and with associated symplectic potential κ and two-form ω

$$\kappa = p_a dq^a, \quad \omega = dq^a \wedge dp_a = -d\kappa. \quad (1)$$

We will refer to such a $2n$ -dimensional symplectic manifold also as M_S .

The two-form ω plays a somewhat similar role in symplectic geometry as the metric g in Riemannian or Lorentzian geometry. As both ω and g are non-degenerate, their inverses exist and, respectively, can be used to lower or raise indices. Metric orthogonality, $g(X, Y) = 0$, corresponds to skew-orthogonality, $\omega(X, Y) = 0$. We shall occasionally make use of the latter relation.

◆2. Besides T^*M , we shall use $T^*M \times \mathbb{R}$ with coordinates (q^a, p_a, u) and (by definition) the “contact” one form

$$\alpha = du - p_a dq^a. \quad (2)$$

We call the “contact manifold”, $(T^*M \times \mathbb{R}, \alpha)$, the “contactification” of T^*M and sometimes use the shorthand M_C for it. A function U on M defines a section of M_C (considered as a bundle with $(n + 1)$ -dimensional fibers and base M) via $u = U(q^a)$, $p_a = \partial_a U$. On such a section, $\alpha = 0$.

[Note that though this construction yields a particular example of a contact manifold, *locally* all contact manifolds can be given this structure.]

◆3. We thus have the “extension” of the $2n$ -dimensional symplectic bundle to the $(2n + 1)$ -dimensional contact bundle. Alternatively one can start with a $(2n + 2)$ -dimensional symplectic bundle and “reduce” it to a $(2n + 1)$ -dimensional contact bundle. [See Remark 5a, at the end of this section.]

◆4. Let M_S be a symplectic manifold. An immersed submanifold, L of M_S is called Lagrangian if it is n -dimensional and if the pull-back of ω to L vanishes. A submanifold of M_S is Lagrangian if and only if its tangent spaces are skew-orthogonal to themselves and have maximal dimension.

♠ a. As a simple example we can construct a Lagrangian submanifold in the following manner; Choose a “generating” function $F = F(q^a)$, then consider the n q^a ’s as the parameters used to parametrically describe an n -manifold L in the $2n$ -dimensional symplectic space by

$$\begin{aligned} p_a &= \partial_a F(q), \\ q^a &= q^a. \end{aligned} \quad (3)$$

One sees immediately, that on L ,

$$\omega = dq^a \wedge dp_a = (\partial_a \partial_b F) dq^a \wedge dq^b \equiv 0.$$

Alternately one could chose $G = G(p)$ as a generating function and define a Lagrangian submanifold by

$$\begin{aligned} q^a &= -\partial^a G(p), \\ p_a &= p_a, \end{aligned}$$

with the notation $\partial^a \equiv \partial/\partial p_a$. In particular, each fiber is a Lagrangian submanifold.

In contrast to the first example, Eq.(3), the new L will in general not be a section of the bundle M_S . Its projection to M need not be everywhere a local diffeomorphism. The resulting singularities will occupy us extensively below.

Other choices include interchanging some of the p 's and q 's in the generating function; for example, let $G = G(p_1, q^2, \dots, q^n)$ with

$$\begin{aligned} q^1 &= -\partial^1 G, \\ p_i &= \partial_i G, \\ p_1 &= p_1, \\ q^i &= q^i, \end{aligned} \tag{4}$$

$i = 2, \dots, n$. ♠

In general there are 2^n different local representations of Lagrangian submanifolds in terms of canonical coordinates. To construct them we divide the set of integers $(1, \dots, n)$ into two disjoint sets with \widehat{A} integers in the first set and \widehat{J} integers in the second set (with $\widehat{A} + \widehat{J} = n$). We then choose \widehat{A} different q 's, i.e., (q^A) and \widehat{J} different p 's, i.e., (p_J) . A generating function is then chosen as $K = K(q^A, p_J)$ and a Lagrangian submanifold is given by

$$\begin{aligned} q^J &= -\partial^J K, \\ p_A &= \partial_A K, \\ p_J &= p_J, \\ q^A &= q^A. \end{aligned} \tag{5}$$

Note that there is never a canonically conjugate pair in the set (q^A, p_J) .

Though it is clear that a submanifold constructed as in Eq.(5) is Lagrangian the converse statement that any Lagrangian submanifold can locally be constructed in this manner must be proved. We now give a derivation of this result.

Though the derivation is not difficult, it does get complicated and the reader might want to skip over the details and simply accept the result or return to the proof later.

Proof: Let L be a Lagrangian submanifold of M_S , with ξ a point on L , and let (p_a, q^a) be a canonical coordinate system. Since L is an immersed submanifold, there exists a subset of n elements of the set (p_a, q^a) , say $v^a \equiv (p_i, q^{i'})$, that provides local coordinates for L near ξ so that L can be represented by $w^a = f^a(v^b)$, with w^a being the remaining n elements of (p_a, q^a) .

The derivation will consist of two parts; we first show that such a subset v^a can always be chosen so that it does not contain a canonical pair p_j, q^j . Then we show that a generating function can be chosen such that locally L is given by Eq.(5).

Instead of giving a proof for arbitrary dimension n of M we take $n = 4$ as a representative (and in physics an important) case. The argument will show how to proceed in general. If the set v^a *does* contain a canonically conjugate pair, we will refer to it as an “unwanted” set; if it does not contain a conjugate pair it will be referred to as a “desired” set. Then, it will obviously suffice to prove: If local coordinates on L are given in the first place in one of the “unwanted” forms

$$(i) v^a = (p_1, q^1, q^2, q^3), \quad (ii) v^a = (p_1, q^1, p_2, q^3), \quad (iii) v^a = (p_1, q^1, p_2, q^2),$$

then one can always transform to a system of the “desired” form.

Consider case (i). Then, near ξ on L , $dp_1 \wedge dq^1 \wedge dq^2 \wedge dq^3 \neq 0$. Since L is Lagrangian, we have on L , $dp_a \wedge dq^a = 0$ (summation convention used). Therefore,

$$dp_1 \wedge dq^1 \wedge dq^2 \wedge dq^3 + dp_4 \wedge dq^4 \wedge dq^2 \wedge dq^3 = 0;$$

and hence the second term, like the first one, does not vanish at ξ . Our assumption (i) implies that, on L , near ξ , $p_4 = f(p_1, q^1, q^2, q^3)$; therefore from

$$dp_4 = \frac{\partial f}{\partial p_1} dp_1 + \frac{\partial f}{\partial q^1} dq^1 + \frac{\partial f}{\partial q^2} dq^2 + \frac{\partial f}{\partial q^3} dq^3 \quad (6)$$

either $\frac{\partial f}{\partial p_1} \neq 0$ and $dp_1 \wedge dq^2 \wedge dq^3 \wedge dq^4 \neq 0$, or $\frac{\partial f}{\partial q^1} \neq 0$ and $dq^1 \wedge dq^2 \wedge dq^3 \wedge dq^4 \neq 0$. From the implicit function theorem, $p_4 = f(p_1, q^1, q^2, q^3)$ can be inverted so that in the former case, (p_1, q^2, q^3, q^4) are a “desired” set, while in the later case (q^1, q^2, q^3, q^4) form the “desired” set.

In case (ii), the Lagrange condition gives

$$dp_1 \wedge dq^1 \wedge dp_2 \wedge dq^3 + dp_4 \wedge dq^4 \wedge dp_2 \wedge dq^3 = 0. \quad (7)$$

Reasoning as above one eliminates in the second product dp_4 in favor of dp_1 or dq^1 , obtaining in each case a “desired” set.

In the third case (iii) one gets

$$dp_1 \wedge dq^1 \wedge dp_2 \wedge dq^2 + dp_3 \wedge dp^3 \wedge dp_2 \wedge dq^2 + dp_4 \wedge dq^4 \wedge dp_2 \wedge dq^2 = 0,$$

so either the second or the third term is nonzero at ξ . Applying again the former reasoning to dp_3 or dp_4 , respectively, one reduces case (iii) to one of the other cases.

One can thus always transform an “unwanted” set to a “desired” set.

Accepting now, for an arbitrary n , the existence of a subset (p_J, q^A) *without* canonical pairs which provides local coordinates on L near ξ , we use the condition $\omega|_L = 0$ in the form

$$0 = d\kappa = d(p_J dq^J + p_A dq^A) = d(-q^J dp_J + p_A dq^A).$$

Thus, there exists locally near ξ a function $K(p_J, q^A)$ such that, on L ,

$$\kappa|_L = -q^J dp_J + p_A dq^A = dK,$$

which means that L is given locally by the n equations, (5), namely

$$q^J = -\partial^J K, \quad p_A = \partial_A K. \quad \blacksquare \text{Q.E.D.} \quad (8)$$

- We note that a generating function can be defined as a potential for the pull-back of κ to a Lagrangian submanifold, $\kappa|_L = dK$.

- Given a point ξ on L , only some the 2^n representations will be valid in its neighborhood. If, for example, $n = 2$ and (q^1, q^2) as well as (q^1, p_2) are permissible, we have

$$dK = p_1 dq^1 + p_2 dq^2 \text{ and } dG = p_1 dq^1 - q^2 dp_2,$$

and the change from (q^1, q^2, K) to (q^1, p_2, G) is a Legendre transformation,

$$K(q^1, q^2) = G(q^1, p_2) + q^2 p_2, \quad p_2 = \partial_2 K.$$

Globally, L can be “given” in terms of an atlas of overlapping charts, each with a representation of the form, Eq.(8), “Legendre-related” in the overlap regions.

- Note also that if we have an invertible transformation, $y^a = Y^a(q^A, p_J)$, the Lagrangian submanifold can be parametrized by the y^a . In applications, this type of situation, where the Lagrangian submanifold is parametrized by coordinates other than the (q^A, p_J) , is very common. In particular it plays a major role in the discussion of Sec. VI and VII, on generating families where the parameters y^a have a physical significance.

♠*b.* We give a simple example of a Lagrangian submanifold generated by a double-valued “function” that is not smooth. The same submanifold can be generated by a smooth (single-valued) function. Consider \mathbb{R} as a configuration space with the generating function

$$F = \pm q^{3/2}$$

so that

$$p = \partial F / \partial q = \pm \frac{3}{2} q^{1/2} \Rightarrow q = \frac{4}{9} p^2.$$

Note that the second derivative of F at $q = 0$ does not exist. The same Lagrangian submanifold is given by the generating function

$$G = -\frac{4}{27} p^3,$$

$$q = -\partial G / \partial p = \frac{4}{9} p^2 \Rightarrow F = pq + G = \frac{8}{27} p^3.$$

and is parametrized by p instead of q . The projection to the base is given by $q = \frac{4}{9}p^2$, with the “critical” point at $p = 0$. ♠

◆5. An important issue is the mapping from an n -dimensional Lagrangian submanifold, L , to the corresponding n -dimensional base space M . This projection, π , is given locally (from Eq. (5)) by

$$\pi : (q^A, p_J) \longmapsto \{q^A, q^J = -\partial^J K(q^A, p_J)\} .$$

For most cases of interest the mapping π is, for almost all points, a diffeomorphism (one-to-one and smooth in both directions). This is the case whenever L is transversal to the fibers. L may, however, have points where the Jacobian matrix (the derivative of π) is degenerate, i.e., has rank lower than n . These are the critical points of π which form the critical set, $CritL$; the image of $CritL$ in M is the caustic set, $\pi(CritL) = CaustL$. In terms of the preceding representation of L the critical points are given by

$$\det(\partial^J \partial^{J'} K) = 0; \tag{9}$$

what matters here is the p_J -dependence of K . Sard’s theorem states that the caustic set has Lebesgue measure zero; the critical set may however, have positive measure.

Note that the amount by which the rank of the Jacobian matrix, $\mathfrak{J} = \pi_*$, drops at critical points is equal to the corresponding decrease in rank for $(\partial^J \partial^{J'} K)$. This integer is an invariant of π , equal to the dimension of the kernel of \mathfrak{J} , i.e., the subspace of the tangent space of L which is annihilated by the projection. The kernel is given by the solutions X_J of $(\partial^J \partial^{J'} K) X_J = 0$.

- Given a point ξ on a Lagrangian submanifold one can choose the coordinate system, (q^a) , near $\pi(\xi)$ such that a “desired” coordinate system has $\widehat{A} = rank \mathfrak{J}$, $\widehat{J} = dim \ker \mathfrak{J}$. The corresponding representation, Eq.(8), contains the largest number of q^j s which is possible at ξ . Then, $\partial^J \partial^{J'} K = 0$ at ξ , and the kernel of \mathfrak{J} is spanned by ∂^J . Such representations are used to give canonical forms of generating functions near singularities of the Lagrangian maps.

♠.c A simple but important example of a Lagrangian submanifold will now be constructed and analyzed. It shows why one introduces bundles and their projections even though one is interested in what is taking place in M : in the bundle functions are unique and regular and the projection allows one to control their singularities.

Consider as base manifold M the Euclidean plane \mathbb{R}^2 with metric δ_{ab} with the associated symplectic manifold $M_S = \mathbb{R}^4$, with coordinates (q^a, p_a) . Now choose a curve C in M , parametrized by $q^a = q_0^a(s)$ in terms of the arc length s . The (unit) tangent vector, t^a and unit normal n^a are defined along C by

$$t^a \equiv \dot{q}_0^a = (t^1, t^2), \quad n^a \equiv -\varepsilon^a_b t^b = (-t^2, t^1), \quad (10)$$

with a dot denoting differentiation with respect to s . The t^a and n^a are related to each other by the (plane) Serret-Frenet equations,¹⁴

$$\dot{t}^a = kn^a \text{ and } \dot{n}^a = -kt^a \quad (11)$$

where $k(s) \equiv \delta_{ab} \dot{t}^a n^b$ and $k^{-1}(s)$ are respectively the curvature and the radius of curvature of C at s . The lines in M normal to C are called rays, and their orthogonal curves, wavefronts.

