May 22, 2012

Solving (Nonlinear) First-Order PDEs

Cornell, MATH 6200, Spring 2012

Final Presentation

Zachary Clawson

Abstract

Fully nonlinear first-order equations are typically hard to solve without some conditions placed on the PDE. In this presentation we hope to present the Method of Characteristics, as well as introduce Calculus of Variations and Optimal Control. The content in the Method of Characteristics section is directly from Evans, sometimes with more detail.

Contents

1 Introduction

Solving nonlinear first-order PDEs in complete generality is something we are only able to do locally, and only most of the time. Smooth solutions may not exist at all points in a specified, 'nice,' domain Ω. We are only able to provide smoothness on a neighborhood of $\Gamma \subseteq \partial\Omega$. Continuity of solutions may also become an issue, and we are able to define *viscosity solutions*, which are analogous to the concept of weak solutions.

The general setting we will be applying ourselves in is solving a PDE on $\Omega \subseteq \mathbb{R}^n$:

$$\begin{cases} F(Du, u, \mathbf{x}) = 0 & \text{in } \Omega \\ u(\mathbf{x}) = g(\mathbf{x}) & \text{on } \partial\Omega. \end{cases} \tag{1.1}$$

We see in the following example that we may not always have smooth solutions in the nonlinear cases, despite smoothness of $\partial\Omega$..

1.1 Example: Eikonal solution on a square

Example 1.1. Consider the domain $\Omega = [0,1]^2 \subsetneq \mathbb{R}^2$

$$\begin{cases} \frac{1}{2}u_x^2 + \frac{1}{2}u_y^2 = \frac{1}{2} & \text{in } \Omega \\ u = 0 & \text{on } \Gamma = \partial\Omega. \end{cases} \tag{1.2}$$

The characteristics of (1.2) are given by $\mathbf{x}(s) = \nabla u(\mathbf{x}(s))$ with $\mathbf{x}(0) \in \partial\Omega$, and hence at any position in space we are traveling in the direction of maximal increase for u. Since the solution to the Eikonal equation gives the distance to $\partial\Omega$, this direction is directly away from the boundary within the divided triangular regions in the Figure.

While the solution here is continuous, we lack smoothness along the diagonals of the square. It is also worth noting we draw the characteristics from the interior to the boundary in Figures 1 and 2. See the following remark.

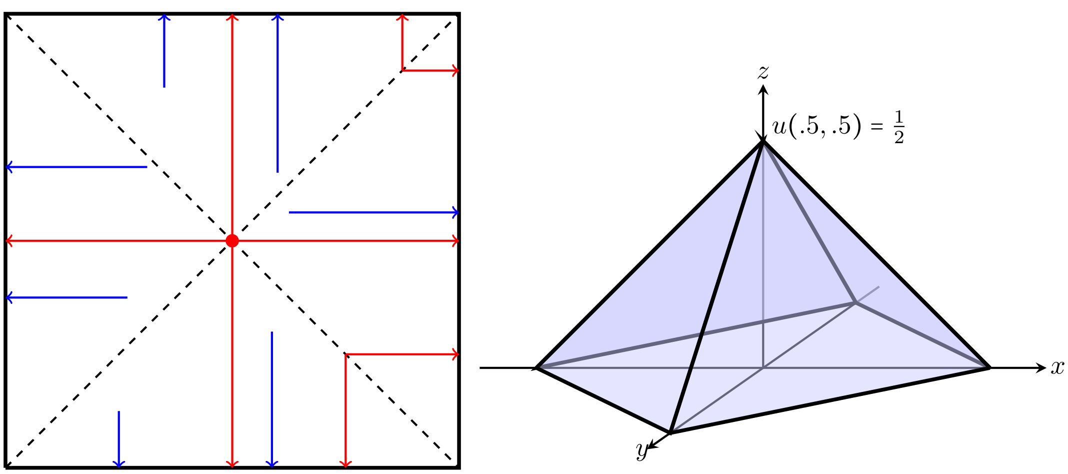

Figure 1: This diagram represents the division of our domain $\Omega = [0,1]^2$ into four regions in which the optimal method of "escape to the boundary" is the straight-line path. The value function u here represents the total time (or distance) it takes for any point to escape to $\partial\Omega$. This gives us an idea of what the solution should look like in 3D – a pyramid. This solution is thus not smooth along the dashed diagonals in the figure.

We also note that in the case that the domain is the unit ball, $\Omega = B(\mathbf{0}, 1) \subsetneq \mathbb{R}^2$, we have a similar resulting discontinuity:

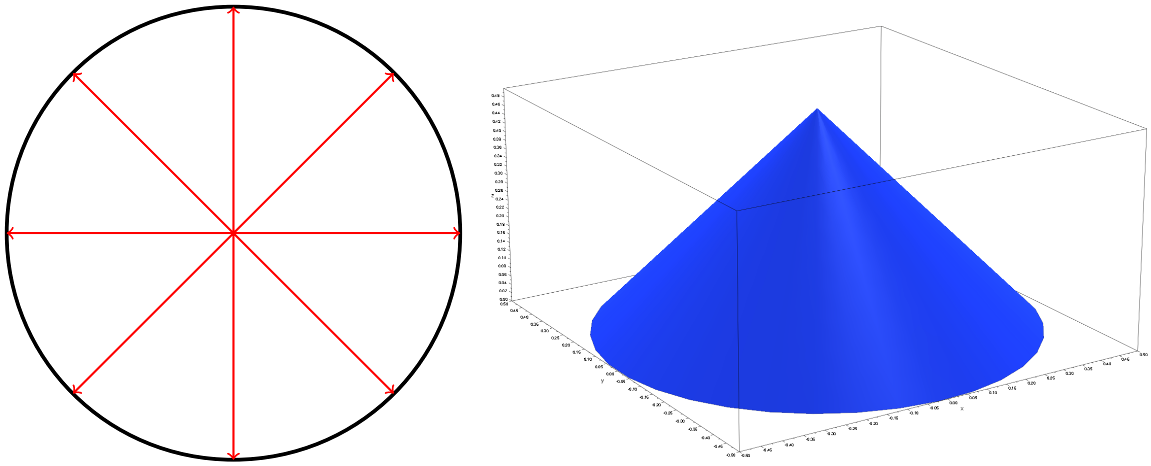

Figure 2: We also see that in the case when we are attempting to escape from a circle, and the boundary data is zero everywhere (same problem) that we also have an ‘unexpected’ discontinuity in the value function.

Remark 1.2. The characteristics drawn in both Figure 1 and Figure 2 above are pointing from an initial point in the domain out to the boundary. The discussion in §2 concerning the Method of Characteristics would produce the opposite – the characteristics would run *from the boundary inwards*.

1.2 Motivational example: Eikonal equation

1.2.1 Optimization lemma

Lemma 1.3 (Parallelogram equality)**.** *If* $\langle\cdot,\cdot\rangle$ *is an inner product on a Hilbert space, then the induced norm satisfies*

$$\|x+y\|^2+\|x-y\|^2=2\left(\|x\|^2+\|y\|^2\right).$$

Proof. Let $\|x\| \triangleq \sqrt{\langle x,x\rangle}$ for $x\in H$, the Hilbert space. Then

$$\begin{aligned}\|x+y\|^2+\|x-y\|^2 &= \langle x+y,x+y\rangle+\langle x-y,x-y\rangle\\ &= \langle x,x+y\rangle+\langle y,x+y\rangle+\langle x,x-y\rangle-\langle y,x-y\rangle\\ &= \langle x,x\rangle+2\langle x,y\rangle+\langle x,x\rangle-\langle x,y\rangle-\langle y,x\rangle+\langle y,y\rangle\\ &= 2\left(\|x\|^2+\|y\|^2\right).\end{aligned}$$

□

Lemma 1.4 (Optimization)**.** *Let* $\Omega\subseteq H$ *be convex and complete, where* H *is a Hilbert space. Then*

$$\forall\ x\in H,\ \exists!\ y_x\in\Omega \text{ such that } \delta_x=\|x-y_x\|=\inf_{z\in\Omega}\|x-z\|, \tag{1.3}$$

where $\|\cdot\|$ *is the induced norm by* H*'s inner product:* $\|x\|=\sqrt{\langle x,x\rangle}$.

Proof. **(a) Existence.** Fix $x\in H$. If $x\in\Omega$, then $y_x=x$ gives the desired result for the infemum. Hence assume $x\in H\smallsetminus\Omega$. Since δ_x is given by an infemum, we have by definition $\exists\{y_n\}_{n\in\mathbb{N}}$ such that

$$\|y_n-x\|\to\delta_x, \tag{1.4}$$

the latter by continuity of the norm (induced by the inner product – see page 138 of [K] for continuity of the inner product). The objective is to now show that $\{y_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence.

