Optimal control for conservation laws in presence of shocks *

Carlos Castro †

May 19, 2009

Abstract

This notes contain a short introduction to the the study of numerical approximation of control problems for conservation laws in presence of shocks. We are mainly interested in aerodynamic applications where problems are formulated in a very complex framework with a large number of constraints. Here we present and analyze simplified models where some of the main difficulties arise and are easier to understand.

The plan is the following: we first consider the main equations in the study of inviscid compressible flows, which are particular examples of a class of systems know as conservation laws. Next we focus on the main properties of the solutions of conservation laws in the simplest scalar one-dimensional case. In the third part we review some difference schemes to approximate conservation laws. In the fourth section we state some optimal control problems for fluids and the main difficulties to approximate them. In the fifth section we briefly describe the most efficient optimization techniques based on gradient methods. Finally, in the last section we analyze the main difficulties arising when trying to approximate the gradient in problems where the dynamic is described by conservation laws.

*This work is supported by the Grant MTM2008-03541 of the MEC (Spain), the DOMINO Project CIT-370200-2005-10 in the PROFIT program of the MEC (Spain).
†Dep. Matemática e Informática, ETSI Caminos, Canales y Puertos, Univ. Politécnica de Madrid, 28040 Madrid, Spain. carlos.castro@upm.es
## Contents

1 Fluid equations for compressible flows 3  
1.1 Euler equations ................................................. 3  
1.2 Inviscid Burgers equation ................................. 5  

2 Solutions of the scalar conservation laws 6  
2.1 The linear advection equation ............................... 6  
2.2 The inviscid Burgers equation ................................. 7  
2.3 Characteristics ................................................... 7  
2.4 Weak solutions ................................................... 8  

3 Numerical approximation for scalar conservation laws 10  
3.1 Conservative schemes ........................................... 11  
3.2 The Lax-Wendroff theorem ..................................... 12  
3.3 Order of accuracy ................................................ 13  
3.4 Stability ......................................................... 13  
3.5 Some examples .................................................... 14  

4 Some optimal control problems for fluids 15  
4.1 An optimal design problem for Euler equations .............. 15  
4.2 The nozzle problem for stationary flows ..................... 16  
4.3 An optimal control problem for Burgers equation ............ 18  
4.4 General remarks ................................................ 19  
4.5 Existence of minimizers ......................................... 20  
4.6 The discrete minimization problem ............................. 22  

5 Gradient calculus 25  
5.1 Unconstrained optimization .................................... 26  
5.2 Gradient Methods ............................................... 27  
5.3 Newton method .................................................... 28  
5.4 Adjoint based methods to compute the gradient ............. 29  

6 Gradient calculus for conservation laws 30  
6.1 The continuous approach without shocks .................... 32  
6.2 The continuous approach in presence of shocks ............... 35  
6.3 Sensitivity of the functional ................................... 40  
6.4 The discrete approach: Differentiable numerical schemes .... 47  
6.5 The discrete approach: Non-differentiable numerical schemes 50
1 Fluid equations for compressible flows

In this section we present the general Euler equations for inviscid flows and two simplified models: The quasi-onedimensional Euler equation which allows to describe the flow of a perfect gas in a nozzle, and the simpler academic example of the scalar inviscid Burgers equation.

1.1 Euler equations

Let $\Omega \subset \mathbb{R}^3$ be the domain occupied by the fluid and $(0, T)$ a time interval. We define the following:

- $\rho(t, x)$, density field,
- $u(t, x)$, density field,
- $p(t, x)$, pressure field,
- $T(t, x)$, temperature field.
- $E(t, x)$, total energy field.