In the four dimensional space M_S of the (q^a, p_a) , we consider the two-dimensional surface, L , associated with a finite section of C where $k > 0$, by

$$q^a = q_0^a(s) + vn^a(s), \quad (12)$$

$$p_a = \delta_{ab} n^b(s) \equiv n_a(s), \quad (13)$$

$$s_1 < s < s_2, \quad 0 < v < \infty. \quad (14)$$

The v and s globally parametrize L ; different values of (v, s) give different points of L . By direct calculation, one sees that the rank of the map from (s, v) to (q^a, p_a) everywhere equals 2; this follows from

$$dq^1 \wedge dq^2 = (1 - vk) ds \wedge dv,$$

$$dq^1 \wedge dp_2 = -(t_2)^2 k ds \wedge dv,$$

$$dq^2 \wedge dp_1 = (t_1)^2 k ds \wedge dv.$$

Thus L is a submanifold of T^*M on which (s, v) are global non-canonical coordinates. One sees that if $(1 - vk) \neq 0$ then (q^1, q^2) are preferred coordinates; elsewhere one can use either (q^1, p_2) or (q^2, p_1) . Moreover one finds that on L , $\kappa = p_a dq^a = dv$ and hence $\omega = dq^a \wedge dp_a = 0$, and so L is Lagrangian.

The projection of L to M is given by

$$q^a = q_0^a(s) + vn^a(s). \quad (15)$$

The Jacobian of this mapping, $(s, p) \rightarrow q^a(s, p)$, obtained using the Serret-Frenet equations, is

$$|\mathfrak{J}| = \begin{vmatrix} (1 - vk)t^1(s), & n^1 \\ (1 - vk)t^2(s), & n^2 \end{vmatrix} = 1 - vk(s). \quad (16)$$

Thus the critical set of the projection is the curve on L given by $|\mathfrak{J}| = 0$ or

$$v = k(s)^{-1}. \quad (17)$$

The caustic is the curve in the base space M given by

$$q_c^a = q_0^a(s) + k(s)^{-1}n^a(s) \quad (18)$$

with, as mentioned earlier, $v = k(s)^{-1}$ the radius of curvature at s of C .¹⁴ After a brief calculation, one finds that the tangent vector to the caustic is

$$\dot{q}_c^a = -\dot{k}k^{-2}n^a(s);$$

hence the rays are tangent to the caustic.

Remark 1 *Perhaps a more intuitive way to characterize the caustic directly in M is to find the points where “neighboring rays intersect”:*

$$q_0^a(s) + vn^a(s) = q_0^a(s + \Delta s) + (v + \Delta v)n^a(s + \Delta s)$$

leads in the limit $\Delta s \mapsto 0$ to

$$\dot{q}_0^a + v\dot{n}^a = -\frac{\Delta v}{\Delta s}n^a = \lambda n^a,$$

and since $\dot{q}_0^a = t^a$ and $\dot{n}^a = -kt^a$ are orthogonal to n^a this gives $\lambda = 0$, and one recovers the earlier caustic condition, $v = k(s)^{-1}$. Note that defining the caustic in this manner is equivalent to the search for zero’s of Jacobi vector fields, i.e., to points conjugate to C on rays.

To further analyze this example we return to the Lagrangian submanifold L , {Eqs.(12), (13) and (14)} and (*first*) define the “lifted rays” by

$$\begin{aligned} q^a &= q_0^a(s) + vn^a(s), \\ p_a &= n_a(s), \end{aligned} \quad (19)$$

with $s = \text{constant}$ and $v = \text{variable}$ and (*second*) the “lifted wave fronts” by $v = \text{constant}$ and $s = \text{variable}$. The two vector fields spanning L , $(\partial/\partial v \equiv \widehat{T}_r, \partial/\partial s \equiv \widehat{T}_w)$, which are tangent respectively to the lifted rays and wave fronts are expressed in the coordinates (q^a, p_a) of M_S by

$$\widehat{T}_r = (n^a, 0), \quad \widehat{T}_r \cdot \widehat{T}_r = 1, \quad (20)$$

$$\begin{aligned} \widehat{T}_w &= (\dot{q}_0^a + v\dot{n}^a, \dot{n}_a) = ((1 - vk)t^a, -kt_a), \\ \widehat{T}_w \cdot \widehat{T}_w &= (k)^2 + (1 - vk)^2 > 0 \end{aligned} \quad (21)$$

with

$$\widehat{T}_w \cdot \widehat{T}_r = 0 \quad (22)$$

where the Serret-Frenet equations have again been used and the scalar product on L is given by $\widehat{U} \cdot \widehat{W} \equiv \delta_{ab}u^a w^b + \delta^{ab}\tilde{u}_a \tilde{w}_b$ with $\widehat{U} = (u^a, \tilde{u}_a)$, etc.

Evaluating $(\widehat{T}_r, \widehat{T}_w)$ on the critical curve, i.e., Eqs.(12) and (13) with $v = k(s)^{-1}$, one obtains

$$\begin{aligned} \widehat{T}_r &= (n^a, 0), \quad \widehat{T}_r \cdot \widehat{T}_r = 1, \\ \widehat{T}_w &= (0, -kt_a), \quad \widehat{T}_w \cdot \widehat{T}_w = k^2 > 0 \end{aligned} \quad (23)$$

with the tangent vector to the critical curve,

$$\widehat{T}_c = \left(\frac{-\dot{k}}{k^2} n^a, -kt_a \right), \quad \widehat{T}_c \cdot \widehat{T}_c = k^2 + \left(\frac{\dot{k}}{k^2} \right)^2 > 0. \quad (24)$$

The projections to M of these vector fields are

$$T_r^a = n^a, \quad T_w^a = (1 - vk)t^a, \quad T_c^a = \frac{-\dot{k}}{k^2} n^a \quad (25)$$

From Eqs.(20), (22), (23) and (24) we see that the lifted rays and lifted wave fronts and the critical curve have no stationary points, (i.e., no zero tangent vectors) while their projections onto M , the wave fronts, do have stationary points (“spikes” or technically cusps, [see remark below]) at the caustic, $(1 - vk = 0)$. \widehat{T}_w spans the kernel of the projection. Note also that at extremals of the curvature ($\dot{k} = 0, k \neq 0$) the caustic curve itself has stationary points - again “spikes” or cusps - provided that $\dot{k} \neq 0$ there. See Figs. 1 and 2.

Remark 2 *To see that indeed the wavefronts and the caustic curve have cusps at the stationary points of their tangent vectors T_w^a and T_c^a , we note the following: if either curve is written as $q^2 = f(q^1)$ their slopes are given, respectively, by $dq^2/dq^1 = t^2/t^1$ and $dq^2/dq^1 = n^2/n^1$ and hence are well defined at their stationary points. However, as the stationary points are smoothly traversed as functions of s , one sees (by expanding about the stationary point) that the vectors T_w^a and T_c^a point in opposite directions on either side of the stationary point if $\dot{k} \neq 0$, giving rise to the spike appearance.*

These local considerations can be applied to and globalized¹⁵ for closed convex curves, C . [See Chapter 8 of Arnold, *Catastrophe Theory*, ref. 2.]

This construction of the normals to a curve in \mathbb{R}^2 is easily extended to higher dimensions. For $M = \mathbb{R}^3$, one could construct the normals to an arbitrary 2-surface in \mathbb{R}^3 . See Sec. VII.

From a slightly different physical model as in this example, the same caustic curve (with the cusps) can easily be observed; it can be seen as the image on a two-surface, of a point source of light reflected by a distorting mirror or passing thru a distorting lens. From a different model in \mathbb{R}^3 , one could visualize the caustics as the “focusing” of light rays from a point source distorted by a mirror, passing through a smoke-filled room. These caustics would form a “two-surface”. We will return to the wave fronts and their singularities shortly via the contact bundle, where their structure is more natural.♠

◆6. Turning now to a $(2n + 1)$ -dimensional contact bundle with local coordinates (q^a, p_a, u) and contact form $\alpha = du - p_a dq^a$, we consider the analogue of a Lagrangian submanifold, namely a Legendrian submanifold, E , defined by the requirement that it be an immersed n -dimensional submanifold in the contact manifold and that the contact form vanishes when pulled back to E .

♠ *d.* A simple example of the construction of an E is to consider any function $F = F(q^a)$. With the q^a acting as the n parameters for the parametrized form of E , E is given by

$$\begin{aligned} p_a &= \partial_a F(q^a), \\ u &= F(q^a), \\ q^a &= q^a. \end{aligned} \tag{26}$$

an n -dimensional submanifold in the $(2n + 1)$ -dimensional contact space.

An alternate form for the construction of an E is to chose the p_a as parameters and take $G = G(p_a)$ as the generating function. One then has for the parametrized form of E

$$\begin{aligned} q^a &= -\partial G(p_a)/\partial p_a \equiv -\partial^a G, \\ u &= G(p_a) - p_a \partial^a G, \\ p_a &= p_a. \end{aligned} \tag{27}$$

- Note that the $u = u(p_a)$ is defined via a Legendre transformation from the $G(p)$.♠

The general forms to represent Legendrian submanifolds in terms of a generating function $G(q^A, p_J)$ are - compare with Eq.(5) ;

$$q^J = -\partial^J G, \tag{28}$$

$$p_A = \partial_A G, \tag{29}$$

$$u = G - p_J \partial^J G. \tag{30}$$

For other types of parametrization of the Lagrangian and Legendrian submanifolds, see Sec.IV.

- Note that if E is a Legendrian submanifold of $T^*M \times \mathbb{R}$, its projection into T^*M is a Lagrangian submanifold. See the previous set of equations and Eqs.(5).
- Also note that if $T^*M \times \mathbb{R}$ is considered as a bundle over $M \times \mathbb{R}$ then its fibers are Legendrian submanifolds.

◆7. The analogue of the Lagrangian mapping of an n -dimensional Lagrangian manifold to the base space M , is a “Legendrian map” of the n -dimensional Legendrian manifold (a projection) to the $(n + 1)$ -dimensional space $M \times \mathbb{R}$, i.e., to the original base M times the one dimensional fiber described by u , with coordinates (q^a, u) ; it is given locally by Eqs.(28) and (30).

After the projection one then has, in general, an n -dimensional “surface” - called a “wave front” - embedded in an $(n+1)$ -dimensional space. The singularities of this map (where the rank of the Jacobian matrix drops below n) are the wave front singularities.

♠e. A simple but very elucidating example (related to ♠c) is again to use \mathbb{R}^2 but now use r and ϕ as coordinates. We first let our configuration space M be a circle S^1 with coordinate ϕ , the 2-dimensional symplectic manifold has coordinates ϕ and p_ϕ . We then identify the contact bundle coordinate u with the radial coordinate r , i.e., we have on our 3-dimensional contact bundle the coordinates, (ϕ, p_ϕ, r) . One way to form a Legendrian submanifold is to take $F = F(\phi)$ and set

$$\begin{aligned} p_\phi &= \partial_\phi F, \\ \phi &= \phi, \\ r &= F(\phi). \end{aligned}$$

Its projection to the (ϕ, r) space is just $r = F(\phi)$, a curve in \mathbb{R}^2 representing a one-dimensional wave front in \mathbb{R}^2 . For the other forms of the generating function (or for multivalued F 's), the associated wave front would in general have singularities, i.e., singular points on the front.

We emphasize in this last example that \mathbb{R}^2 is *not the configuration space* but is the extension of the configuration space S^1 . The configuration space is S^1 , the symplectic manifold is the S^1 with its cotangent vectors and finally the contact bundle is the \mathbb{R}^2 with the cotangent vectors over the S^1 .

This construction is easily generalized to higher dimensions. Consider the configuration space M to be a closed 2-surface in R^3 with local coordinates (θ, ϕ) and

$$u = r = \sqrt{(x^2 + y^2 + z^2)}$$

so that the contact bundle has coordinates $(\theta, \phi, p_\theta, p_\phi, r)$. A generating function of the form $F = F(\theta, \phi)$ yields the Legendrian submanifold

$$\begin{aligned}
p_\theta &= \partial_\theta F, \\
p_\phi &= \partial_\phi F, \\
\phi &= \phi, \\
\theta &= \theta, \\
r &= F(\theta, \phi),
\end{aligned}
\tag{31}$$

with projection $r = F(\theta, \phi)$, a two-dimensional wave front in \mathbb{R}^3 . Again different forms of the generating function lead to singularities of the wave front (in general, curves). {There are other ways of thinking of these 2-dimensional wave fronts; in the above example we could think of a null three-surface intersecting a $t = \text{constant}$ slice of Minkowski space-time yielding the wavefront, or having the null surface intersecting a time-like tube or intersecting a null cone. The latter case is what occurs in the version of GR where the basic variables are the light-cone cuts of null-infinity.^{7,8}}

Or consider $M = \mathbb{R}^3$, with $(x, y, z, p_x, p_y, p_z, t)$ as the contact coordinates. Again with $F = F(x, y, z)$ we have the projection into the four space given by

$$t = F(x, y, z). \tag{32}$$

Arnold² calls this particular wave front, an example of a “big wave front”. In the context of Lorentzian optics (where, of course, the dynamics determines F), Eq.(32) describes a null or characteristic surface. The singularities of the “big wave front” are two-dimensional “surfaces”.