Show $\{\mathbf{y_n}\}$ is Cauchy. Let $v_n=y_n-x$, so $\|v_n\|=\|y_n-x\|$. Then

$$\|v_n+v_m\|=\|y_n+y_m-2x\|=2\left\|\frac{1}{2}(y_n+y_m)-x\right\|=2\left\|y'-x\right\|\geq 2\delta_x,$$

where $y' = \frac{1}{2}(y_n + y_m) \in \Omega$ by convexity. Now using that $y_n - y_m = v_n - v_m$ and the Parallelogram equality (second equality), we have

$$\|y_n - y_m\| = \|v_n - v_m\| = -\|v_n + v_m\|^2 + 2\left(\|v_n\|^2 + \|v_m\|^2\right) \leq -(2\delta_x)^2 + 2\left(\|y_n - x\|^2 + \|y_m - x\|^2\right).$$

Since we have both $\|y_n - x\| \to \delta_x$ and $\|y_m - x\| \to \delta_x$ by (1.4), the above gives that the RHS converges to 0 as $n, m \to \infty$. This then gives $\|y_n - y_m\| \to 0$ as $n, m \to \infty$, and hence $\{y_n\}_{n\in\mathbb{N}}$ is Cauchy.

With this in hand, using the completeness of Ω, we now have that $\exists\ y_x \in \Omega$ such that $y_n \to y_x$ and $\delta_x = \|x - y_x\|$. We have this since

$$\|x - y_x\| \leq \|x - y_n\| + \|y_n - y_x\| = \delta_{y_n} + \|y_n - y\| \to \delta_x + 0 = \delta_x \text{ as } n \to \infty.$$

Since the LHS is independent of n, this proves that $\|x - y_x\| = \delta_x$.

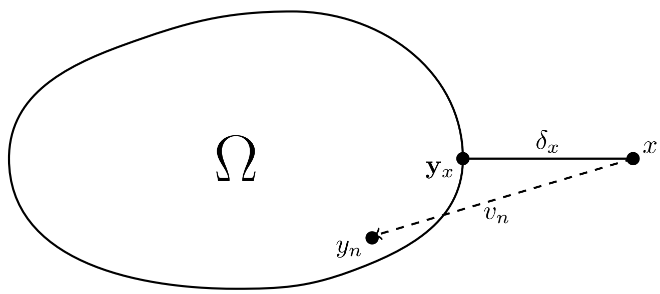

Figure 3: This diagram shows the convex space Ω with a fixed $\mathbf{x} \in H \setminus \bar{\Omega}$ provides a unique minimizer for $\delta_{\mathbf{x}}$ given by $\mathbf{y_x}$. Remember that $\delta_{\mathbf{x}}$ is the minimal distance from $\mathbf{x}$ to Ω (i.e. $\partial\Omega$).

(b) Uniqueness. Assume both y_x and y'_x both satisfy (1.3), and hence $\delta_x = \|x - y_x\| = \|x - y'_x\|$. Since H is an inner product space, we have the Paralellogram equality:

$$\begin{aligned}
\|y_x - y'_x\|^2 &= \|(y_x - x) - (y'_x - x)\|^2 \\
&= 2\|y_x - x\|^2 + 2\|y'_x - x\|^2 - \|(y_x - x) + (y'_x - x)\|^2 \\
&= 2\delta_x^2 + 2\delta_x^2 - 4\left\|\frac{1}{2}(y_x + y'_x) - x\right\|^2 \\
&= 4\delta_x^2 - \|Y_x - x\|^2,
\end{aligned}$$

where $Y_x = \frac{1}{2}(y_x + y'_x) \in \Omega$ by convexity. Since $Y_x \in \Omega$, we then have $\|Y_x - x\| \geq \delta_x$. Therefore the above results in

$$\|y_x - y'_x\|^2 \leq 4\delta_x^2 - 4\delta_x^2 = 0 \implies \|y_x - y'_x\| = 0 \implies y_x = y'_x.$$

This provides uniqueness. □

Remark 1.5. If $x \in H \setminus \Omega$, then $\delta_x = \inf_{z\in\partial\Omega} \|x - z\|$. That is, the unique minimizer is actually on the boundary, which makes sense geometrically.

Remark 1.6. There are other ways to pose this problem, for example if Ω is a subspace instead of convex the theorem also holds true (but convexity does not imply that a set is a subspace). Closure of Ω is equivalent to completeness here.

1.2.2 Example: Lemma applied to Eikonal equation

Eikonal equation. The Eikonal equation on a space X is given by ($\mathbf{x} \in X$)

$$\begin{cases} |Du(\mathbf{x})| = f(\mathbf{x}) & \text{in } X \\ u(\mathbf{x}) = g(\mathbf{x}) & \text{on } \Gamma \subseteq \partial X. \end{cases} \tag{1.5}$$

Theorem 1.7 (Arnold)**.** *Given $\Omega \subsetneq \mathbb{R}^n$ is a closed and convex subset, defining $X = \mathbb{R}^n \setminus \Omega$ with $f(\mathbf{x}) \equiv 1$ on $\mathbb{R}^n$ and $g(\mathbf{x}) \equiv 0$ on $\partial X = \partial\Omega = \Gamma$, we have the value function $u(\mathbf{x})$ gives the minimal distance to Ω.*

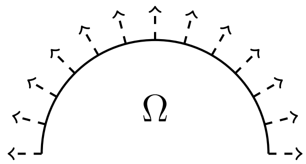

Figure 4: The characteristic curves here are pointing directly outward from $\Gamma = \partial\Omega$ when solving the problem on $H \setminus \Omega$, and they record the distance (time) traveled.

Proof. By the preceding Lemma and following remark, we have that for every $\mathbf{x} \in \mathbb{R}^n \setminus \Omega$, $\exists!\ \mathbf{y}_\mathbf{x} \in \partial\Omega$ such that $\delta_\mathbf{x} = \|\mathbf{x} - \mathbf{y}_\mathbf{x}\|$. Therefore, we define $u(\mathbf{x}) = d(\mathbf{x}, \mathbf{y}_\mathbf{x})$. To prove that u satisfies (1.5) we see that

$$u_{x_i}(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}_\mathbf{x}\|^{-1}(x_i - y_{\mathbf{x},i}) = \frac{x_i - y_{\mathbf{x},i}}{\|\mathbf{x} - \mathbf{y}_\mathbf{x}\|},$$

so

$$|Du|^2 = \sum_{i=1}^n u_{x_i}^2 = \sum_{i=1}^n \frac{(x_i - y_{\mathbf{x},i})^2}{\|\mathbf{x} - \mathbf{y}_\mathbf{x}\|} = 1.$$

This proves the Theorem. Note $y_{\mathbf{x},i} = (y_\mathbf{x})_i$. □

Remark 1.8. A solution of the Eikonal equation in this setting will always be a constant plus the distance to the curve $\partial\Omega$.

Example 1.9. Let $\Omega = B(\mathbf{0}, 1)$, i.e. the unit ball in $\mathbb{R}^n$. Then

$$u(\mathbf{x}) = \left\| \mathbf{x} - \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \right\| = \|\mathbf{x}\|_2 - 1$$

is a solution of (1.5) with $g \equiv 0$ on all of $\partial\Omega$, and $f \equiv 1$. We may transform any region Ω to this region and use this as a solution.

Corollary 1.10 (Arnold)**.** *Any solution to* (1.5) *with* $f(\mathbf{x}) \equiv C$ *for some* $C > 0$ *is locally the sum of a constant and the distance to some curve.*

2 Method of Characteristics

This section sets up the Method of Characteristics exactly as Evans does in his text but gives extra detail in some cases. The method of characteristics is one approach to solving the Eikonal equation (1.5) and first order fully nonlinear PDEs.

2.1 Method of Characteristics statement

Our goal is to solve a PDE given by

$$\begin{cases} F(Du, u, \mathbf{x}) = 0 & \text{on } \Omega \\ u = g & \text{in } \Gamma \subseteq \partial\Omega, \end{cases} \tag{2.1}$$

for some open $\Omega \subseteq \mathbb{R}^n$.

Our notation lets $z(s) = u(\mathbf{x}(s))$ and $p_i(s) = u_{x_i}(\mathbf{x}(s))$, so we consider $F(\mathbf{p}, z, \mathbf{x})$. Then the system that we solve provided by the method of characteristics is given by

$$\begin{cases} \text{(a)} & \dot{\mathbf{p}}(s) & = & -D_x F(\mathbf{p}(s), z(s), \mathbf{x}(s)) - D_z F(\mathbf{p}(s), z(s), \mathbf{x}(s))\mathbf{p}(s) \\ \text{(b)} & \dot{z}(s) & = & D_p F(\mathbf{p}(s), z(s), \mathbf{x}(s)) \cdot \mathbf{p}(s) \\ \text{(c)} & \dot{\mathbf{x}}(s) & = & D_p F(\mathbf{p}(s), z(s), \mathbf{x}(s)), \end{cases} \tag{2.2}$$

with

$$F(\mathbf{p}(s), z(s), \mathbf{x}(s)) \equiv 0.$$

The above holds for all s in an interval $I = [0, T]$ for some $T > 0$ (i.e. the solutions will exist locally).

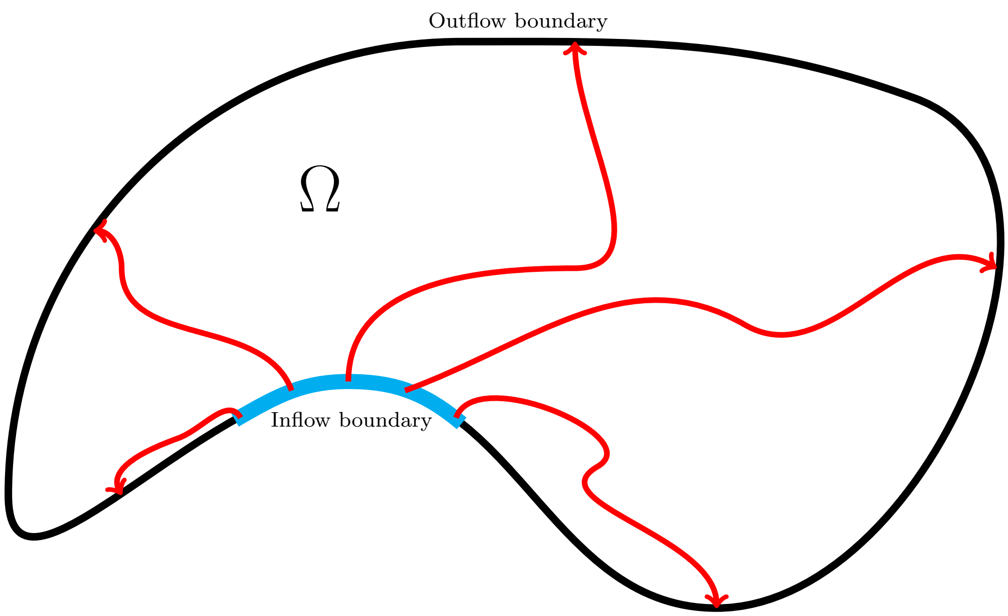

Figure 5: Characteristics emanating from the *inflow boundary* to the *outflow boundary*. The inflow boundary isn't necessarily Γ, but it is the smallest set Γ can be so that we can get full solutions in the above scenario (assuming none of the characteristics curves intersect one another).

2.2 Method of Characteristics derivation

We seek a path $\mathbf{x}(s)$ in Ω defined by the nonlinear PDE, for which we have $u(\mathbf{x}(s))$ solves the PDE along THAT line. The goal is to create these trajectories with initial conditions $\mathbf{x}(0) = x^0 \in \Gamma \subseteq \partial\Omega$ so that we may gain intuition concerning the behavior of u. This technique allows us to 'patch' solutions together within some local neighborhood of a portion of Γ.

We now formally begin the derivation by letting $u \in C^2(\Omega)$ be a solution to (2.1) and once again stating that we set

$$z(s) \triangleq u(\mathbf{x}(s)) \text{ and } \mathbf{p}(s) \triangleq Du(\mathbf{x}(s)),$$

where we aim to choose a $\mathbf{x}(s)$ that allow us to 'easily' compute $z(s)$ and $\mathbf{p}(s)$. We write $p^i(s) = u_{x_i}(\mathbf{x}(s))$. Following Evans, we exam $\dot{\mathbf{p}}(s)$ and use the "chain rule in higher dimensions" to get

$$\dot{p^i}(s) = \sum_{j=1}^{n} u_{x_i x_j}(\mathbf{x}(s))\dot{x}^j(s). \tag{2.3}$$

As Evans remarks this is rather discouraging since it involves second derivatives. Our objective will be to get rid of the second derivative terms with a 'smart selection' of $\mathbf{x}(s)$. Now simply compute $\frac{\partial}{\partial x_i} F(Du, \mathbf{x})$ to get

$$\sum_{j=1}^{n} F_{p_j}(Du, u, \mathbf{x}) u_{x_i x_j} + F_z(Du, u, \mathbf{x}) u_{x_i} + F_{x_i}(Du, u, \mathbf{x}) = 0. \tag{2.4}$$

Evaluating (2.4) at $\mathbf{x} = \mathbf{x}(s)$, while moving the right-two terms onto the other side, we now have

$$\sum_{j=1}^{n} u_{x_i x_j}(\mathbf{x}(s)) F_{p_j}(\mathbf{p}(s), z(s), \mathbf{x}(s)) = -F_{x_i}(\mathbf{p}(s), z(s), \mathbf{x}(s)) - F_z(\mathbf{p}(s), z(s), \mathbf{x}(s)) p^i. \tag{2.5}$$

The LHS of the above is a familiar expression, seen in (2.3). If we were to select $\dot{x}^j(s) = F_{p_j}(\mathbf{p}(s), z(s), \mathbf{x}(s))$ then combining (2.3) and the above will produce that

$$\boxed{\mathbf{p}(s) = -D_x F(\mathbf{p}(s), z(s), \mathbf{x}(s)) - F_z(\mathbf{p}(s), z(s), \mathbf{x}(s))\mathbf{p}.}$$

We now compute $\dot{z}(s)$ by direct differentiation:

$$\dot{z}(s) = \sum_{j=1}^{n} u_{x_j}(\mathbf{x}(s))\dot{x}^j(s) = \sum_{j=1}^{n} p^j(s) F_{p_j}(\mathbf{p}(s), z(s), \mathbf{x}(s)) = D_p F(\mathbf{p}(s), z(s), \mathbf{x}(s)) \cdot \mathbf{p}(s).$$

That is,

$$\boxed{z(s) = D_p F(\mathbf{p}(s), z(s), \mathbf{x}(s)) \cdot \mathbf{p}(s).}$$

2.2.1 Why do we only define boundary data on $\Gamma \subseteq \partial\Omega$?

Suppose a characteristic passes through two different points on the boundary. If we are, for instance, in the case when constant function values propagate along characteristics (e.g. the transport equation), we must be positive that the boundary data matches. That is, if $\mathbf{x}(s)$ is a characteristic and $x^0 \neq \mathbf{x}(s_1)$ with both $x^0, \mathbf{x}(s_1) \in \partial\Omega$, we must ensure that $g(x^0) = g(\mathbf{x}(s_1))$, otherwise we have a conflict. This will be seen in following examples.

2.2.2 PDE solution satisfies ODEs

These equations are extremely useful as they allow us to form a smooth system of solvable ODEs, where all of the second derivative terms have vanished due to our 'smart' selection of $\mathbf{x}(s)$. We now verify that these equations truly do produce solutions for Du and u:

Theorem 2.1 (Structure of characteristics ODEs)**.** *Let $u \in C^2(\Omega)$ solve the nonlinear, first-order partial differential equation* (1.1) *in* Ω. *Assume* $\mathbf{x}(\cdot)$ *satisfies* (2.2)*(c). If* $\mathbf{p}(\cdot) = Du(\mathbf{x}(\cdot))$ *and* $z(\cdot) = u(\mathbf{x}(\cdot))$*, then* $\mathbf{p}(\cdot)$ *and* $z(\cdot)$ *solve the ODEs* (2.2)*(a) and* (2.2)*(b), respectively.*

Proof. We have $p^i(s) = u_{x_i}(\mathbf{x}(s))$. Differentiating with respect to s yields

$$\dot{p^i}(s) = \sum_{j=1}^{n} u_{x_i x_j}(\mathbf{x}(s))\dot{x}^j(s) = \sum_{j=1}^{n} u_{x_i x_j} F_p(\mathbf{p}(s), z(s), \mathbf{x}(s)).$$

By the identity we derived (2.5) we now have

$$\dot{p^i}(s) = -F_{x_i}(\mathbf{p}(s), z(s), \mathbf{x}(s)) - F_z(\mathbf{p}(s), z(s), \mathbf{x}(s)) p^i.$$

We finally write this in vector notation to get

$$\mathbf{p}(s) = -D_x F(\mathbf{p}(s), z(s), \mathbf{x}(s)) - F_z(\mathbf{p}(s), z(s), \mathbf{x}(s))\mathbf{p}(s),$$

which matches exactly with (2.2)(a). We do the same for $z(s) = u(\mathbf{x}(s))$:

$$\dot{z}(s) = \sum_{j=1}^{n} u_{x_j}(\mathbf{x}(s))\dot{x}^j(s) = \sum_{j=1}^{n} p^j(s)F_{p_j}(\mathbf{p}(s), z(s), \mathbf{x}(s)) = D_pF(\mathbf{p}(s), z(s), \mathbf{x}(s)) \cdot \mathbf{p}(s).$$

This completes the proof. □

2.3 Boundary conditions

Boundary conditions are of interest since we aim to solve the ODE system to recover information about u. In the following subsection we illustrate that we may WLOG assume that our boundary is 'flat' near our initial point $x^0 \in \Gamma \subseteq \partial\Omega$.