♠*f*. In the context of Legendrian submanifolds and maps we return, for a moment, to example ♠*c*. By contactifying T^*M to $T^*M \times \mathbb{R}$, i.e., by adding the coordinate $u \equiv t$, to the set (q^a, p_a) , we obtain the 5-dimensional contact manifold with coordinates, (q^a, p_a, t) . The extended base space, $M \times \mathbb{R} \equiv \mathbb{R}^2 \times \mathbb{R}$, can be interpreted as a $(2 + 1)$ -dimensional flat “space-time” with coordinates (q^a, t) . The Legendrian submanifold, E , is constructed from the Lagrangian submanifold L , by simply “adding” $t = v$ to L ; i.e., E is given by

$$q^a = q_0^a(s) + vn^a(s), \tag{33}$$

$$p_a = n_a(s),$$

$$t = v. \tag{34}$$

On E the contact form, $\alpha = dt - p_a dq^a$ vanishes since $\kappa = p_a dq^a = dv$; $\alpha = dt - dv = 0$.

The projection of E to the space-time is the “null surface”, (optical wavefront or “big wavefront”)

$$\begin{aligned} q^a &= q_0^a(s) + vn^a(s), \\ t &= v \end{aligned} \tag{35}$$

possessing a variety of singularities. [See the discussion in \blacklozenge 9.] Note that, from Eq.(18), the caustic is a non-geodesic null curve in space-time.. A similar remark applies in higher dimensions \spadesuit

\blacklozenge 8. It must be emphasized that in the case of $\spadesuit a$ and $\spadesuit e$ we have chosen (for simplicity) the generating function to depend only on the configuration space coordinates q^a . If the F , so chosen, is a (single-valued) function of the q^a there will be no caustics or wave front singularities. The caustics and wave front singularities arise from the alternate forms of the generating function or from “multivalued functions” $F(q)$, as illustrated by examples $\spadesuit b$ and $\spadesuit f$.

\blacklozenge 9. One of the most remarkable insights achieved in the theory of Lagrangian and Legendrian maps is that (in the cases of configuration spaces, with dimension $n \leq 5$) there is a simple and complete classification of the associated stable singularities of the maps. (Stable means that the singularities retain their qualitative, differential-topological properties under all small perturbations of the generating function. An example of an unstable singularity is provided by an L which is a fiber, $q^a = G^a(p_b) = const.$) This classification, which really lies at the heart of the work of Arnold and coworkers, is based on the idea of using the allowed fiber preserving canonical coordinate freedom on the symplectic or contact spaces to put the generating functions into different inequivalent canonical forms.

We conclude this section with a discussion and a list of the stable singularities for low dimensions of the base manifold M . We will include both the singularities of the Lagrangian and Legendrian maps. In particular we will discuss in detail all the cases of dimension $n = 1$ and 2 and just skim over the case of $n = 3$. The notation used to describe the different cases, i.e., A_i and D_i is that of Arnold and arises from the observation that the classification of singularities is closely related to the classification of semisimple Lie groups where that notation is used. Also for simplicity we have excluded from the summary several closely related cases that differ by signs.

The following material is complicated (though not difficult) and is not essential for the further understanding of this work. On first reading one might want to skip it and go straight to the remarks at the end of this section.

For typographic reasons - the conflict between superscript indices and powers - we will use the coordinates (x, y, z) for the base space coordinates and the contact coordinate instead of the customary (q^a, u) for the remainder of this section.

(a) 1-dimensional M with local coordinate x :

A_1 : The trivial case of a neighborhood of a non-critical point of a Lagrangian submanifold has the canonical choice of generating function $G = -p^2$, and thus $\{x = 2p, p = p\}$ represents the Lagrangian submanifold locally.

A_2 : The only other case in 1-dimension is the fold singularity of the Lagrangian map. Again let x be the coordinate of M and p the momentum coordinate. A canonical choice for the generating function is $G = -p^3$. The Lagrangian submanifold is given by $\{x = 3p^2, p = p\}$ and the projection to M is $x = 3p^2$ which has a “fold” singularity at $p = 0$, with rank $\mathfrak{J} = 0$. Near the fold, the Lagrangian submanifold covers the base twice for $x > 0$ and not at all for $x < 0$; p is a coordinate near the singularity, x is not.

Extending this to a (three dimensional) contact manifold where the contact coordinate is $u = y$ then the Legendrian submanifold is given, via $y = G + xp$, as $\{y = 2p^3, x = 3p^2, p = p\}$ with the Legendrian projection given by $\{y = 2p^3, x = 3p^2\}$ which is a curve in the (x, y) plane having a cusp at $p = 0 \Leftrightarrow (x = 0, y = 0)$. Alternately the curve can be given by $4x^3 = 27y^2$. (Note: the Legendrian projection is a homeomorphism, but not a diffeomorphism near that point.) This is the most general stable (local) form that a singular one-dimensional wave front in \mathbb{R}^2 can take. This cusp is what we referred to earlier, in $\blacklozenge 5c$, as a “spike” in the wave front. See Fig. 1.

Before proceeding we introduce a convention:

When specifying a generating function G for $n > 1$ and we write down only terms containing p_a 's; it is to be understood that besides those variables which are “visible”, G is always taken to depend trivially on as many configuration variables as are needed for a “good” coordinate system on L . If, e.g., $n = 3$ and we write $G = -p_x^3$, (as in the case $n = 1$, A_2) we mean $G(p_x, y, z) = -p_x^3$, so that the corresponding representation of L , with coor-

dinates (p_x, y, z) , reads

$$x = 3p_x^2 \quad , \quad p_y = 0 \quad , \quad p_z = 0$$

and the projection is given by

$$(p_x, y, z) \rightarrow (3p_x^2, y, z).$$

All the action is in the (x, p_x) pair while the other coordinates are dummies. This convention is obviously useful; in particular when proceeding to dimension $n + 1$, it is not necessary to list again all the cases for $m \leq n$ augmented by dummies. (Note that the amount by which the rank drops is not affected by dummies.)

(b) 2-dimensional $M = \mathbb{R}^2$ with local coordinates (x, y) .

Again we have the cases A_1 and A_2 , augmented as described above. In the A_2 Legendrian case, p_x and y can be taken as coordinates on E , which is the 2-surface in (x, y, p_x, p_y, z) -space generated by $G = -p_x^3$, $z = G + xp_x$ so that the Legendrian submanifold (with $u = z$) is given by

$$\{x = 3p_x^2, \quad y = y, \quad z = 2p_x^3, \quad p_x = p_x, \quad p_y = 0\}$$

whose image under projection to $\mathbb{R}^3 = M \times \mathbb{R} = (x, y, z)$ is the “product” of the algebraic curve $4x^3 = 27z^2$ (considered above) with the y -axis. This is a two-dimensional wave front in \mathbb{R}^3 that has a “cusp ridge” singularity along the y -axis. The fold singularity of the Lagrangian map *becomes* the cusp ridge of the Legendrian map; see Fig. 3.

A_3 : The third canonical type of Lagrangian submanifold in $n = 2$ dimensions (which has a new form) is given by $G = -(p_x)^4 + y(p_x)^2$ yielding

$$\{x = 4(p_x)^3 - 2yp_x, \quad y = y, \quad p_y = (p_x)^2, \quad p_x = p_x\} \quad (36)$$

with the Lagrangian map $\{x = 4(p_x)^3 - 2yp_x, y = y\}$. The critical points of the Lagrangian submanifold are given by the curve $y = 6(p_x)^2$ which when projected to M yields the caustic curve, $\{x = -8(p_x)^3, y = 6(p_x)^2\}$ which is a cusp in the (x, y) plane; the rank \mathfrak{J} drops by one.

[Note that in the two-dimensional plane the only two types of stable singularities are the folds of the A_2 maps and the cusps of the A_3 maps. They can be made physically manifest by the poor focusing of light from a simple source (e.g., by a point source) onto a plane by a distorting mirror or

glass of water. Their general appearance is often complicated by the fact that multiple sources often give rise to several different Lagrangian submanifolds and their respective caustics overlap. Another very important physical manifestation of these caustics is their appearance in the “source” plane in the theory of gravitational lensing. We will say more about this in a later section.]

The Legendrian submanifolds obtained from the A_3 generating function have the form

$$\{x = 4(p_x)^3 - 2yp_x, y = y, z = 3(p_x)^4 - yp_x^2, p_x = p_x, p_y = (p_x)^2\} \quad (37)$$

and the Legendrian map to $\mathbb{R}^3 : (x, y, z)$, given by

$$\{x = 4(p_x)^3 - 2yp_x, y = y, z = 3(p_x)^4 - y(p_x)^2\}. \quad (38)$$

This is a two-dimensional surface (wavefront) in \mathbb{R}^3 , parametrized by the (y, p_x) , known as the swallow tail. See Fig. 4. Its critical points are given by $y = 4(p_x)^2$ which map to the wavefront singularities on the curve $\{x = -5(p_x)^3, y = 4(p_x)^2, z = -(p_x)^4\}$.

(c) 3-dimensional $M:(x, y, z)$; there are the same cases as in (a) and (b), i.e., $A_1, A_2,$ and A_3 , plus three new cases namely

$$A_4 : G = -(p_x)^5 + z(p_x)^3 + y(p_x)^2$$

and

$$D_4^\pm : G = \mp(p_x)^2 p_y + (p_y)^3 + z(p_y)^2.$$

The A_4 caustics (two-surfaces in 3-space) are swallowtails and the D_4 caustics are the so-called elliptic umbilic and hyperbolic umbilic singularities. (See Arnold³ for their definitions.) The Legendrian singularities associated with the A_4 and D_4 maps (e.g., the singularities of the “big wave front”, the null surfaces in space-time, are far more complicated. The five singularities listed above, (i.e., A_2, A_3, A_4, D_4^\pm), applied to the spatial projections of the big wave-fronts in $(3+1)$ -space-time (i.e., the three dimensional caustic and its singularities) have been treated and shown to be stable.¹⁶

Remark 3 : *In both cases A_2 and A_3 the rank of the corresponding Jacobian \mathfrak{J} drops by one at the critical curve. The direction of the kernel of \mathfrak{J} is tangent to the critical curve only at the cusp point in the A_3 case, while it is transverse to the critical curve in the other case. Similar invariant criteria can be used to characterize the other singularities. (To see this in the context of lens*

theory, see Ref.[9].) An advantage of such criteria is that they can be applied without having to transform to the normal form of the generating function. In the examples $\spadesuit c$ and $\spadesuit f$ one easily verifies that a critical point at which $\dot{k} \neq 0$ corresponds to a fold and one at which $\dot{k} = 0$ corresponds to a cusp.

Remark 4 : We want to emphasize that all the different types of caustics in the low dimensional cases have been observed in optical experiments.¹⁷

Remark 5 : In the treatment of symplectic and contact manifolds and their associated Lagrangian and Legendrian submanifolds that we have given here, we began with a base space of dimension n , then introduced its cotangent bundle T^*M of dimension $2n$ (the phase space) and defined Lagrangian submanifolds as special n -dimensional submanifolds in T^*M . We then introduced an additional dimension \mathbb{R} , obtaining locally $T^*M \times \mathbb{R}$ (with the contact coordinate u on \mathbb{R} and contact form $\alpha = du - p_a dq^a$) and thus obtained the $(2n + 1)$ -dimensional contact manifold as an “extension” (the contactification) of the phase space with its n dimensional Legendrian submanifolds. We then considered the projections of the Lagrangian submanifolds onto the n -dimensional space M (Lagrangian maps) and the projections of the Legendrian submanifolds onto the $(n + 1)$ -dimensional space $M \times \mathbb{R}$ (Legendrian maps). We want to point out two aspects of this:

a. We could have started in an alternate way and introduced a different base space \widetilde{M} (configuration space) of dimension $(n + 1)$ and its phase space $T^*\widetilde{M}$ of dimension $(2n + 2)$; then by considering the projective cotangent space $PT^*\widetilde{M}$ (i.e., non-zero covectors of \widetilde{M} up to scale) we would have obtained a $(2n + 1)$ -dimensional contact space. The contact structure of $PT^*\widetilde{M}$ arises as follows: the symplectic potential $\widetilde{\kappa}$ of $T^*\widetilde{M}$ defines on $PT^*\widetilde{M}$ a one-form up to a non-zero factor. Thus the corresponding null vector spaces (i.e., the annihilators) of the one-forms are unique: they form the “field of contact hyperplanes” on $PT^*\widetilde{M}$. In suitable coordinates $\{q^a, q^0, p_a, (p_0 = -1)\}$, the one-form $dq^0 - p_a dq^a$ generates the field of the above contact hyperplanes. Although the intrinsic and global structures of $PT^*\widetilde{M}$ and $M_C = T^*M \times \mathbb{R}$ are different, their dimensions $1+2n$, are the same and they play the same role for local considerations. We can choose local coordinates (q^a, q^0, p_a) with $p_0 = -1$ on (part of) $PT^*\widetilde{M}$ and identify them with local coordinates $(q^a, q^0 = u, p_a)$ on M_C : Then the hypersurface elements given by $dq^0 - p_a dq^a = 0$ correspond

to those given by $du - p_a dq^a = 0$. Therefore, if the objects of interest are these elements and not the 1-forms themselves, one may locally work with either $PT^*\widetilde{M}$ or M_C and their Legendrian submanifolds and projections. This applies in particular to the local study of null hypersurfaces in space-time, the latter being represented either as \widetilde{M} or $M \times \mathbb{R}$. For their global analysis, $PT^*\widetilde{M}$ is the appropriate setting.

b. Though it is often natural to think of the configuration space M as the physical space (e.g., when one discusses caustics of families of light rays), nevertheless it is equally often useful to think of the $(n + 1)$ space $M \times \mathbb{R}$ as the physical space or space-time (e.g., when studying wavefronts and their singularities.) Sometimes the relations between these two interpretations can get quite confusing. Depending on the physical situation the relationships could be quite different.