2.3.1 Straightening the boundary

Fixing $x^0 \in \Gamma$, we "find" (Evans) smooth mappings $\mathbf{\Phi}, \mathbf{\Psi} : \mathbb{R}^n \to \mathbb{R}^n$ with $\mathbf{\Psi} = \mathbf{\Phi}^{-1}$ and $\mathbf{\Phi}$ 'straightens out' $\partial\Omega$ near x^0.

We are now seeking a solution $u : \Omega \to \mathbb{R}$ to (1.1). Write $\Upsilon = \mathbf{\Phi}(\Omega)$, setting

$$v(\mathbf{y}) \triangleq u(\mathbf{\Psi}(\mathbf{y})) \quad \forall\ \mathbf{y} \in \Upsilon,$$

and then

$$u(x) = v(\mathbf{\Phi}(\mathbf{x})) \quad \forall\ \mathbf{x} \in \Omega.$$

We now illustrate that the PDE, call it G, associated with v is of the "same form" as the PDE for u. Calculating the derivatives of u provides this information:

$$u_{x_i}(\mathbf{x}) = \sum_{j=1}^{n} v_{y_j}(\mathbf{\Phi}(\mathbf{x}))\Phi^j_{x_i}(\mathbf{x}) \quad \forall\ i = 1, \ldots, n.$$

Writing this in vector-matrix form, we see that

$$Du(\mathbf{x}) = Dv(\mathbf{y})D\mathbf{\Phi}(\mathbf{x}),$$

where $\mathbf{y} = \mathbf{\Phi}(\mathbf{x})$ and $\mathbf{x} = \mathbf{\Psi}(\mathbf{y})$. Plugging this into (1.1) then provides us that

$$G(Dv(\mathbf{y}), v(\mathbf{y}), \mathbf{y}) \triangleq F(Dv(\mathbf{y})D\mathbf{\Phi}(\mathbf{\Psi}(\mathbf{y})), v(\mathbf{y}), \mathbf{\Psi}(\mathbf{y})) = F(Du(\mathbf{x}), u(\mathbf{x}), \mathbf{x}) = 0.$$

Defining $v(\mathbf{y}) = h(\mathbf{y}) \triangleq g(\mathbf{\Psi}(\mathbf{y}))$ on $\Theta = \mathbf{\Phi}(\Gamma)$, we therefore have

$$\begin{cases} G(Dv, v, \mathbf{y}) = 0 & \text{in } \Upsilon \\ v = h & \text{on } \Theta. \end{cases}$$

2.3.2 Compatibility conditions

As shown in the previous section we may assume that for a given fixed $x^0 \in \Gamma$, that it is flat on the boundary – that is, it lies in the plane $\{x_n = 0\}$.

How do we define our initial conditions correctly so that the ODE system truly produces solutions along characteristics? Carefully. We are given

$$\mathbf{p}(0) = p^0, \quad z(0) = z^0, \quad \mathbf{x}(0) = x^0,$$

but only currently have defined x^0. Since $z(s) = u(\mathbf{x}(s))$, it seems natural that $z(0) = u(x^0) = g(x^0)$. That is,

$$z^0 = g(x^0). \tag{2.6}$$

The more complicated question we aim to answer is how to define p^0. We know that $u(\mathbf{x}) = g(\mathbf{x})$ on $\Gamma \subseteq \partial\Omega$. Since we assume that x^0 lived in the plane with $\{x_n = 0\}$, we then have that

$$u_{x_i}(x^0) = g_{x_i}(x^0) \quad \forall\ i = 1, \ldots, n-1,$$

and since we also want the PDE to hold we finally have our compatibility conditions on p^0:

$$\begin{cases} p^i_0 = g_{x_i}(x^0) & \forall\ i = 1, \ldots, n-1 \\ F(p^0, z^0, x^0) = 0. \end{cases} \tag{2.7}$$

This gives n equations with n unknowns. Together (2.6) and (2.7) gives us what we call our *compatibility conditions*.

2.3.3 Non-characteristic boundary data

The objective of this section is to provide some groundwork for when we may solve a first-order PDE using the method of characteristics, locally. We start with a fixed point $x^0 \in \Gamma$. From (2.7) we have a way to define p^0 given x^0. We now pose conditions to solve for this initial value problem locally for all $\mathbf{y} \in \Gamma$ near x^0. That is, we seek a function $\mathbf{q}(\cdot)$ such that $\mathbf{q}(x^0) = p^0$ and $\mathbf{q}(\mathbf{y})$ satisfies the comp ability conditions.

That is, we fix $\mathbf{y} \in \Gamma$ near x^0 and we hope to solve the characteristic ODEs (2.2) subject to the initial conditions

$$\mathbf{p}(0) = \mathbf{q}(\mathbf{y}), \quad z(0) = g(\mathbf{y}), \quad \mathbf{x}(0) = \mathbf{y}.$$

We then must figure out how to select $\mathbf{q}(\cdot)$. It must satisfy $\mathbf{q}(x^0) = p^0$ as well as the compatibility conditions (2.7). To summarize, we must have

$$\mathbf{q}(x^0) = p^0, \qquad \begin{cases} q^i(\mathbf{y}) = g_{x_i}(\mathbf{y}) \quad (i = 1, \dots, n-1) \\ F(\mathbf{q}(\mathbf{y}), g(\mathbf{y}), y) = 0 \end{cases} \tag{2.8}$$

for all $\mathbf{y} \in \Gamma$ close to x^0.

Lemma 2.2 (Noncharacteristic boundary conditions)**.** *There exists a unique solution* $\mathbf{q}(\cdot)$ *of* (2.8) *for all* $\mathbf{y} \in \Gamma$ sufficiently close to x^0, *provided*

$$F_{p_n}(p^0, z^0, x^0) \neq 0. \tag{2.9}$$

Remark 2.3. The triple (p^0, z^0, x^0) is said to be *noncharacteristic* if (2.9) holds.

Proof of Lemma 2.2. We are automatically provided the first $n-1$ components of $\mathbf{q}(\mathbf{y})$ by (2.8), so the goal is to solve for $q^n(\mathbf{y})$. Remembering that $x_0^n = 0$, since we assumed flatness of the boundary, it makes sense that $F_{p_n}(p^0, z^0, x^0) \neq 0$ should be an appropriate condition to 'push' solutions out of the $\{x_n = 0\}$ plane.

We aim to solve for $q^n(\cdot)$ about (p^0, z^0, x^0). To do do we must consider $\frac{\partial F}{\partial p_n}(p^0, z^0, x^0)$ and show it is nonzero. This is given by the condition in (2.9), and hence by the implicit function theorem we may then solve for $\mathbf{q}(\cdot)$ locally about x^0. ☐

2.4 Local existence

Theorem 2.4 (Local invertibility)**.** *Assume we have the noncharacteristic condition* $F_{p_n}(p^0, z^0, x^0) \neq 0$. *Then* $\exists$ *interval* I *such that* $0 \in I \subseteq \mathbb{R}$, *and a neighborhood* W *of* x^0 *in* $\Gamma \subseteq \mathbb{R}^{n-1}$, *and a neighborhood* V *of* x^0 *in* $\mathbb{R}^n$, *and a neighborhood* V *of* x^0 *in* $\mathbb{R}^n$, *such that for each* $x \in V$ *there exist unique* $s \in I, y \in W$ *such that*

$$x = \mathbf{x}(y, s).$$

The mappings $x \mapsto s, y$ *are* C^2.