Remark 6 : *In this section we have mainly tried to give an exposition of the mathematics of Lagrangian and Legendrian submanifolds and their maps to M and $M \times \mathbb{R}$, with occasional digressions to their connections with physics. In particular we have just explained that a variety of generating functions can be used to obtain a variety of Lagrangian and Legendrian submanifolds and their maps but we have essentially avoided describing how they are to be physically chosen. We have done this for two reasons; pedagogically we thought it best to first describe the mathematics and second because the variety of different physical uses could in their own way be confusing. We think that there is however one essential idea that is common to all (or at least most) uses; when a Lagrangian (Legendrian) submanifold is chosen it should be thought of as a particular ensemble of states of a physical system, i.e. for each point of the submanifold there is a particle (photon or light-ray in the case of geometric optics) with a particular position and momentum. Sometimes the submanifold will be thought of as representing the initial conditions for the ensemble, other times it will represent the evolution of a smaller ensemble. Later we will discuss these ideas in the context of the Hamilton-Jacobi equation with an emphasis on the eikonal equation, i.e., the massless Hamilton-Jacobi equation, and the beautiful theory of generating families.*

Remark 7 : *We point out a fascinating historic fact that seems not to be well known; Einstein¹⁸ as early as 1917 in his investigations (involving both*

improvements and serious criticisms) of the Sommerfeld-Epstein quantum rules, very clearly came across the existence of Lagrangian submanifolds; he clearly saw that generating functions of the form $G(q^a)$ could, in general, only be given locally or as multivalued functions and that in regions there could be lower dimensional subspaces of critical points.

III. An Example from Dynamics

In this section we will give a simple but very illustrative example of the physical use of the mathematical ideas described in the previous section. The example comes from a verbal discussion given by Arnold in reference [2] and worked out in reference [10].

Consider a one-dimensional configuration space (and its associated phase space) with a free particle Hamiltonian,

$$H = \frac{1}{2}p^2.$$

We want to treat the evolution of an ensemble of free particles, or equivalently a pressureless fluid, with some given initial conditions. This problem can be treated either directly via the particle motions, ($x = x_0 + pt$), or via the Hamilton-Jacobi (H-J) equation

$$\frac{\partial S}{\partial t} + \frac{1}{2}\left(\frac{\partial S}{\partial x}\right)^2 = 0. \quad (39)$$

We select the latter method since it illustrates the material of both the previous and later sections.

We choose, as an example, the following *initial* momentum distribution

$$p = \frac{1}{1+x^2}, \quad (40)$$

defining a Lagrangian submanifold in the (x, p) phase space. Since $p = \frac{\partial S}{\partial x}$, we obtain the initial value of the action function (or velocity potential)

$$S_0 = \tan^{-1}x. \quad (41)$$

Simply from the physical situation we expect the faster moving particles eventually to overtake the slower ones and that the single valued momentum field, Eq.(40), should change to a multivalued one. At points where the “multi-valuedness” starts or ends, we expect to find the focusing or caustic points of the projection map. [See fig.5]

Using, from Eq.(40), $x \equiv x_0 = \pm\sqrt{p^{-1} - 1}$ for the initial position as a function of the momenta, the equations for the particle motions, namely

$$g_{\pm} = x - pt - x_0 \equiv x - pt \mp \sqrt{p^{-1} - 1} = 0, \quad (42)$$

implicitly define a function $p(x, t)$ in the strip $0 \leq t < \frac{8}{9}\sqrt{3} \equiv t_c$, (for the meaning of t_c , see below), $-\infty < x < \infty$. (To see this, consider $g_{\pm} = 0$ for fixed (x, t) in $0 < p \leq 1$. Then $g_+ = 0$ ($g_- = 0$) has a unique solution for p if $x \geq t$ ($x \leq t$), and for $x = t$ the solutions coincide, $p(x, x) = 1$.) Using this function, $p(x, t) = \frac{\partial S(x, t)}{\partial x}$, we can write down the solution of the time dependent H-J equation,(39)

$$S(x, t) = \frac{1}{2}p^2t + \tan^{-1}(x - pt) \quad (43)$$

where $p = p(x, t)$ is that defined above in Eq.(42), in the aforementioned strip. Note that this solution can be directly obtained by integrating the H-J equation with the initial data, Eq.(41); see Secs.IV and VII.

At *all times* t , the particle states (x, p) are given by the (cubic) algebraic curve, (obtained from $g_+g_- = 0$),

$$p(x - pt)^2 + p - 1 = 0. \quad (44)$$

which represents a family $L(t)$ of Lagrangian submanifolds in the (x, p) phase space. For $0 \leq t < t_c$, $L(t)$ is generated by $S(x, t)$,

$$p = \frac{\partial S(x, t)}{\partial x}, \quad x = x \quad (45)$$

with a trivial projection (diffeomorphism). But for $t > t_c$, there is a time-dependent open interval $x_1 < x < x_2$ where Eq.(44) has three solutions, p_i , i.e., $L(t)$ has two folds; see Fig.4. Near the folds S does not generate $L(t)$. However near the folds we can introduce an alternative generating function $G(p, t)$, obtained, first, for $t < t_c$ by the Legendre transformation

$$G(p, t) = S(x, t) - px = -\frac{1}{2}p^2t + \tan^{-1}(\sqrt{p^{-1} - 1}) - p\sqrt{p^{-1} - 1} \quad (46)$$

in the domain $x > t$ and then continued to $t \geq t_c$. Then $L(t)$, including the critical points, is given by

$$x = -\frac{\partial G}{\partial p} = pt + \sqrt{p^{-1} - 1}, \quad p = p \quad (47)$$

Dropping the trivial part, $p = p$, Eq.(47) gives the projection onto the x -space.

The critical points (where the folds are) are given by those values of p where

$$\frac{dx}{dp} = 0 = t - \frac{1}{2p^2\sqrt{p^{-1} - 1}}. \quad (48)$$

Eq.(48) can be rewritten as

$$f(p, t) \equiv p^4 - p^3 + \frac{1}{(2t)^2} = 0. \quad (49)$$

Thought of as a function of p , $f(p, t)$ has a minimum at $p = 3/4$, and a point of inflection with a double extremum at $p = 0$. One sees that at the *minimum*, for $t < t_c \equiv \frac{8}{9}\sqrt{3}$, $f(3/4, t)$ is positive and hence $f(p, t)$ does not vanish for any p while for $t > \frac{8}{9}\sqrt{3}$ there are always two solutions; one solution lies between 0 and 3/4 while the other lies between 3/4 and 1. As $t \Rightarrow \infty$, the two roots approach respectively 0 and 1. The values of x , in Eq.(47), associated with these critical points are the caustics which move to the right along the x -axis with increasing t .

As we mentioned earlier, physically we can think of this example as representing an ensemble of free particles moving to the right with an initial momentum distribution. After some time the faster ones catch up with the slower ones and the distribution becomes triple valued with two caustics. If there had been some initial smooth density $\rho = \rho_0(x_0)$, the density at later times is given by $\rho(x, t) = \rho_0(x_0)(dx_0/dx)$, where the initial position x_0 as a function of x and t is obtained by inserting the p from Eq.(40) as a function of x_0 , i.e., $p = 1/(1 + x_0^2)$, into $g_+ = 0$ of Eq.(42), yielding $x = X(x_0, t) = x_0 + t/(1 + x_0^2)$. After the critical time $t_c = \frac{8}{9}\sqrt{3}$ there will be caustics at points $x_1(t)$ and $x_2(t)$; for each x between these two points there will be three values of p on the Lagrangian submanifold which correspond to three initial positions, x_{i0} [$i = 1, 2, 3$]. Associated with these three different p 's there will be three density distributions $\rho_i(x, t)$, which turn out to be

$$\rho_i(x, t) = \frac{(1 + x_{i0})^2 \rho_0(x_{i0})}{(1 + x_{i0})^2 - 2tx_{i0}}.$$

At t_c the “flow” splits at the first caustic point x_c into three partial flows, and thereafter there are two moving infinite “density waves” at the caustic positions.

The singularities at $x_1(t)$ and $x_2(t)$ are folds, while the singularity at the “trifurcation” point x_c , (the point at the critical time where the caustic first begins) is an unstable one if considered as belonging to the Lagrangian projection at fixed t , while it is stable as a singularity of the family of maps with variable t , called a metamorphosis (perestroikas).

We mention in passing that Eq.(44) can alternatively be interpreted as defining a Lagrangian submanifold in the $(x, t; p, -E)$ phase space over the (x, t) space-time as base. In that interpretation the fold curves $x_i(t)$ meet at the (stable) cusp point (x_c, t_c) where the caustic begins.

The ideas described here can (in principle) be extended to H-J theory with arbitrary Hamiltonians. From a complete solution of the H-J equation on an n -dimensional configuration space (i.e., one that depends on n independent constants) it is possible to construct a solution $S(x^a, t)$ from arbitrary Cauchy data, $S_0(x^a)$ and study the evolving Lagrangian submanifold with the development of the critical points and density waves. In fact the use of this idea has been proposed and extensively developed in order to account for the origin of large scale structure in the early universe.^{11,19,20} In Secs.VI and VII in the discussion of generating families, we will return to this issue.

IV. Multiple Uses for the Eikonal

Another interesting and useful application of the ideas of Sec.II is to wave propagation in arbitrary space-times. The treatment is essentially kinematic, the dynamics enters in the fact that we are assuming that we can solve the eikonal equation and that we can produce families of solutions at will. For simplicity of presentation, we will take a rather unsophisticated approach to the solution of algebraic equations, assuming that almost always a solution exists. Later, in Section VI, we will give a more sophisticated treatment of the same issues

We begin with an arbitrary four dimensional manifold, M^4 , with a Lorentzian metric g given in some local coordinate system x^a , by $g_{ab}(x)$. We want to consider the null hypersurfaces of g in M^4 ; e.g., the hypersurfaces of constant phase in the geometric optics (high frequency) limit of the Maxwell equations on M^4 . These hypersurfaces can be described as the level

surfaces of functions $S = S(x^a)$ satisfying the eikonal equation, (or massless H-J equation) namely

$$g^{ab}\partial_a S\partial_b S = 0. \quad (50)$$

A solution $S = S(x^a)$ will be referred to as an eikonal. Though later we will discuss (in a special case) the problem of generating solutions to Eq.(50), at the start we will assume that we have been given a solution $S(x^a)$ that is continuous but perhaps only piece-wise smooth. The level “surfaces” of S might have self-intersections and have sharp edges, as for example in Figs.3 and 4.

We will consider several different uses for the $S(x^a)$.

1. First we write $S(x^a) = S(x^A, r, t)$, $A = 1, 2$, having made an arbitrary decomposition of the four space into a one parameter family of three dimensional spaces ($t = \text{constant}$); these three surfaces are in turn foliated by families of two dimensional surfaces, M^2 , with x^A as coordinates. These 2-surfaces are parametrized by r on each $t = \text{constant}$ three-surface. (One might think of $t = \text{constant}$, as space-like surfaces, with r as a radial coordinate and the x^A as the local angular coordinates on the two-surface, though there are many alternate pictures one could make.) We now consider the cotangent bundle over each M^2 , with $\omega = dx^A \wedge dp_A$ and construct a Lagrangian submanifold, L^2 , on it in the following manner. From $S = S(x^A, r, t)$, fix $S = s_0$ and $t = t_0$ and solve for r , obtaining

$$r = R(x^A, t_0, s_0). \quad (51)$$

Using this as a generating function for L^2 , we have that

$$\begin{aligned} p_A &= \partial_A R(x^A, t_0, s_0), \\ x^A &= x^A \end{aligned}$$

defines a Lagrangian submanifold.

2. Any one of these symplectic manifolds, M_S^2 , can now be contactified by adding the coordinate r . The contact form is then $\alpha = dr - p_A dx^A$.

A Legendrian submanifold E^2 in $M_S^2 \times \mathbb{R}$ is defined by

$$p_A = \partial_A R(x^A, t_0, s_0), \quad (52)$$

$$\begin{aligned}x^A &= x^A, \\r &= R(x^A, t_0, s_0),\end{aligned}$$

with the projection to the (x^A, r) space (i.e., the three dimensional space $t = t_0$) given by $r = R(x^A, t_0, s_0)$. This describes a wave front constructed from the intersection of the null “surface”, $s_0 = S(x^a)$ with the $t = t_0$ three-surface. Note that by the assumptions in this construction the wavefronts will not have any singularities; however if evolved to later times they in general do develop singularities.

3. An alternate way of looking at this evolution is to go back to $S(x^a)$ and view it as $S(x^i, t)$, with $x^i = (x^A, r)$, and again take $S = s_0$ and then solve for $t = T(x^i, s_0)$. Now we treat a manifold (a time slice), M^3 , with coordinates x^i , as the base space of a symplectic bundle, (x^i, p_i) with form

$$\omega = dx^i \wedge dp_i.$$

A Lagrangian submanifold can be obtained from the generating function, $T(x^i, s_0)$, with projection to M^3 . Again the construction used here precludes caustics; however generalizations to be considered later do lead to caustics which consists of the singularities of the evolving wave fronts.

Remark 8 : *Note that there is a completely different Lagrangian submanifold in the same, (x^i, p_i) , symplectic space also constructed from $S(x^a)$; it arises from allowing the value of S to vary but keeping t_0 constant. A generating function is $S = S(x^i, t_0)$. Its projection (and the associated caustics) to M^3 are completely different from those of the generating function $t = T(x^i, s_0)$. They are the caustics associated with families of null surfaces studied at one instant of time; contrasted against the previous case of the projection of one null surface $S = S_0$ in space-time into the three space of the x^i . This distinction often is the source of considerable confusion.*

4. This Lagrangian structure, obtained from $T(x^i, s_0)$, can be contactified by adding to M^3 the coordinate t , with contact form $\alpha = dt - p_i dx^i$. Now taking the generating function as $t = T(x^i, s_0)$, we obtain a Legendrian submanifold of (x^i, p_i, t) – space. Its projection to space-time, $M^4 = (x^a) = (x^i, t)$ is a null “three-surface”, referred to as a “big wave front” by Arnold. This is the same “surface” as described by $s_0 = S(x^a)$; a level surface of the eikonal.