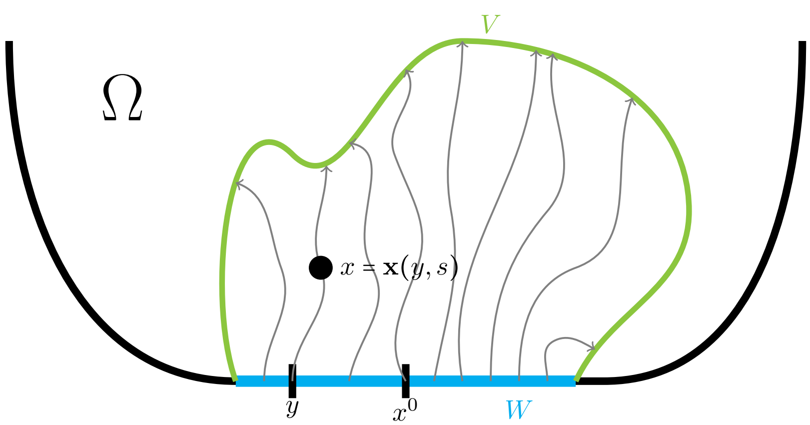

Figure 6: This diagram shows the local existence given by the inverse function theorem. We have a neighborhood W of x^0 on $\Gamma \subseteq \partial\Omega$. We also then have a neighborhood V of x^0 living in $\Omega \subseteq \mathbb{R}^n$. This solvability for the solution $u \in C^2$ will be given on V, and as such we have that the characteristics will not intersect in V.

Proof. With our notation we have $\mathbf{x}(x^0,0) = x^0$. We will apply the Inverse Function Theorem which gives the result that we may solve for $x = \mathbf{x}(y,s)$.

First use that
$$\mathbf{x}(y,0) = (y,0) \quad (y \in \Gamma),$$
which gives
$$x^j_{y_i}(x^0,0) = \begin{cases} \delta_{ij} & (j = 1,\dots,n-1) \\ 0 & (j = n). \end{cases}$$
We will use this to produce $D_y\mathbf{x}$, an $(n-1)\times(n-1)$ matrix. In the s-component for each of these, we have that derivative is simply zero. We also have, directly from our characteristic equations (2.2)(c), that
$$x^j_s(x^0,0) = F_{p_j}(p^0,z^0,x^0).$$
With all of this in hand we may write the Jacobian of $\mathbf{x}$ to prove the solvability of $\mathbf{x}$ with regard to the other variables of F. Since $\mathbf{x}$ depends on y and s, we have computed it's derivative with respected to those, and now collected them into the full Jacobian:

$$D\mathbf{x}(x^0,0) = \begin{pmatrix} 1 & & & F_{p_1}(p^0,z^0,x^0) \\ & \ddots & & \vdots \\ & & 1 & \vdots \\ 0 & 0 & 0 & F_{p_n}(p^0,z^0,x^0) \end{pmatrix}_{n\times n}.$$

The determinant of the above is obviously nonzero so long as $F_{p_n}(p^0,z^0,x^0) \neq 0$, our characteristic condition. The inverse function theorem now gives the result. ☐

Remark 2.5. Assuming we have some $y \in \Gamma$ close to x^0, we will use the following notation to show the dependence of solutions to the characteristic ODE not only on s, but on y as well. That is, not only on the time elapsed along the characteristic from the boundary, but also where that characteristic started from (which makes sense).

$$\begin{cases} \mathbf{p}(s) &=& \mathbf{p}(y,s) \\ z(s) &=& z(y,s) \\ \mathbf{x}(s) &=& \mathbf{x}(y,s). \end{cases}$$

Sometimes it may also be written that $x^0 \in \mathbb{R}^{n-1}$ due to having assumed a straightened boundary (i.e. $\{x_n = 0\}$).

Remark 2.6. The above lemma provides us that for each $x \in V$ that we may uniquely solve

$$\begin{cases} x = \mathbf{x}(y,s) \\ \text{for } y = \mathbf{y}(x), \;\; s = s(x). \end{cases} \tag{2.10}$$

Now we may solve

$$\begin{cases} u(x) &\triangleq& z(\mathbf{y}(x), s(x)) \\ \mathbf{p}(x) &\triangleq& \mathbf{p}(\mathbf{y}(x), s(x)), \end{cases} \tag{2.11}$$

for $x \in V$ and s, y as in (2.10).

Theorem 2.7 (Local Existence Theorem)**.** *The function u defined above is C^2 and solves the PDE*
$$F(Du(\mathbf{x}), u(\mathbf{x}), \mathbf{x}) = 0 \quad (\mathbf{x} \in V),$$
with the boundary condition
$$u(\mathbf{x}) = g(\mathbf{x}) \quad (x \in \Gamma \cap V).$$

Proof. We attack the proof as follows:

(1) First we set up the problem: Fix $y \in \Gamma$ near x_0. Solve the characteristic ODEs (2.2) and write $\mathbf{p}(s) = \mathbf{p}(y,s)$, $z(s) = z(y,s)$, and $\mathbf{x}(s) = \mathbf{x}(y,s)$ according to the previous work.

(2) Now we want to show that the PDE (1.1) is satisfied along the characteristics. That is, for $y \in \Gamma$ 'close enough' to $x^0 \in \Gamma$, then we want to show that

$$f(y,s) \triangleq F(\mathbf{p}(y,s), z(y,s), \mathbf{x}(y,s)) = 0 \quad (s \in I). \tag{2.12}$$

We will proceed to show $f(y,0) = 0$ and $f_s(y,s) = 0$, which directly implies $f(y,s) = 0$. First,

$$f(y,0) = F(\mathbf{p}(y,0), z(y,0), \mathbf{x}(y,0)) = F(\mathbf{q}(y), u(y), y) = 0$$

by our compatibility condition (2.7). Also,

$$\begin{aligned} f_s(y,s) &= \sum_{j=1}^{n} F_{p_j}\dot{p}^j + F_z\dot{z} + \sum_{j=1}^{n} F_{x_j}\dot{s}^j \\ \text{plug in (2.2)} \rightarrow \quad &= \sum_{j=1}^{n} F_{p_j}\left[-F_{x_j} - F_z p^j\right] + F_z\left(\sum_{j=1}^{n} F_{p_j}p^j\right) + \sum_{j=1}^{n} F_{x_j}F_{p_j} \\ &= 0. \end{aligned}$$

This shows, as mentioned, that $f(y,s) = 0$ as desired.

(3) Now by Lemma 2 and the resulting equations (2.10) and (2.11) along with the result (2.12) we finally have

$$F(\mathbf{p}(x), u(x), x) = 0 \quad \forall x \in V.$$

Recall that V is the neighborhood about x_0 for which we were able to solve (2.10) and (2.11). The above looks like what we desire, but we must formally prove that $\mathbf{p}(x)$ solved via the Local Invertibility Theorem is in fact $Du(x)$. That is, we must show that

$$\mathbf{p}(x) = Du(x) \quad (x \in V) \tag{2.13}$$

To do so is slightly computationally complicated so we take a momentary detour.

Precomputations for final result. Let $s \in I$ and $y \in W$. We have directly from the characteristic ODEs that (don't get confused here, $\dot{z} = z_s$ and $\dot{x}^j = x_s^j$ since we now have two variables (y,s) technically)

$$z_s(y,s) = \sum_{j=1}^{n} p^j(y,s)x_s^j(y,s). \tag{2.14}$$

We also will establish that

$$z_{y_i}(y,s) = \sum_{j=1}^{n} p^j(y,s)x_{y_i}^j(y,s) \quad (i = 1,\dots,n-1). \tag{2.15}$$

To do so define

$$r^i(s) \triangleq z_{y_i}(y,s) - \sum_{j=1}^{n} p^j(y,s)x_{y_i}^j(y,s)$$

for each $i = 1,\dots,n-1$. Taking the derivative of $r^i(s)$ with respect to s gives

$$\dot{r}^i(s) = z_{y_i s} - \sum_{j=1}^{n}\left(p_s^j x_{y_i}^j + p^j x_{y_i s}^j\right) \tag{2.16}$$

Taking (2.14) and differentiating with respect to y_i gives

$$z_{sy_i} = \sum_{j=1}^{n}\left(p_{y_i}^j x_s^j + p^j x_{sy_i}^j\right).$$

Placing this into (2.16) gives

$$\dot{r}^i(s) = \sum_{j=1}^{n}\left(p_{y_i}^j x_s^j - p_s^j x_{y_i}^j\right) = \sum_{j=1}^{n}\left[p_{y_i}^j F_{p_j} - (-F_{x_j} - F_z p^j)x_{y_i}^j\right]. \tag{2.17}$$

We now differentiate (2.12) with respect to y_i to get

$$\sum_{j=1}^{n} F_{p_j}p_{y_i}^j + F_z z_{y_i} + \sum_{j=1}^{n} F_{x_j}x_{y_i}^j = 0.$$

Plugging this into (2.17) finally gives us that

$$\dot{r}^i(s) = F_z\left(\sum_{j=1}^{n} p^j x_{y_i}^j - z_{y_i}\right) = -F_z r^i(s).$$

Since $r^i(s)$ solves the above ODE with initial condition $r^i(0) = g_{x_i}(y) - q^i(y) = 0$, this clearly gives $r^i(s) \equiv 0$ for all $s \in I$ and $i = 1,\dots,n-1$. This proves (2.15).

Now we may proceed. To keep notation simple, let $j = 1, \dots, n$, and using (2.11) we have

$$\begin{aligned} u_{x_j} &= z_s s_{x_j} + \sum_{i=1}^{n-1} z_{y_i} y^i_{x_j} \\ \text{by previous work} \rightarrow &= \sum_{k=1}^{n} p^k x^k_s s_{x_j} + \sum_{i=1}^{n-1} \sum_{k=1}^{n} p^k x^k_{y_i} y^i_{x_j} \\ &= \sum_{k=1}^{n} p^k \left(x^k_s s_{x_j} + \sum_{i=1}^{n-1} x^k_{y_i} y^i_{x_j} \right) \\ &= \sum_{k=1}^{n} p^k x^k_{x_j} = \sum_{k=1}^{n} p^k \delta_{jk} = p^j. \end{aligned}$$

That is, $\mathbf{p} = Du$.

□

2.5 Examples

2.5.1 Eikonal equation

The Eikonal equation can be written as

$$F(\mathbf{p}, z, \mathbf{x}) = -f(\mathbf{x}) + \frac{1}{2} \sum_{i=1}^{n} p_i^2 = 0,$$

so we may solve it via the method of characteristics. We have $\dot{x}_i = p_i$, $\dot{p}_i = -f_{x_i}$, and $\dot{z} = |\mathbf{p}|^2$ for $i = 1, \dots, n$.

This gives that

$$\mathbf{x}(s) = Du(\mathbf{x}(s)),$$

so we may think of a characteristic emanating from $\Gamma \subseteq \partial\Omega$ to be moving in the "direction of optimal motion," the direction of the gradient at that position. The exact dynamics are controlled by our choice of f. For polynomial f of degree 2 or less, the characteristic ODE system is fully linear (except for $\dot{z}$, but that isn't involved in the coupling between $\dot{x}$ and $\dot{\mathbf{p}}$) and we may solve it using standard ODE techniques. Once f becomes nonlinear, the problem obviously becomes harder.

2.5.2 Examples: Deriving PDEs from conditions on characteristics

Let $U \subseteq \mathbb{R}^n$ be open with ∂U smooth. Suppose we have $F : \mathbb{R}^n \times \mathbb{R}\bar{U}$, $F(\mathbf{p}, z, \mathbf{x})$ is smooth and $g : \Gamma \to \mathbb{R}$ is also smooth on $\Gamma \subseteq \partial\Omega$.

From the Local Existence Theorem of the Method of Characteristics, we have that $\exists V \subseteq U$, a neighborhood of Γ so that the method of characteristics has produced a C^2 solution, $u(\mathbf{x})$ satisfying (1.1).

For these problems we consider that our vector field

$$\mathbf{v} : \mathbb{R}^n \to \mathbb{R}^n, \ \mathbf{v} = \mathbf{v}(\mathbf{x})$$

is smooth and transversal to Γ.

Example 2.8 (Time-optimal problem)**.** We seek solutions to (1.1) such that the time traveled along a characteristic gives the value of the function. That is, $u(\mathbf{x}(s))$ is the time traveled along $\mathbf{x}(s)$ since entering U. This gives us obvious boundary data: $g \equiv 0$ on $\Gamma = \partial U$.

First we suppose that our characteristics follow according to the vector field $\mathbf{v}(\mathbf{x})$. This places the condition on F that

$$\mathbf{v}(\mathbf{x}) = D_p(Du(\mathbf{x}), u(\mathbf{x}), \mathbf{x}) \tag{2.18}$$

immediately (it places the condition along each characteristic, so assuming it for the whole space is okay).

To start, we write

$$z(s) = u(\mathbf{x}(s)) = s \implies \dot{z}(s) = 1,$$

and since $\dot{z}(s) = \dot{\mathbf{x}} \cdot \mathbf{p} = \mathbf{v} \cdot \mathbf{p}$, we therefore arrive at

$$\mathbf{v} \cdot \mathbf{p} = 1, \tag{2.19}$$

as an obvious condition our PDE must satisfy. Remember that $\mathbf{v}(\mathbf{x}) = D_p F(Du(\mathbf{x}), u(\mathbf{x}), \mathbf{x})$ is something F must satisfy. After careful inspection notice that $F(Du(\mathbf{x}), u(\mathbf{x}), \mathbf{x}) = \mathbf{v}(\mathbf{x}) \cdot Du(\mathbf{x}) - 1$ is an obvious candidate satisfying both (2.18) and (2.19) so we have derived *a* PDE.

Note that if we select $\mathbf{v}(\mathbf{x}) = Du(\mathbf{x})$ we rederive the Eikonal equation.

Homework problem 1. Using the same approach, derive a PDE such that $u(\mathbf{x}(s))$ is the total distance traveled from $\mathbf{x}(0) \in \Gamma$ along each characteristic $\mathbf{x}(s)$.

Homework problem 2. Can you do the same thing to derive a PDE that satisfies $u(\mathbf{x}(s)) = g(\mathbf{x}(0))$ on each characteristic?

2.5.3 Examples of how characteristics flow

The following examples come directly from Evans. We provide some discussion about certain simple scenarios in which we could construct PDEs to satisfy the given properties.

Flow to an attracting point. If we consider a situation in which the characteristics flow to an attracting point, we may think of our vector field $\dot{\mathbf{x}}(s) = \mathbf{v}(\mathbf{x}(s)) = D_p F(Du(\mathbf{x}(s)), u(\mathbf{x}(s)), \mathbf{x})$ as having a single critical point and on the entire vector field we have $\mathbf{v} \cdot \nu < 0$ where ν is the *outward pointing normal*.

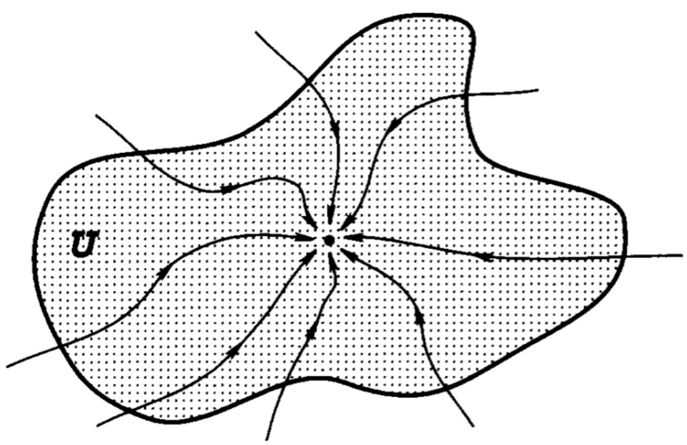

Figure 7: Flow to an attracting point. Figure copied directly from Evans.

It is a fair question to ask if solutions constructed using the method of characteristics will be smooth, or even continuous. In this case the answer is no, that they will not be smooth, and unless U itself is a 'very' nice shape (e.g. a circle) we will likely even have discontinuities in different values leading up to the critical point.

Further, consider the preceding example where the solution u returns the amount of time traveled along each characteristic. The characteristics themselves never reach the critical point (standard ODE theory), and hence it takes infinite time. Thus if $\mathbf{w}$ is the critical point, then $u(\mathbf{x}) \to \infty$ as $\mathbf{x} \to \mathbf{w}$.

On the other hand, if we were to instead consider the PDE giving the amount of distance traveled, this would produce finite, but likely discontinuous solutions (certainly not smooth).

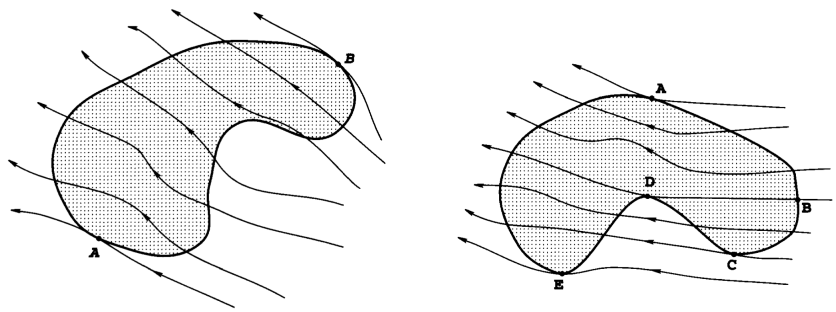

Figure 8: Flow across a domain (left) and a characteristic point given by D (right). Figures copied directly from Evans.