5. As the last use of the eikonal, we mention that taking $S = S(x^a)$ as the generating function for a Lagrangian submanifold in the symplectic manifold over M^4 given by $p_a = \partial_a S(x^a)$ there will, in general, be three dimensional caustics associated with its projection to M^4 . We do not know of a geometric use for this construction.

♠ Often in a physical discussion one is interested in a steady state situation where a light source (say a point or a two-surface) would light up and remain on as a source (in time) of families of wave fronts. {We assume in this discussion that the metric (or a conformally related metric) in the eikonal equation does not depend on time.} The families of wave fronts would look exactly alike at every instant of time t . The problem is to solve the eikonal equation so that the initial or Cauchy data, $S(x^i, t_0) = S_0(x^i)$, corresponds to the evolution of one wave front obtained from its normal evolution from the given source surface, that is projected back to the $t = t_0$ three-surface. This can be accomplished by returning to item #3, where the evolving wave fronts on the three manifold M^3 of the (x^i) , was described by the level 2-surfaces of $t = T(x^i, s_0)$. Treating the two-surface defined by $t_0 = T(x^i, s_0)$ as the source surface and ignoring s_0 since it is a given constant, we can define the Cauchy data by $S_0 = T(x^i)$; the level surfaces of this function S_0 are the wave fronts at $t = t_0$ of light emitted by the source at earlier times. This is the situation that arises in the discussion of gravitational lensing; it is assumed that there is a fixed source in a (conformally stationary) space-time that continuously emits light.

A closely associated point of view to this is to consider the “*time-independent*” eikonal equation (defined only in conformally stationary space-times), namely

$$g^{ij}\partial_i\widehat{T}\partial_j\widehat{T} - 2g^{i0}\partial_i\widehat{T} + g^{00} = 0. \quad (53)$$

This equation can be obtained by substituting the ansatz, $S = t - \widehat{T}(x^i)$ into the original eikonal equation. This equation is satisfied by the Cauchy data $S_0 = T(x^i)$ of the previous paragraph if we take T , defined implicitly by

$$s_0 = S(x^i, T) \quad (54)$$

where $S(x^i, t)$ satisfies the eikonal equation, $g^{ab}\partial_a S\partial_b S = 0$. Indeed, differentiating Eq.(54) with respect to x^i , we obtain $0 = \partial_i S + (\partial S/\partial t)\partial_i T$ or

$$\partial_i T = -\frac{\partial_i S}{\partial_t S}$$

which when substituted into Eq.(53) leads back to the eikonal equation. Thus our “stationary” Cauchy data is a *solution* of the *time-independent* eikonal equation.

In the case of static space-times where the $g^{i0} = 0$, Eq.(53) is often written in terms of the Hamiltonian $H = g^{ij}p_i p_j - n^2(x^i) = 0$ with $n^2 = -g^{00}$, reinterpreted as an effective index of refraction and $p_i = \partial_i T$. This point of view leads to Fermat’s Principle of least time.

◆ We give a powerful example of how the eikonal equation can be solved with given Cauchy data. We will assume that a three parameter family of solutions of the eikonal equation is known. In principle there always exists such a three parameter family of solutions, though in practice it is generally very hard to find them exactly. Perturbation techniques might be needed to approximate them. Nevertheless for the general discussion we will assume that there exists a solution S^* of the eikonal equation that depends on three independent parameters, i.e.,

$$S^* = S^*(x^i, t, \alpha_i). \quad (55)$$

This is called a complete integral. We show that a “general integral” (which involves an arbitrary function of three variables) can be constructed from the complete integral in the following fashion: we first add to it an arbitrary function of the α_i , i.e., we consider

$$S^{**} = S^*(x^i, t, \alpha_i) - F(\alpha_i) \quad (56)$$

which trivially still satisfies the eikonal equation.

We next form the equations

$$\partial S^{**} / \partial \alpha_i = \partial S^* / \partial \alpha_i - \partial F / \partial \alpha_i = 0. \quad (57)$$

For the present we assume that it has a solution of the form $\alpha_i = A_i(x^i, t)$. (This is an example of our unsophisticated treatment. Tacitly we are referring to the implicit function theorem, assuming that the determinant of $\partial^2 S^{**} / \partial \alpha_i \partial \alpha_j$ is different from zero. We will return later, in Secs.VI and VII, to the issue of solving Eq.(57) when the determinant vanishes.)

Finally, via $\alpha_i = A_i(x^i, t)$, the α_i are eliminated in S^{**} yielding

$$S(x^i, t) = S^*(x^i, t, A_i(x^i, t)) - F(A_i(x^i, t)). \quad (58)$$

It is not difficult to show that this S satisfies the eikonal equation.²¹ The x^a derivatives of $S(x^i, t)$ involve both the explicit x^a dependence and the dependence via the $A_i(x^i, t)$; the latter dependence however drops out because of Eq.(57). Since the eikonal equation is satisfied as far as the explicit dependence then so does Eq.(58). This solution now depends on an arbitrary function of three variables, namely $F(\alpha_i)$.

The task is now to determine $F(\alpha_i)$ in terms of Cauchy data, $S_0(x^i)$. This is accomplished as follows; consider the function $\mathfrak{s}(x^i, \alpha_i) \equiv S^*(x^i, t_0, \alpha_i) - S_0(x^i)$ and then construct the three equations

$$\frac{\partial \mathfrak{s}(x^j, \alpha_j)}{\partial x^i} \equiv \frac{\partial (S^* - S_0)}{\partial x^i} = 0.$$

Because S^* is a complete solution, they can be solved for

$$\alpha_i = A_i(x^j).$$

We now assume that the Cauchy data was chosen so that the last equation can be algebraically inverted, i.e.,

$$x^i = X^i(\alpha_i).$$

At $t = t_0$ we have that

$$S_0(x^i) = S^*(x^i, t_0, A_i(x^i)) - F(A_i(x^i)). \quad (59)$$

Replacing all the A_i by α_i and all the x^i by $X^i(\alpha_j)$, we have that

$$F(\alpha_i) = S^*(X^i(\alpha_i, t_0), \alpha_i) - S_0(X^i(\alpha_i)), \quad (60)$$

i.e., the free $F(\alpha_i)$ is now expressed in terms of the free Cauchy data, $S_0(x^i)$ and the complete solution.

{The construction described here for the solution of the Eikonal equation in terms of Cauchy data can be easily extended to the H-J equation. See Sec.VII.}

♠ In Minkowski space, the plane waves provide a complete integral;

$$S^* = x^i \alpha_i - t \sqrt{(\alpha_i)^2}.$$

This allows us to find (in principle - modulo algebraic inversions) all solutions of the flat-space Eikonal equation with arbitrary Cauchy data. ♠

V. A Caveat

As we pointed out in examples, the principal strength and importance of the theory of Lagrangian and Legendrian projections lies in its ability to treat places where the projections are not diffeomorphisms.

Though in many cases it is possible and perhaps even intuitively useful to treat the projections as being almost always diffeomorphisms with lower dimensional regions as the exception. One could try to handle the singularities formally by allowing the generating functions to be multivalued and only piece-wise smooth and then approach the critical points as limits of regular ones. However this creates difficulties: the projections to the base space near critical points are difficult to treat and the structure of the caustics or wave front singularities are often hard to “see”. Even when possible, this approach certainly is inelegant and ill-defined mathematically. One might have hoped that the cases with critical points would be exceptions but the opposite is true; the existence of critical points is generic and one must be able to construct the proper type of generating function. In Sec. IV where we dealt with solutions to the eikonal equation we always tacitly assumed that the relevant equations could be solved for certain specific variables. This was true for most regions but not in the regions where certain Jacobians vanished and where the critical points existed. In fact it is the existence of the critical points that was the obstruction to solving the algebraic equations.

An important question then is; does there exist some general procedure applicable to physical problems for the construction of Lagrangian or Legendrian submanifolds with the associated projection maps - including singularities - so that the issue of finding appropriate choices for the generating function does not arise and where the associated projection maps are given in some natural systematic manner. We saw, in examples $\spadesuit c$ and $\spadesuit f$ how to obtain the Lagrangian and Legendrian submanifolds (with critical points) without a generating function and then in Sec.III, discussing the free particle H-J equation, the evolution of the H-J equation itself suggested the Legendre transformation to a proper single-valued generating function. We will see in the next section that there indeed is a systematic method based on the concept of “generating families”. We will see that it allows us to construct general Lagrangian and Legendrian submanifolds associated with either H-J or eikonal evolution based on arbitrary Cauchy data. However if wanted or needed, it is possible to construct generating functions, which are

local objects, from generating families which serve to define the submanifolds globally.

VI. Generating Families

There is a remarkably beautiful method^{1,4} for the construction of single valued (local) generating functions - easily applied in many physical situations - using what are called “*generating families*”. Actually one can bypass the generating function construction and go directly, via the generating *families*, to the Lagrangian and Legendrian submanifolds and associated maps. [Though in the literature of catastrophe theory what we are calling generating *families* are frequently called generating functions, we will stay with the terminology adopted by Arnold.]

We first outline the mathematical ideas behind “generating families” and then show how it can be applied to various physical situations or problems.

We give two methods for the construction of generating families; the first begins with a generating *function*, a special class of generating *families* being constructed from it. Second, from certain observations concerning the first construction, the procedure can be generalized to what becomes the full theory of generating families.

For the first method we start with the cotangent bundle T^*M equipped with canonical coordinates (q^a, p_a) . For step one, we assume a Lagrangian submanifold L to be given, with a point ξ on L and a generating function $G(q^A, p_J)$ near ξ . The local embedding of L into T^*M is then given by

$$q^J = -\partial^J G, \quad p_A = \partial_A G. \quad (61)$$

We now define a function F in a neighborhood U of ξ in T^*M by

$$F(q^a, p_a) = G(q^A, p_J) + q^J p_J. \quad (62)$$

Since F does not depend on p_A nor G on q^J , it is trivially seen that F identically satisfies the equations

$$\partial^A F = 0, \quad \partial_J F = p_J. \quad (63)$$

The remaining equations

$$\partial^J F = 0, \quad \partial_A F = p_A, \quad (64)$$

which are equivalent to Eqs.(61), define the Lagrangian submanifold.

The construction of F from G implies that

$$\text{rank}(\partial^a F = 0, \partial_a F - p_a) = n, \quad (65)$$

since the $2n$ equations in Eq.(65) have, by construction, the unique solution (61).

We henceforth want to forget the G from which F was constructed and claim: If a function $F(q^a, p_a)$ satisfies the rank condition Eq.(65), then the $2n$ equations

$$\partial^a F = 0, \quad \partial_a F = p_a \quad (66)$$

can be solved for some n of the set (q^a, p_a) . This uniquely yields an embedded Lagrangian submanifold. Indeed, according to the implicit function theorem, one can then express n of the variables (q^a, p_a) - say (q^J, p_A) - in terms of the remaining ones; in other words Eqs.(66) implies that

$$q^J = Q^J(q^A, p_J), \quad p_A = P_A(q^A, p_J)$$

hold on the n -manifold L with (local) coordinates (q^A, p_J) . If the identity

$$dF = \partial_a F dq^a + \partial^a F dp_a$$

on T^*M is pulled back to L , then because of Eqs.(66), one has the result that: on L

$$dF = p_a dq^a = \partial_a F dq^a = \kappa, \quad (67)$$

i.e., the restriction of F to L is a potential for the one-form κ on L and hence $\omega = 0$. Thus L is, in fact, Lagrangian. Moreover, on L , we see that $F - q^J p_J \equiv G$ obeys Eqs.(61).

This formulation in terms of F , - i.e., F being any function obeying the rank condition, Eq.(65), - allows the construction of Lagrangian submanifolds which may have regular points as well as critical points and which might require, for their local descriptions, several different generating functions. In this sense, the description in terms of F via Eqs.(65) and (66) is more general

and “less local” than the one in terms of G and Eq.(61). Any Lagrangian submanifold can locally be obtained from some F .

The foregoing argument generalizes immediately to the Legendre case. We simply add to F , (which is a function on T^*M) the additional variable $u \in \mathbb{R}$, and form $u = F$. Then a Legendrian submanifold, E , on $T^*M \times \mathbb{R}$ is given by

$$u = F, \quad p_a = \partial_a F, \quad \partial^a F = 0. \quad (68)$$

The complete theory of generating families now arises as a generalization of the preceding construction. At first sight there appear to be substantial differences but on closer observation we see that it really is a generalization. We will show later how the previous case is a specialization of the general theory.

The basic idea is to start with a configuration space, M^n , of dimension n and then enlarge it to a space $M^{n+m} = M^n \times M^m$ of dimension $n + m$, with local coordinates (q^a, s^J) , $a = 1, \dots, n$ and $J = 1, \dots, m$. The dimensions n and m are arbitrary. A “generating function” $\mathfrak{F}(q^a, s^J)$ defined on the large space, M^{n+m} , is then chosen, e.g., by physical or geometric arguments (examples of which will be given shortly).

$\mathfrak{F}(q^a, s^J)$ is arbitrary except for the following rank condition: The equations

$$\frac{\partial \mathfrak{F}}{\partial s^J} \equiv \partial_J \mathfrak{F} = 0 \quad (69)$$

admit solutions for some m of the set (q^a, s^J) , and whenever they hold, the $(n + m) \times m$ matrix

$$[\mathfrak{F}_{Ja} \equiv \frac{\partial^2 \mathfrak{F}}{\partial s^J \partial q^a}, \quad \mathfrak{F}_{JK} \equiv \frac{\partial^2 \mathfrak{F}}{\partial s^J \partial s^K}] \quad (70)$$

has rank m .