2.6 Characteristics for Hamilton-Jacobi equations

The characteristics for the Hamilton-Jacobi equations provide a coupling in the characteristic equations between the $\dot{\mathbf{x}}$ and $\dot{\mathbf{p}}$, with no dependence on $z = u$. The general Hamilton-Jacobi PDE is given by

$$G(Du, u_t, u, x, t) = u_t + H(Du, x) = 0,$$

with $Du = D_x u$. If we let $y = (x, t)$, $q = (p, p_{n+1})$, this gives

$$G(q, z, y) = p_{n+1} + H(p, x)$$

for the PDE we'll apply the Method of Characteristics to. We result in

$$\begin{cases} \dot{x}^i(s) &= H_{p_i}(\mathbf{p}(s), \mathbf{x}(s)) \quad (i = 1, \dots, n) \\ \dot{x}^{n+!}(s) &= 1. \end{cases}$$

Since H does not depend on u, we get

$$\left\{ \begin{array}{rcl} \dot{p}^i(s) &=& -H_{x_i}(\mathbf{p}(s), \mathbf{x}(s)) \quad (i = 1, \dots, n) \\ \dot{p}^{n+1}(s) &=& 0, \end{array} \right\}$$

and

$$\dot{z}(s) = D_p H(\mathbf{p}(s), \mathbf{x}(s)) \cdot \mathbf{p}(s) + 1 \cdot p^{n+1} = D_p H(\mathbf{p}(s), \mathbf{x}(s)) - H(\mathbf{p}(s), \mathbf{x}(s)).$$

This provides us with the characteristic equations for the Hamilton-Jacobi equations, given by

$$\begin{cases} \dot{\mathbf{p}}(s) &= -H_{x_i}(\mathbf{p}(s), \mathbf{x}(s)) \\ \dot{z}(s) &= D_p H(\mathbf{p}(s), \mathbf{x}(s)) - H(\mathbf{p}(s), \mathbf{x}(s)) \\ \dot{\mathbf{x}}(s) &= H_{p_i}(\mathbf{p}(s), \mathbf{x}(s)). \end{cases}$$

The first and third of these give *Hamilton's equations*. These equations give a nice coupling between $\mathbf{p}$ and $\mathbf{x}$, with no z terms. z depends on $\mathbf{p}$ and $\mathbf{x}$ exclusively, so once those are solved we have solved the problem.

2.7 One true MoC example

Just to show how we may apply the Method of Characteristics to a real and given problem, we consider Problem 5(c), a quasi-linear PDE:

$$uu_{x_1} + u_{x_2} = 1 \qquad \text{with } u(x_1, x_1) = \frac{1}{2}x_1.$$

To solve this write the PDE in standard form:

$$F(p, z, x) = zp_1 + p_2 - 1 = 0.$$

Using the method of characteristics we determine

$$\dot{z} = \dot{\mathbf{x}} \cdot \mathbf{p} = (z, 1) \cdot (p_1, p_2) = 1, \qquad \begin{cases} \dot{p}_1 &= -0 - p_1 \cdot p_1 \\ \dot{p}_2 &= -0 - p_1 p_2, \end{cases} \qquad \begin{cases} \dot{x}_1 &= z \\ \dot{x}_2 &= 1. \end{cases}$$

Selecting $x_0 = (t,t)$, we then have $g(x_0) = \frac{1}{2}t$. for $t \in \mathbb{R}$. We may solve that $z(s) = s + z_0 = s + g(x_0) = s + \frac{1}{2}t$, so

$$\dot{x}_1 = s + \frac{1}{2}t \implies x_1(s) = \frac{1}{2}s^2 + \frac{1}{2}st + x_1^0 = \frac{1}{2}s^2 + \frac{1}{2}st + t,$$

and $x_2(s) = s + x_2^0 = s + t$. Now, we have

$$\left(\frac{1}{2}s^2 + \frac{1}{2}st + t, s + t\right) = (x_1(s), x_2(s)),$$

and solving for s, t we arrive at

$$(s,t) = \left(\frac{2x_2 - 2x_1}{2 - x_2}, \frac{2x_1 - x_2^2}{2 - x_2}\right).$$

This finally allows us to get rid of the s to produce

$$u(x_1, x_2) = s + \frac{1}{2}t = \frac{-x_1 + 2x_2 - x_2^2/2}{2 - x_2}.$$

3 Calculus of Variations vs. Optimal Control

This section aims to provide a (more or less) definition of what *calculus of variations* and *optimal control* exactly are and how they are related. Calculus of variations seeks to minimize a functional over a set of trajectories in the state space starting at one location and ending at another, in finite time.

Optimal control problems are a subset of calculus of variations problems, yet we have a dynamical “control” placed on the derivatives of all elements in the admissible class $\mathcal{A}$.

3.1 Calculus of Variations

We let $L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a smooth function named the *Lagrangian*. Fix any two points $x, y \in \mathbb{R}^n$ with $x \neq y$ and select $t > 0$. We define the *action functional*

$$I[\mathbf{w}(\cdot)] \triangleq \int_0^t L\left[\dot{\mathbf{w}}(s), \mathbf{w}(s)\right] ds,$$

where the functions $\mathbf{w}$ belong to the *admissible class*

$$\mathcal{A} \triangleq \left\{\mathbf{w}(\cdot) \in C^2\left([0,t];\mathbb{R}^n\right) \mid \mathbf{w}(0) = y,\ \mathbf{w}(t) = x\right\}.$$

A problem in the calculus of variations is to minimize this action functional over $\mathcal{A}$. I.e. we seek $\mathbf{x}(\cdot)$ so that

$$I\left[\mathbf{x}(\cdot)\right] = \min_{\mathbf{w}(\cdot)\in\mathcal{A}} I[\mathbf{w}(\cdot)].$$

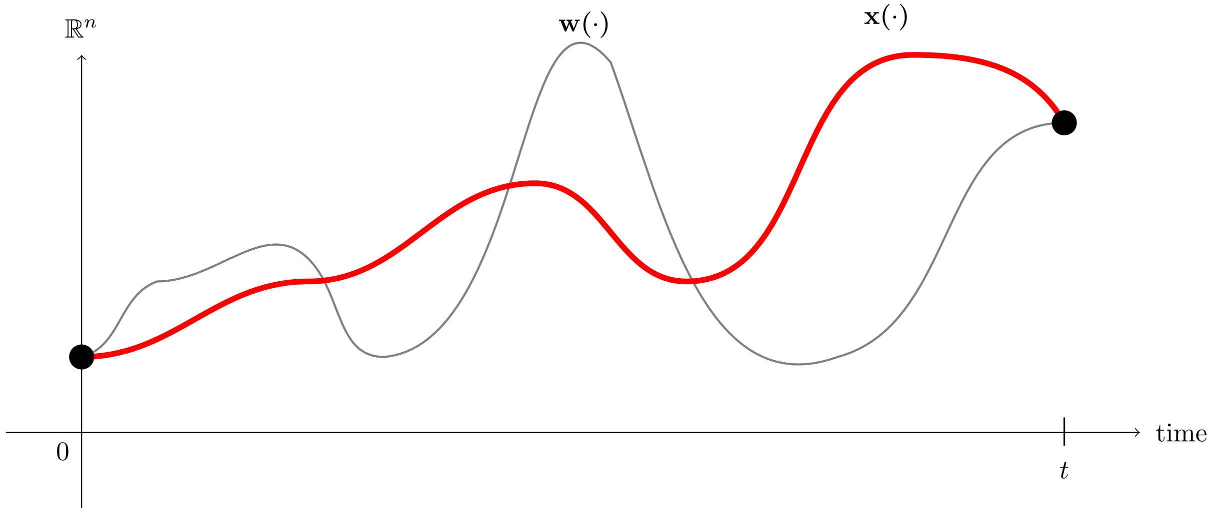

Figure 9: A problem in calculus of variations is to minimize a functional $I[\mathbf{w}(\cdot)] = \int_0^t L(\dot{\mathbf{w}}(s), \mathbf{w}(s))ds$ over all possible candidates $\mathbf{w}(\cdot) : [0,t] \to \mathbb{R}^n$ satisfying that $\mathbf{w} \in C^2\left([0,t]\right)$ and $\mathbf{w}(0) = \mathbf{x}$ and $\mathbf{w}(t) = \mathbf{y}$. Here $\mathbf{x}(\cdot)$ is the minimizer.

3.2 Optimal Control

In optimal control we now prescribe our set $\mathcal{A}$ of admissible candidates to satisfy an additional dynamic control. Any $\mathbf{w}(\cdot) \in \mathcal{A}$ must now also satisfy the system

$$\begin{cases} \mathbf{w}'(s) = \mathbf{f}(\mathbf{w}(s), \mathbf{a}(s)) & \text{in } \Omega \subseteq \mathbb{R}^n \\ \mathbf{w}(0) = \mathbf{x} & \text{for some fixed } \mathbf{x} \in \Omega. \end{cases}$$

Our goal is still the same. We are starting at an initial point $\mathbf{x}$ at time zero and traveling to some other point, $\mathbf{w}(t) = \mathbf{y}$.