From $\mathfrak{F}(x^a, s^J)$, by an ingenious method, one can then either construct appropriate generating *functions* on the cotangent bundle over M^n and hence a Lagrangian submanifold or instead, directly construct the Lagrangian submanifold from the generating *family*.

Since the considerations are essentially local we can consider $M^{n+m} = R^n \times R^m$.

We first state the main result; namely how to construct an n -dimensional (Lagrangian) submanifold from the generating family $\mathfrak{F}(x^a, s^J)$. This is then followed by the proof that the submanifold so constructed is in fact Lagrangian.

We first use the function $\mathfrak{F}(x^a, s^J)$ as a generating function to generate a Lagrangian section \widehat{L} in the cotangent space over M^{n+m} ,

$$\begin{aligned} p_a &= \partial\mathfrak{F}/\partial q^a, & q^a &= q^a \\ \Pi_J &= \partial\mathfrak{F}/\partial s^J, & s^J &= s^J. \end{aligned} \quad (71)$$

We then define a subset of M^{n+m} by imposing the extremal condition

$$\Pi_J = \partial\mathfrak{F}/\partial s^J(q^a, s^J) = 0. \quad (72)$$

According to the rank condition, the solutions of this equation form an n -dimensional submanifold of M^{n+m} , that can be expressed by

$$q^a = Q^a(y^b), \quad s^J = S^J(y^b) \quad (73)$$

When these are substituted into $p_a = \partial_a\mathfrak{F}$, one obtains $p_a = P_a(y^b)$: the equations

$$q^a = Q^a(y^b), \quad s^J = S^J(y^b), \quad p_a = P_a(y^b), \quad \Pi_J = 0 \quad (74)$$

define an n -dimensional submanifold N of the large phase space. By its construction, N is the intersection of \widehat{L} and the submanifold P of T^*M^{n+m} defined by $\Pi_J = 0$.

The submanifold of T^*M^n defined by

$$q^a = Q^a(y^b), \quad p_a = P_a(y^b) \quad (75)$$

is Lagrangian.

What follows is the proof of this contention. As the proof is rather technical and difficult the reader might simply prefer to accept the contention and bypass the proof. Doing so does not greatly affect the understanding of, or the ability to use, generating families. The proof is given for completeness.

Proof: Let ξ be a point of N . The dimensions of the tangent spaces $N_\xi, \widehat{L}_\xi, P_\xi$, are $n, n + m, 2n + m$, respectively. Since $(n + m)$ independent

vectors at \widehat{L}_ξ , (obtained from the derivatives of Eq.(71)) using $\Pi_J = \partial_J \mathfrak{F}$ and $p_a = \partial_a \mathfrak{F}$) have the form

$$V_{(a)} = (\partial_{ab} \mathfrak{F}) \partial^b + \mathfrak{F}_{aK} \partial^K + \partial_a, \quad V_{(J)} = \mathfrak{F}_{bJ} \partial^b + \mathfrak{F}_{JK} \partial^K + \partial_J$$

and vectors at P_ξ have the form

$$Y = Y^a \partial_a + Y^J \partial_J + Y_a \partial^a,$$

(with Y^a, Y^J, Y_a arbitrary and $Y_J = 0$) we see immediately that $\widehat{L}_\xi + P_\xi$ spans the tangent space of T^*M^{n+m} at ξ and hence $\dim(\widehat{L}_\xi + P_\xi) = 2n + 2m$; i.e., \widehat{L} and P intersect transversely. (This statement is the geometric reformulation of the rank condition.)

The critical point to be established next is that the projection of N into the small phase space T^*M^n , a projection along the s^J -direction is everywhere a local diffeomorphism, so that the image L is an n -dimensional submanifold of T^*M^n , given by, Eq.(75),

$$q^a = Q^a(y^b), \quad p_a = P_a(y^b).$$

To prove that, one has to show that no (non-vanishing) vector tangent to N is annihilated by the projection. Following Arnold, this can be done elegantly as follows.

We first note that the kernel of the projection consists of all vectors of the form $X = X^J \partial_J$ (i.e., vectors in the s^J -directions) and then observe that, (from the skew-orthogonal product of tangent vector of T^*M^{n+m} , defined by $[X, Y] \equiv X^J Y_J - X_J Y^J + X^a Y_a - X_a Y^a$), the kernel is skew-orthogonal to all the vectors Y tangent to P , i.e., from $Y = Y^a \partial_a + Y^J \partial_J + Y_a \partial^a$, $[X, Y] = 0$. Suppose now that X is in the kernel and tangent to N . Then X is also tangent to \widehat{L} since $N \subset \widehat{L}$. Therefore X is skew-orthogonal to both P and \widehat{L} (since \widehat{L} is Lagrangian all vectors in \widehat{L}_ξ are skew-orthogonal). But since the tangents to P and \widehat{L} together span the total tangent space of T^*M^{n+m} - i.e., transversality - X is skew-orthogonal to “everything”, and thus $X = 0$, which was to be shown.

The submanifold L given by Eqs.(75) is, in fact, Lagrangian. This again follows by pulling back to L the identity

$$d\mathfrak{F} = \partial_a \mathfrak{F} dq^a + \partial_J \mathfrak{F} ds^J$$

which results in

$$d\mathfrak{F} = p_a dq^a = \kappa. \quad \text{QED} \quad (76)$$

Note that any Lagrangian submanifold of T^*M^n can be obtained by the foregoing construction. Suppose that L is given locally by $K(q^A, p_J)$ as in Eq.(8). Then the generating family (of the type considered in Eq.(62)).

$$\mathfrak{F}(q^a, p_J) = K(q^A, p_J) + q^J p_J,$$

(with $p_J = s^J$), reproduces L , as is easily verified.

The projection of L to the base is, of course given by

$$q^a = Q^a(y^b).$$

Taking into account how $Q^a(y^b)$ was obtained via Eq.(72), one can see that the kernel of that projection is determined by the solution X^K of

$$\mathfrak{F}_{JK} X^K = 0,$$

thus the critical points of L are given by

$$D \equiv \left| \frac{\partial^2 \mathfrak{F}}{\partial s^J \partial s^K} \right| = 0. \quad (77)$$

We may summarize and geometrically interpret the preceding construction as follows: For each fixed s^J , $p_a = \partial_a \mathfrak{F}(q^a, s^J)$ defines a singularity-free Lagrangian submanifold of T^*M^n , i.e., \mathfrak{F} , acting as a generating function, defines an m -parameter family of “regular” Lagrangian submanifolds. By solving $\partial_J \mathfrak{F} = 0$, i.e., Eq.(72), and inserting them into $p_a = \partial_a \mathfrak{F}(q^a, s^J)$, we obtain $p_a = P_a(y^b)$, which with $q^a = Q^a(y^b)$, provides another Lagrangian submanifold, the *envelope* of the former family. This Lagrangian submanifold has the projection map $\pi : y^b \rightarrow q^a = Q^a(y^b)$. Its critical points are given as those points where the rank of the Jacobian matrix $\partial_a Q^b$ is not maximal or equivalently, where Eq.(77) holds.

Now we can also see that the previous construction via Eq.(66) is included in the general case. If $m = n$ and if the first matrix in Eq.(70), i.e., \mathfrak{F}_{Ka} , has rank m , then one can express the s^J as functions of the p_a , and $F(q^a, p_a) \equiv \mathfrak{F}(q^a, s^J(p_a))$ is a generating family of the former kind.

Eq.(75) represents the Lagrangian submanifold in terms of some coordinates y^b . Due to the implicit function theorem, the y^b can always (locally) be chosen as subsets of the (q^a, s^J) .

We now consider the possible cases:

#1. Let us first assume that at a solution point, (q_0^a, s_0^J) ,

$$D \equiv |\mathfrak{F}_{JK}| \neq 0. \quad (78)$$

Then, Eqs.(72) can be solved uniquely for all the s^J ,

$$s^J = S^J(q^a). \quad (79)$$

This result can be inserted into $\mathfrak{F}(q^a, s^J)$ so that

$$\mathfrak{F}(q^a, s^J) \Rightarrow G(q^a) = \mathfrak{F}(q^a, S^J(q^a)) \quad (80)$$

yields a generating function $G(q^a)$ for a Lagrangian submanifold. From the general theory

$$p_a = \partial G / \partial q^a, \quad q^a = q^a \quad (81)$$

with a trivial (diffeomorphism) Lagrangian map.

Conversely when $D = 0$, at (q_0^a, s_0^J) , the Lagrangian projection is not a diffeomorphism in any neighborhood of the point, i.e., we have a Lagrangian singularity there as noted in connection with Eq.(77). The vanishing of D is thus the necessary and sufficient condition for the occurrence of a caustic at the point in question.

#2. The other case to consider is when the m equations $\partial_J \mathfrak{F} = 0$ can be algebraically solved for a mixture of some $q^{a'}$'s and some $s^{J'}$'s, i.e., where the solutions have the form

$$q^{a'} = Q^{a'}(q^{a''}, s^{K''}), \quad s^{J'} = S^{J'}(q^{a''}, s^{K''}), \quad (82)$$

with m variables $(q^{a'}, s^{J'})$ and n variables $(q^{a''}, s^{K''})$ such that at least one $s^{K''}$ occurs. (The set of $q^{a''}$ might be empty.) The Lagrangian submanifold, parametrized by the n variables $(q^{a''}, s^{K''})$, is now given by

$$\begin{aligned} q^{a'} &= Q^{a'}(q^{a''}, s^{K''}) \\ q^{a''} &= q^{a''}, \\ p_a &= \partial_a \mathfrak{F} = P_a(q^{a''}, s^{K''}). \end{aligned} \quad (83)$$

The generating function

$$S(q^{a''}, p_{a'}) = \mathfrak{F}(Q^{a'}(q^{a''}, s^{K''}), q^{a''}, S^{J'}(q^{a''}, s^{K''}), s^{K''}) - p_{a'} Q^{a'}(q^{a''}, s^{K''}), \quad (84)$$

which does not depend on $s^{K''}$, yields the same submanifold as do Eqs.(83). To see that S , in fact, does not depend on $s^{K''}$, one first treats the right side as a function of $(q^{a''}, p_{a'}, s^{K''})$ and then by differentiating with respect to $s^{K''}$, and using Eqs.(83), one sees that the derivative vanishes.

Since, from generating functions for Lagrangian submanifolds one can construct a contact coordinate (see Eqs.(26) and (27)) and hence a Legendrian submanifold and Legendrian map, the construction of the Lagrangian submanifolds via generating *families* rather than generating *functions*, is easily extended (see Eq.(68)) to the Legendrian submanifolds and maps via

$$u = \mathfrak{F}(q^a, s^J), \quad \partial_J \mathfrak{F} = 0, \quad p_a = \partial_a \mathfrak{F}.$$

VII. Applications of Generating Families

Since many or perhaps most of the applications in physics of Lagrangian and Legendrian submanifolds and maps are associated with dynamical or optical systems and appear to be either directly or indirectly associated with Hamilton-Jacobi theory or the related eikonal equation we will confine our discussion to showing how generating families can be constructed for specific physical situations.

♠ We begin with a simple but important physical model. Consider four dimensional Minkowski space-time foliated by the standard $t = \text{const.}$, space-like three surfaces $\sum_t \Leftrightarrow \mathbb{R}^3$, with Cartesian coordinates x^i . We choose at $t = 0$, an arbitrary two surface, \mathfrak{S} , in \sum_0 that “lights-up”, i.e., that is to be a source of light, with local coordinates (s^J) , $J = 1, 2$, i.e., $x^i = x_0^i(s^J)$, which describes \mathfrak{S} parametrically. The (x^1, x^2, x^3) in \mathbb{R}^3 are the points of physical space (observation points) that will be reached by light rays from \mathfrak{S} . At each instant of time t , the light fills a region bounded by two “small wavefronts” - from the “incoming” and “outgoing” rays. In space-time these small wavefronts, as time evolves, form a pair of null hypersurfaces (“big” wavefronts), whose intersection is \mathfrak{S} . We wish to find these small wavefronts.

Let the function $t = \mathfrak{F}(x^1, x^2, x^3, s^1, s^2)$ represent the time it takes for light to go from \mathfrak{S} to the observation point, x^i . From the constancy of the speed of light, $c = 1$, we have that

$$t = \mathfrak{F}(x^1, x^2, x^3, s^1, s^2) = \sqrt{(x^i - x_0^i(s^J))(x^i - x_0^i(s^J))} \quad (85)$$

which we will write as

$$\mathfrak{F} = \sqrt{(\mathbf{r} - \mathbf{r}_0(s^J)) \cdot (\mathbf{r} - \mathbf{r}_0(s^J))}.$$

First of all we define, in accordance with Eq.(71)

$$\mathbf{p} = \frac{\partial \mathfrak{F}}{\partial \mathbf{r}} = \frac{(\mathbf{r} - \mathbf{r}_0(s^J))}{|\mathbf{r} - \mathbf{r}_0(s^J)|}. \quad (86)$$

From $\Pi_J = \partial \mathfrak{F} / \partial s^J = 0$, we have that

$$\Pi_J = -\frac{(\mathbf{r} - \mathbf{r}_0(s^J))}{|\mathbf{r} - \mathbf{r}_0(s^J)|} \cdot \mathbf{T}_J = -\mathbf{p} \cdot \mathbf{T}_J = 0 \quad (87)$$

with $\mathbf{T}_J(s^K) = \partial \mathbf{r}_0 / \partial s^J$, the two tangent vectors to the surface. (The physical meaning of $\partial \mathfrak{F} / \partial s^J = 0$ is that, since $t = \mathfrak{F}(\mathbf{r}, \mathbf{r}_0(s^J))$, the travel time of a ray leaving from the surface at $\mathbf{r}_0(s^J)$ and arriving at \mathbf{r} is an extremal (minimum). We see, below, that to satisfy this condition, rays must leave normal to the surface, \mathfrak{S} .