Rather than dwell on the details, let’s see this in action:

3.2.1 Deriving the Eikonal equation (Optimal Control)

Deriving the Eikonal equation from an optimal control problem is a simple yet elegant result.

Define the **trajectory of motion** to be given by $\mathbf{y}(s)$ for $s \in [0, T]$ for some $T > 0$, where

$$\begin{cases} \mathbf{y}'(s) = f(\mathbf{y}(s))\mathbf{a}(s) & \text{in } \Omega \\ \mathbf{y}(0) = \mathbf{x} & \text{for some fixed } \mathbf{x} \in \Omega. \end{cases}$$

That is, $\mathbf{y}(s)$ is a trajectory living in Ω, where $\mathbf{a}(\cdot)$ is our 'control' with $\mathbf{a} : \mathbb{R} \to A \subseteq \mathbb{R}^n$, with A compact and contained in the unit sphere. That is, $|\mathbf{a}(\cdot)| = 1$ for all $\mathbf{a}(\cdot) \in \mathcal{A}$.

We consider that as the trajectory $\mathbf{y}(s)$ moves it accrues a cost presented by the *cost function* $K(\mathbf{x}, \mathbf{a}(\cdot))$, where $K : \Omega \times \mathbb{R} \to \mathbb{R}$. Once $\mathbf{y}(s) \in \partial\Omega$ for the first time, we incur an *exit-cost* given by the function $q : \partial\Omega \to \mathbb{R}$.

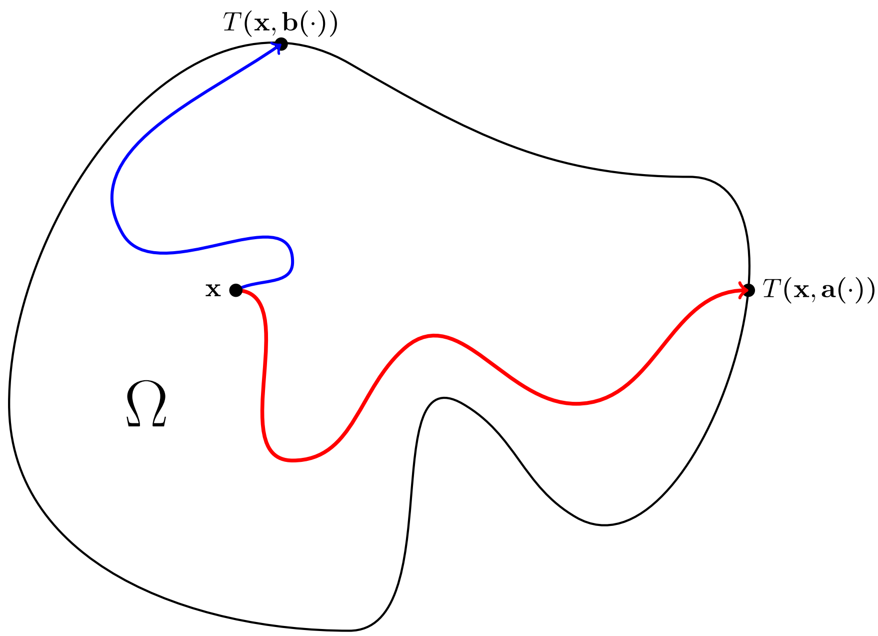

Figure 10: The curve in red represents $\mathbf{y}(t)$ given by the *trajectory of motion*. The speed at any point in Ω is controlled by f, and direction by $\mathbf{a}(\cdot)$. We will later see that we will choose $\mathbf{a}(\mathbf{x}) = -\frac{\nabla u(\mathbf{x})}{|\nabla u(\mathbf{x})|}$ for all $\mathbf{x} \in \Omega$, i.e. the direction of maximal decrease. We may think of the curve in red as the curve with this select of control. Should we pick another control $\mathbf{b}(\cdot)$ we would get a different curve, e.g. the blue one.

Define the **final exit time starting at x** to be given by

$$T(\mathbf{x}, \mathbf{a}(\cdot)) = \min \{ t \geq 0 \mid y(t) \in \partial\Omega \}.$$

Now, the **total cost to exit starting at x** is then given by

$$\mathcal{J}(\mathbf{x}, \mathbf{a}(\cdot)) = \int_0^{T(\mathbf{x}, \mathbf{a}(\cdot))} K(\mathbf{y}(s), \mathbf{a}(s)) ds + q(\mathbf{y}(T(\mathbf{x}, \mathbf{a}(\cdot)))).$$

Finally, we choose to infemize over all possible controls $\mathbf{a}(\cdot)$ in order to produce the **value function**, which returns the minimal exit cost starting at $\mathbf{x}$:

$$v(\mathbf{x}) = \inf_{|\mathbf{a}(\cdot)|=1} \mathcal{J}(\mathbf{a}, \mathbf{a}(\cdot)).$$

This looks like an intimidating beast to tackle right off the bat. We appeal to *Bellman's optimality principle*, which gives us the following. Let $\tau > 0$ be a 'small' number. By the principle if we travel along $\mathbf{y}(s)$ for τ seconds, we accrue cost according to K, and we wish to add on the minimal-cost-to-exit at $\mathbf{y}(\tau)$, given by $b(\mathbf{y}(\tau))$. To summarize,

$$v(\mathbf{x}) = \inf_{|\mathbf{a}(\cdot)|=1} \left[\int_0^\tau K(\mathbf{y}(s), \mathbf{a}(s))ds + v(\mathbf{y}(\tau)) \right].$$

Since $\tau > 0$ is small, the integral term may be approximated using geometry:

$$\int_0^\tau K(\mathbf{y}(s), \mathbf{a}(s))ds \;\approx\; \tau K(\mathbf{y}(0), \mathbf{a}(0)) + O(\tau^2) \;=\; \tau K(\mathbf{x}, \mathbf{a}(0)) + O(\tau^2).$$

The Taylor expansion for $v(\mathbf{y}(\tau))$ about $\tau = 0$ should produce a 'good' approximation. This is given by

$$\begin{aligned}
v(\mathbf{y}(\tau)) &= v(\mathbf{y}(0)) + \frac{\frac{d}{ds}\left[v(\mathbf{y}(s))\right]|_{s=0}}{1!}\tau^1 + O(\tau^2) \\
&= v(\mathbf{y}(0)) + \tau \nabla v(\mathbf{y}(0)) \cdot \mathbf{a}(0) f(\mathbf{y}(0)) + O(\tau^2) \\
&= v(\mathbf{x}) + \tau f(\mathbf{x}) \nabla v(\mathbf{x}) \cdot \mathbf{a}(0) + O(\tau^2).
\end{aligned}$$

Plugging all of this into the equation we now have

$$\begin{aligned}
v(\mathbf{x}) &= \inf_{|\mathbf{a}(\cdot)|=1} \left[\tau K(\mathbf{x}, \mathbf{a}(0)) + v(\mathbf{x}) + \tau f(\mathbf{x}) \nabla v(\mathbf{x}) \cdot \mathbf{a}(0) + O(\tau^2) \right] \\
&= v(\mathbf{a}) + \tau \inf_{|\mathbf{a}(\cdot)|=1} \left[K(\mathbf{x}, \mathbf{a}(0)) + f(\mathbf{x}) \nabla v(\mathbf{x}) \cdot \mathbf{a}(0) + O(\tau) \right].
\end{aligned}$$

Canceling $v(\mathbf{x})$ on each side and dividing through by τ, we are left with (inf = min since A is compact)

$$\min_{|\mathbf{a}(\cdot)|=1} \left[K(\mathbf{x}, \mathbf{a}(0)) + f(\mathbf{x}) \nabla v(\mathbf{x}) \cdot \mathbf{a}(0) \right] + O(\tau).$$

We may assume that $K = K(\mathbf{x})$ (i.e. it only depends on physical location) in order to deduce that selecting $\mathbf{a}^*(0) = -\frac{\nabla v(\mathbf{x})}{|\nabla v(\mathbf{x})|}$ is the minimizer. Taking $\tau \downarrow 0$ we are finally left with

$$K(\mathbf{x}) - |\nabla v(\mathbf{x})| f(\mathbf{x}) = 0 \implies \boxed{|\nabla v(\mathbf{x})| = K(\mathbf{x})/f(\mathbf{x}).}$$

Taking $K \equiv 1$ is a standard procedure.

4 References

[A] Lectures on Partial Differential Equations, by Vladimir Arnold

[E] Partial Differential Equations, by Lawrence Evans (second edition)

[K] Introductory Functional Analysis with Applications, by Kreyszig

[S] 300 Years of Optimal Control: From The Brachystochrone to the Maximum Principle, by Sussmann and Willems.