We can solve the Eqs.(87) by introducing the unit normal to \mathfrak{S} , given by

$$\mathbf{n}(s^J) = \frac{\mathbf{T}_1 \times \mathbf{T}_2}{|\mathbf{T}_1| |\mathbf{T}_2|} \quad (88)$$

and using the fact that Eq.(87) implies that

$$\mathbf{r} = \mathbf{r}_0(s^J) + v \mathbf{n}(s^J). \quad (89)$$

Thus from Eq.(86),

$$\mathbf{p} = \mathbf{n}(s^J) \quad (90)$$

i.e., if Eq.(87) holds, the momentum is the unit normal to the surface \mathfrak{S} .

Eq.(89), for each fixed value of v , defines a small wavefront with the two signs of v yielding the incoming and outgoing fronts. For sufficiently large $|v|$, these fronts could develop singularities. For examples, see Figs. 6 and 7.

Eqs.(89) and (90) define a (three dimensional) Lagrangian submanifold in the six dimensional phase space of (\mathbf{r}, \mathbf{p}) , in terms of the parameters v and s^J , while the Lagrangian map π is given by Eqs.(89).

Now with the use of generating families, this example generalizes (from 2 to 3 dimensional configuration spaces), the construction $\spadesuit c$ from Sec.II.

The extension of this construction to a Legendrian submanifold, E , consists of simply adding t as the contact coordinate and using $t = v$ with Eqs.(89) and (90) to define E , i.e.,

$$\mathbf{r} = \mathbf{r}_0(s^J) + v\mathbf{n}(s^J), \quad (91)$$

$$\mathbf{p} = \mathbf{n}(s^J), \quad (92)$$

$$t = v, \quad (93)$$

while the projection, the Legendrian map, to the space-time, (\mathbf{r}, t) , is given by

$$\mathbf{r} = \mathbf{r}_0(s^J) + v\mathbf{n}(s^J), \quad (94)$$

$$t = v. \quad (95)$$

(Compare with $\spadesuit f$ of Sec.II.) Qualitatively these examples can be generalized to arbitrary four dimensional Lorentzian space-times.^{13,16,22}

As was stated earlier the critical points of the Lagrangian map can be calculated either from the vanishing of the Jacobian of that map or from

$$D(s^J, v) \equiv \left| \frac{\partial^2 \mathfrak{F}}{\partial s^J \partial s^K} \right| = 0.$$

Directly from the latter expression we have, after a brief calculation, that

$$\frac{\partial^2 \mathfrak{F}}{\partial s^J \partial s^K} = v^{-1}(g_{JK} - v h_{JK}) \quad (96)$$

where $g_{JK} = \mathbf{T}_J \cdot \mathbf{T}_K$ and $h_{JK} = \mathbf{n}(s^J) \cdot \partial \mathbf{T}_K / \partial s^J$ are respectively the first and second fundamental forms (or respectively, the induced metric and extrinsic

curvature tensors) of the surface, \mathfrak{S} . The critical points (determined by the vanishing of the determinant D , of Eq.(96)) are then given by the values of $v = |\mathbf{r} - \mathbf{r}_0|$ such that

$$v^2 D = g + v(g_{11}h_{22} - 2g_{12}h_{12} + g_{22}h_{11}) + v^2 h = 0 \quad (97)$$

where g and h are the determinants of g_{JK} and h_{JK} . The two roots $v_1(s^J)$ and $v_2(s^J)$, of Eq.(97), can be recognized from the differential geometry of 2-surfaces in \mathbb{R}^3 , as defining the two principal radii of curvature of \mathfrak{S} , and we have the classical result that:

- The caustic of a two-surface \mathfrak{S} , acting as a light-source, consists of the loci of the principal centers of curvature of that surface and are given by

$$\mathbf{r}(s^J) = \mathbf{r}_0(s^J) + \mathbf{n}v_1(s^J), \quad (98)$$

$$\mathbf{r}(s^J) = \mathbf{r}_0(s^J) + \mathbf{n}v_2(s^J). \quad (99)$$

In other words, it consists of two different two-surfaces, each given parametrically in terms of s^J by Eqs.(98) and (99). These two surfaces touch each other whenever $v_1(s^J) = v_2(s^J)$; in other words on the normals from *umbilic* points of \mathfrak{S} where the two radii of curvature coincide.^{23,24} On the caustic point, associated with the umbilic point, there occurs what is called an “umbilic” singularity. Other “singularities” of the caustic surfaces, which are cusp ridges and swallowtails, can be analyzed in terms of the local differential geometry of the surface \mathfrak{S} . They are associated with extremals of the curvatures ($k_1 = v_1^{-1}$, $k_2 = v_2^{-1}$) on the curves of a principal curvature coordinate system.

[An alternative way to obtain Eq.(97) is to calculate and set to zero the Jacobian of Eq.(89);

$$J = \left| \frac{\partial \mathbf{r}}{\partial p}, \frac{\partial \mathbf{r}}{\partial s^1}, \frac{\partial \mathbf{r}}{\partial s^2} \right| = \mathbf{n} \cdot \left\{ (\mathbf{T}_1 + v \frac{\partial \mathbf{n}}{\partial s^1}) \times (\mathbf{T}_2 + v \frac{\partial \mathbf{n}}{\partial s^2}) \right\} = 0. \quad (100)$$

By using \mathbf{n} from Eq.(88) and the identity $(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = (\mathbf{A} \cdot \mathbf{C})(\mathbf{B} \cdot \mathbf{D}) - (\mathbf{A} \cdot \mathbf{D})(\mathbf{B} \cdot \mathbf{C})$ with the definition of the first and second fundamental form, Eq.(100) is seen to be identical to Eq.(97).]

Remark 9 We mention, without entering into the details, that from Eqs.(84),(85) and (87) one can construct one of several generating functions for this case. A typical one, valid if $n_z \neq 0$, takes the form

$$G(z, p_x, p_y) = zn_z(s^J) - \mathbf{r}_0(s^J) \cdot \mathbf{n}(s^J) \quad (101)$$

where the s^J are given implicitly as functions of the (p_x, p_y) by $(p_x, p_y) = (n_x, n_y)$.♠

A much larger class of examples to which generating families can be applied is given by the following:

♠. Consider any (autonomous) Hamiltonian system with phase space coordinates (q^a, p_a) and Hamiltonian

$$H = H(q^a, p_a), \quad a = 1 \dots n, \quad (H : T^*M \rightarrow \mathbb{R})$$

and associated (H-J) equation

$$\frac{\partial S}{\partial t} + H(q^a, \frac{\partial S}{\partial q^a}) = 0 \quad (S : M \times \mathbb{R} \rightarrow \mathbb{R}). \quad (102)$$

(The following considerations apply equally to the general relativistic H-J equation,

$$g^{ab}\partial_a S \partial_b S + m^2 = 0.)$$

We use the existence of an n parameter family, (s^a) , of solutions to the H-J equation, a “complete solution”,

$$S = S^*(q^a, s^b, t),$$

i.e., a solution depending on n parameters s^b , such that the equation

$$\frac{\partial S^*}{\partial s^a} = \alpha_a$$

can be solved uniquely with respect to the variables q^a .

We now define what is to be our generating family, the function $\mathfrak{F}(q^a, s^b, t)$ by

$$\mathfrak{F} = S^*(q^a, s^b, t) - F(s^b) \quad (103)$$

where $F(s^b)$ is an arbitrary function such that \mathfrak{F} obeys the rank condition. We now require, from the theory of generating families, the extremal condition, i.e., Eq.(72), that

$$\partial\mathfrak{F}/\partial s^b \equiv \partial S^*/\partial s^b - \partial F/\partial s^b = 0. \quad (104)$$

From the general theory we infer that Eq.(104) can be solved for either s^b or q^a or some combination of them or alternatively it allows us to describe the solution parametrically, i.e.,

$$s^a = S^a(y^b, t), \quad q^a = Q^a(y^b, t).$$

Moreover, the resulting equations

$$\begin{aligned} q^a &= Q^a(y^b, t), \\ p_a &= \frac{\partial}{\partial q^a} \mathfrak{F}(q^a, s^b, t) = P_a(y^b, t), \end{aligned}$$

define a one-parameter family of Lagrangian submanifolds of T^*M , parametrized by t .

They also define a Lagrangian submanifold of the phase space over $M \times \mathbb{R} = \{q^a, t\}$ with canonical coordinates $(q^a, t, p_a, -E)$, contained in the “physical hypersurface” (constraint submanifold, \mathfrak{E}) given by $E = H(q^a, p_a)$. A “classical” solution $S(q^a, t)$ to the H-J equation may be geometrically characterized as the generating function of a Lagrangian section of $T^*(M \times \mathbb{R})$ contained in \mathfrak{E} . Therefore, it is reasonable to call any generating function of any Lagrangian submanifold L , contained in \mathfrak{E} , [explained earlier in connection with generating families, see Eq.(84)] a “generalized solution” of the H-J equation since it extends thru singularities the Lagrangian submanifold defined (locally) by a “classical” solution S . [Such an extension is unique since L is ruled by phase trajectories determined by Hamilton’s equations and initial conditions.]

Note that indeed a generating function of this type does satisfy the generalized H-J equation;

$$H(q^a, p_a) \equiv H(q^A, q^J, p_A, p_J) \Rightarrow H(q^A, -\partial^J K, \partial_A K, p_J) + \frac{\partial K}{\partial t} = 0,$$

where we have used Eq.(5).

In this sense, the construction, via Eqs.(103) and (104), provides “generalized” solutions of the H-J equation in that it extends an ordinary solution thru singular points.¹⁶ It is also a “general” solution, in the sense that by a suitable choice of a complete solution $S^*(q^a, s^b, t)$ of the H-J equation and the function $F(s^b)$, it can be adapted to any Cauchy data. If one begins with arbitrary Cauchy data, $S_0(q^a)$, it is possible to convert it into an expression for $F(s^b)$. See the corresponding argument for the eikonal equation in Sec.IV.

This procedure allows us to choose an ensemble (of non-interacting particles in three space) for a classical system and then see how the entire ensemble evolves and study the density waves; for example the high density in the neighborhood of caustics. See Sec.III. ♠

♠An example closely related to the preceding one is that of the eikonal equation in an arbitrary four dimensional Lorentzian manifold, namely

$$g^{ab}(x^c)\partial_a S\partial_b S = 0, \quad (105)$$

where g^{ab} is the inverse metric and x^c are some local coordinates. This time we start from a two parameter family of solutions,

$$S = Z(x^c, \zeta, \bar{\zeta}) \quad (106)$$

with $(\zeta, \bar{\zeta})$ the parameters, chosen as the complex stereographic coordinates on S^2 in order to label the sphere of null directions at the space-time point x^c . (We write expressions in terms of both $(\zeta, \bar{\zeta})$ in order to point out that the functions used are not holomorphic in ζ . Also it is convenient to employ the independent directional derivatives with respect to $(\zeta, \bar{\zeta})$.) It is often difficult or even practically impossible to find such solutions though there are perturbation techniques to construct approximations to such solutions.

We now define our generating family by^{25,22}

$$\mathfrak{F}(x^c, \zeta, \bar{\zeta}) = Z(x^c, \zeta, \bar{\zeta}) - F(\zeta, \bar{\zeta}) \quad (107)$$

with $F(\zeta, \bar{\zeta})$ an arbitrary function. (Often F is chosen as a regular function on S^2 , though this is not necessary.) The extremal condition, Eq.(72), is now

$$\partial\mathfrak{F}/\partial\zeta = 0, \quad \partial\mathfrak{F}/\partial\bar{\zeta} = 0. \quad (108)$$

If these equations can be solved by $\zeta = \zeta(x^a)$, then $\mathfrak{F}[x^a] = \mathfrak{F}(x^c, \zeta(x^a), \bar{\zeta}(x^a))$ also satisfies the eikonal equation. (See Sec. IV for the flat -space version of

this with three parameters.) Note that this construction *does not* allow the construction of the general solution; to do that $F(\zeta, \bar{\zeta})$ would have to depend on three parameters rather than two. However this procedure does allow the construction of any arbitrary single null hypersurface.²² The first example of this section is an illustration of this construction. A much more valuable example is the construction²⁶ of the light-cone of some arbitrary but fixed space-time point, x_0^a . This can be used to generalize the usual treatment of gravitational lensing. In fact in this case, Eq.(109) below, becomes the lens equation when two of the coordinates, the radial distance from an observer and the time, are held constant.²⁷

If \mathfrak{F} is held constant (say zero) then the three equations, Eqs.(107) and (108), can be solved for *three*, (x^i) , of the four coordinates, x^a , in terms of the fourth one, x^* , i.e., they have the form

$$x^i = X^i(x^*, \zeta, \bar{\zeta}). \quad (109)$$

Eq.(109) represents the set of null geodesics that generate the big wavefront, $\mathfrak{F} = 0$; they are labeled by the $(\zeta, \bar{\zeta})$ and parametrized by one of its coordinates, x^* .♠

♠As a final example, we show how the N-plane lens map (in the standard weak field, thin lens approximation) of gravitational lens theory can be obtained naturally via a generating family, as a Lagrangian projection.

Suppose that the lightrays emitted from some point source \mathfrak{s} on a source plane P are consecutively gravitationally deflected by N thin mass distributions, M_i before they reach an observer \mathfrak{D} . The M_i are represented by surface mass densities in N planes P_i orthogonal to a straight line going thru \mathfrak{D} and perpendicular to the source plane P. A virtual light path is represented as a polygon figure from \mathfrak{D} to \mathfrak{s} with vertices on the P_i . The influence of the M_i on light can be expressed in terms of two-dimensional potentials Ψ_i on the planes P_i . The travel time of a lightray depends not only on its geometrical path length, but also on the gravitational Shapiro-time delay suffered when the rays passes a “lens” M_i . If the positions \mathfrak{s}_i of a virtual ray on the plane P_i and the position $\mathfrak{q} \equiv \mathfrak{s}_{N+1}$ of \mathfrak{s} on a source plane P, are scaled suitably, the (variable part) of the travel time has the form⁹

$$\mathfrak{F}(\mathfrak{s}_1, \mathfrak{s}_2, \dots, \mathfrak{s}_N, \mathfrak{q}) = \sum_{i=1}^N C_i \left[\frac{1}{2} (\mathfrak{s}_i - \mathfrak{s}_{i+1})^2 - \beta_i \Psi_i(\mathfrak{s}_i) \right] \quad (110)$$

where the constants C_i, β_i depend on the distances of \mathfrak{s} and the M_i from \mathfrak{D} . Fermat's principle (which singles out the "real" from the virtual rays) in this idealization takes the form

$$\frac{\partial \mathfrak{F}}{\partial \mathfrak{s}_i} = 0. \quad (111)$$

We consider $\mathfrak{F}(\mathfrak{s}_1, \mathfrak{s}_2, \dots, \mathfrak{s}_N, \mathfrak{q})$ as a generating family with - in the notation of Sec.V -

$$s^J = (\mathfrak{s}_1, \mathfrak{s}_2, \dots, \mathfrak{s}_N), \quad q^a = \mathfrak{q} = \mathfrak{s}_{N+1},$$

i.e., in this case we have $n = 2$ and $m = 2N$. The rank condition is satisfied; indeed the solution from Eq.(111) has the form

$$\mathfrak{s}_2 = f_2(\mathfrak{s}_1), \quad \mathfrak{s}_3 = f_3(\mathfrak{s}_2, \mathfrak{s}_1) = f_3[\mathfrak{s}_1], \quad \dots, \quad \mathfrak{q} = f_{N+1}(\mathfrak{s}_1).$$

If we put

$$\mathfrak{p} = \frac{\partial \mathfrak{F}}{\partial \mathfrak{q}} = P(\mathfrak{s}_1)$$

then according to the general theory the equations

$$\mathfrak{q} = f_{N+1}(\mathfrak{s}_1), \quad \mathfrak{p} = P(\mathfrak{s}_1)$$

describe a Lagrangian submanifold of $T^*\mathbb{R}^2 = \{\mathfrak{q}, \mathfrak{p}\}$, parametrized by the ray direction (corresponding to \mathfrak{s}_1) at the observer \mathfrak{D} . The associated Lagrangian projection is given by the lens map

$$\mathfrak{s}_1 \mapsto \mathfrak{q} = f_{N+1}(\mathfrak{s}_1)$$

which takes a ray direction \mathfrak{s}_1 at \mathfrak{D} to the source position \mathfrak{q} . (Note that this is a gradient map if $N=1$ but a more general Lagrangian map for $N > 1$.) Critical curves, caustics, types of singularities then can be analyzed according to the general theory.♠

We mention, with no discussion, that Lagrangian submanifolds play a role in the characterization and construction of physical states of (linear) quantum fields on (classical) curved, globally hyperbolic space-times (M, g) . Such (Hadamard) states can be characterized by the "wave front sets" of their two-point "functions", subsets of $T^*(M \times M)^{28,29}$ which have been shown to be contained in Lagrangian submanifolds of $T^*(M \times M)$.

VIII. Epilogue

The study of caustics and wave front evolution has a rich history; it dates back to the early studies of Newton and Huygens, Cayley studied the normal wavefront evolution from the triaxial ellipsoid in the middle of the 19th century. The contemporary study of generic wavefront and caustic behavior arose in the mid-century via the classification studies of the singularities of functions and mappings. It arose mainly via the efforts of the mathematicians, H. Whitney, R. Thom and V.I. Arnold; the work of the latter on Lagrangian maps has been the main concern here. With several notable exceptions, in particular M. Berry and Ya. Zeldovich, physicists seem to have largely ignored the subject even though it has implications for a wide range of physical applications; all forms of wave propagation, both classical and quantum mechanical; from geometric optics thru to physical optics; intensity distributions in interference and diffraction phenomena³⁰ (e.g., evaluations of the Fresnel and Airy integrals); gravitational lensing; structure formation in the early universe and in galaxies via density waves; finite size image disruption³¹; Hamilton-Jacobi theory; stability problems; thermodynamics; elasticity theory; and states of quantum fields in curved space-times.²⁹

We have only attempted to give the most rudimentary treatment of the basic mathematical ideas that lie at the origin of this large subject and to introduce several potential applications of the general theory to physics. For a variety of reasons we have avoided completely several relevant topics, e.g., global topological questions, the theory of classification of critical points of functions, the surprising relationship between the classification of functions and the Weyl groups, etc.

There are several articles and books which contain extensive bibliographies and historical surveys of the origin and development of singularity theory. The book *Catastrophe Theory*,² besides being a wonderful introduction to the subject, contains both a brief history and an extensive annotated bibliography to both the theory and its many applications. The article in Russian Math. Surveys,³² dedicated to Arnold on his 60th birthday, contains a complete list of Arnold's publications while Arnold's article,¹⁵ on large scale issues in wave propagation (and as a delightful aside, a discussion on Mathematics Education), also has a large bibliography as does Arnold's article⁵ in Vol.V in Dynamical Systems of the Enc. of Math. Sciences. Though Arnold's book³³ "*Huygens & Barrow, Newton & Hooke*" only touches on the details

of singularity theory it must be mentioned for its wealth of fascinating historical observations. We point out that though most of our references are to books and articles published later than 1980 almost all of the fundamental mathematical work was completed by the mid 1970's. We list several of the principle early references.^{34,35,36,37}

V.I. Arnold on Mathematicians

“It is almost impossible for me to read contemporary mathematicians who instead of saying

‘Petya washed his hands’ write simply

‘There is a $t_1 < 0$ such that the image of t_1 under the natural mapping $t_1 \mapsto \text{Petya}(t_1)$ belongs to the set of dirty hands and a $t_2, t_1 < t_2 < 0$, such that the image of t_2 under the above mentioned mapping belongs to the complement of the set defined in the preceding sentence’ ”.

IX. Acknowledgments

ETN thanks Simonetta Frittelli and Carlos Kozameh for the many stimulating and enlightening “fights” and discussions which played a major role in clarifying his understanding of this subject. He also thanks the NSF for support under research grants # PHY 92-05109 and PHY 97-22049. JE is grateful to Peter Schneider and his group for introducing him to lensing, to Helmut Friedrich for clarifying remarks on the theory of wavefronts and to Thomas Buchert for sharing with him thoughts on applications of singularity theory to cosmology.

We extend a special thanks to V.I. Arnold for his help in clarifying certain historical issues and for help in greatly expanding our bibliography.

In particular we offer an extra special thanks to Simonetta Frittelli for the preparation of the figures used here.

References

- ¹V. I. Arnold, S.M. Gusein-Zade, A. N. Varchenko, *Singularities of Differentiable Maps, Vol. I*, Birkhauser, Boston, Basel, Stuttgart, (1985).
- ²V. I. Arnold, *Catastrophe Theory*, 3rd revised edition, Springer-Verlag, Berlin, Heidelberg, NY, Tokyo, (1992). This contains an excellent history and bibliography of the subject.
- ³V. I. Arnold, *Mathematical Methods of Classical Mechanics*, 2nd Edition, Springer-Verlag, Berlin, Heidelberg, NY, (1989).
- ⁴V. I. Arnold, S.P. Novikov, eds., Enc. of Math. Sciences, *Dynamical Systems, IV*, Springer-Verlag, Berlin, Heidelberg, NY, (1990).
- ⁵V.I. Arnold, eds., Enc. of Math. Sciences, *Dynamical Systems, V*, Springer-Verlag, Berlin, Heidelberg, NY, (1994).
- ⁶V. I. Arnold, *Singularities of Caustics and Wave Fronts*, Kluwer, Dordrecht, (1990).
- ⁷S. Frittelli, C. Kozameh, E. Newman, *GR Via Characteristic Surfaces*, J. Math. Phys. **36** , 4984, (1995).
- ⁸S. Frittelli, C. Kozameh, E. Newman, *On the Dynamics of Characteristic Surfaces*, J. Math. Phys. **36** , 6397, (1995).
- ⁹P. Schneider, J. Ehlers, E.E. Falco, *Gravitational Lenses*, Springer-Verlag, New York, Berlin, Heidelberg, (1992).
- ¹⁰A.O. Petters, *Arnold's Singularity Theory and Gravitational Lensing*, J. Math.Phys, **34**, 3555, (1993).
- ¹¹Ya. B. Zeldovich, *Gravitational Instability: An Approximate Theory for Large Density Perturbations*, Astron.Astrophys.5 (1970).
- ¹²A rather abstract but general reference to much of the material here is *Symplectic Geometry and Analytic Mechanics* by Paulette Libermann and Charles-Michel Marle, D. Reidel Publishing Co. Dordrecht, Holland, (1997).

- ¹³The treatment by Friedrich, H. and Stewart, J., *Characteristic Initial Data and Wavefront Singularities in General Relativity*, Proc. Roy. Soc. A, **385**, 345-371, (1983) of some of the same material is also quite abstract.
- ¹⁴J.W. Bruce and P.J. Giblin, *Curves and Singularities*, Cambridge University Press, 2nd edition, (1992). The Russian edition (1988) contains an historical survey.
- ¹⁵V.I. Arnold, *Topological Problems of the Theory of Wave Propagation*, Russian Math. Surveys **51**:1 1-47, (1996).
- ¹⁶W. Hasse, M. Kriele, V. Perlick, *Caustics of Wavefronts in General Relativity*, Class. Quantum Grav. **13**, 1161 (1996).
- ¹⁷M.V. Berry and C. Upstill, *Catastrophe Optics: Morphologies of Caustics and Their Diffraction Patterns*, ed. E Wolf, Progress in Optics XVIII, North Holland (1980).
- ¹⁸A. Einstein, *On the Quantum Theory of Sommerfeld and Epstein*, (1917), in the collected works of A. Einstein, Vol.VI, Doc.45.
- ¹⁹V. I. Arnold, S.F. Shandarin, Ya. B. Zeldovich, *The Large Scale Structure of the Universe, I, General Properties; One and Two-Dimensional Models*, Geophys. Astrophys. Fluid Dyn. 20, (1982).
- ²⁰S. F. Shandarin and Ya. B. Zeldovich, *The Large-Scale Structure of the Universe: Turbulence, Intermittency, Structures in a Self-Gravitating Medium*, Rev. Mod. Phys., **61**, 185, (1989).
- ²¹L. Landau and Lifschitz, *Classical Mechanics*, Pergamon Press, Headington Hill Hall, Oxford, 4 & 5 Fitzroy Sq. London, W.1, (1960).
- ²²S. Frittelli, E. Newman, G. Silva-Ortigoza, *The Eikonal Equation in Asymptotically Flat Space-Times*, J. Math. Phys., **40**, 1041,(1999).
- ²³W. Klingenberg, *A Course in Differential Geometry*, Springer-Verlag, New York, Berlin, Heidelberg, (1978).
- ²⁴W. Blaschke, *Differentialgeometrie*, Springer, Berlin, (1945), especially chapter 9.

- ²⁵S. Frittelli, E. Newman, G. Silva-Ortigoza, *The Eikonal Equation in Flat Space-Times*, I,
J. Math. Phys., **40**, 383, (1999).
- ²⁶S. Frittelli, E. Newman, *A Universal Lensing Equation*, to be published,
Phys. Review D.
- ²⁷J. Ehlers, S. Frittelli, E. Newman, *Gravitational Lensing from a Space-Time Perspective*, to be published in a volume of the Boston Studies in the Philosophy of Science; Festschrift for John Stachel, Ed. Juergen Renn, Kluwer Academic Press, Dordrecht, (2000)
- ²⁸M.J. Radzikowski, Ph.D.thesis, Princeton University, October 1992
- ²⁹W. Junker, Ph.D.thesis, Hamburg University, July 1995
- ³⁰M.V. Berry, *Twinkling Exponents in Catastrophe Theory of Random Short Waves*, in *Wave Propagation and Scattering*, ed. B.J. Uscinski, Clarendon Press, Oxford, (1986).
- ³¹M.V. Berry, *Disruption of Images; the Caustic Touching Theorem*, J. Opt. Soc. Amer., A4, 561, (1987).
- ³²*Vladimir Igorevich Arnol'd on his Sixtieth Birthday*, Russian Math. Surveys, **52:5**, 235-255, (1997).
- ³³V.I. Arnold, "*Huygens & Barrow, Newton & Hooke*", Birkhauser Verlag, Basel, Boston, Berlin, (1990)
- ³⁴V.I. Arnold, *On a Characteristic Class Entering into Conditions of Quantification*, Functional Anal.Appl. **1**,(1967).
- ³⁵V.I. Arnold, *Integrals of Rapidly Oscillating Functions and Singularities of the Projections of Lagrange manifolds*, Functional Anal.Appl. **1**,(1972).
- ³⁶V.I. Arnold, *Normal Forms of Functions near Degenerate Critical Points, the Weyl Groups A_k, D_k, E_k* , Functional Anal.Appl. **1**,(1972).
- ³⁷V.I. Arnold, *Wave Front Evolution and Equivariant Morse Lemma* , Comm. Pure and Applied Math. **29**,(1976).