The continuity equation establishes the mass conservation,

$$\partial_t \rho + \nabla \cdot (\rho u) = 0.$$  \hspace{1cm} (1)

The conservation of momentum provides the following vectorial equation

$$\rho(\partial_t u + u \nabla u) + \nabla p - \nabla \cdot \sigma = f.$$ \hspace{1cm} (2)

which is equivalent to the momentum equation

$$\partial_t (\rho u) + \nabla \cdot (\rho u \otimes u) + \nabla \cdot (pI - \sigma) = f,$$  \hspace{1cm} (3)

where $I$ is the identity $3 \times 3$ matrix and $\sigma$ is the stress tensor. For Newtonian flows $\sigma$ is assumed to be

$$\sigma = \mu (\nabla u + \nabla u^T) + \left(\xi - \frac{2}{3}\mu\right)I \nabla \cdot u,$$  \hspace{1cm} (4)

where $\mu$ and $\xi$ are the first and second viscosities of the fluid. For inviscid flows, $\mu = \xi = 0$ and the momentum equation reads

$$\partial_t (\rho u) + \nabla \cdot (\rho u \otimes u) + \nabla p = f.$$  \hspace{1cm} (5)
Finally, we establish the conservation of energy,

$$\partial_t (\rho E) + \nabla \cdot (\rho u E + p) = \nabla \cdot (\rho u \sigma + \kappa \nabla T) + f \cdot u,$$

(6)

where $\kappa$ is the thermal conductivity. For inviscid flows $\sigma = 0$. If we also assume that changes in the temperature can be neglected we have

$$\partial_t (\rho E) + \nabla \cdot (\rho u H) = f \cdot u,$$

(7)

where $H$ is the stagnation enthalpy,

$$H = E + \frac{p}{\rho}.$$

(8)

The system of equations is closed with the equation of state

$$\frac{p}{\rho} = RT,$$

where $R$ is the ideal gas constant. This last equation can be written as

$$e = \frac{p}{\rho(\gamma - 1)},$$

(9)

where $\gamma$ is a physical constant and $e$ is the internal energy. The energy $E$ and the internal energy $e$ are related by

$$E = \frac{1}{2}u^2 + e.$$

In the two dimensional case without external forces, if we write $u = (u, v)$ and $x = (x, y)$, the Euler system for compressible flows can be written as

$$\partial_t U + \nabla \cdot F = \partial_t U + \partial_x F_x + \partial_y F_y = 0, \text{ in } \Omega$$

(10)

$$U = \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ \rho E \end{pmatrix}, \quad F_x = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ \rho u H \end{pmatrix}, \quad F_y = \begin{pmatrix} \rho v \\ \rho uv \\ \rho v^2 + p \\ \rho v H \end{pmatrix}$$

(11)

where

$$p = (\gamma - 1)\rho \left( E - \frac{1}{2}(u^2 + v^2) \right), \quad H = E + \frac{p}{\rho}.$$  

(12)
Equations of a perfect gas on a duct with variable cross sectional area.

We consider the one-dimensional version of the Euler equations for a flow in a duct of variable cross sectional area $A$ (see figure 1). In this case, the domain is an interval $(0, x_e)$ and the flow is assumed to be one-directional. The velocity is then a scalar function $u$.

$$\partial_t(AU) + \partial_x(AF) = \frac{dA}{dx} P, \quad x \in (0, x_e),$$  \hspace{1cm} (13)

where,

$$F = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho uH \end{pmatrix}, \quad P = \begin{pmatrix} 0 \\ p \\ 0 \end{pmatrix}, \quad U = \begin{pmatrix} \rho \\ \rho u \\ \rho E \end{pmatrix},$$  \hspace{1cm} (14)

$$H = e + \frac{p}{\rho} = \frac{\gamma - 1}{\gamma} p + \frac{1}{2} u^2.$$  \hspace{1cm} (15)

Variables: $x_e$ length of the duct, $A(x)$ cross sectional area of the duct, $u$ velocity, $p$ pressure, $\rho$ density, $H$ total enthalpy, $e$ energy, $\gamma$ constant.

Other important quantities: $M = \frac{u}{c}$ mach number, $c = \sqrt{\gamma p/\rho}$ sound velocity, $p_0 = p \left(1 + \frac{\gamma - 1}{2} M^2 \right)^{\frac{\gamma}{\gamma - 1}}$ stagnation pressure.

### 1.2 Inviscid Burgers equation

The simplest model that contains some of the main features of the above systems is the Burgers equation,

$$\partial_t u + \partial_x (u^2/2) = 0, \quad x \in \Omega \subset \mathbb{R}.$$
This is a particular case of a general conservation law
\[ \partial_t u + \partial_x f(u) = 0, \quad x \in \Omega \subset \mathbb{R}, \]
with \( f(u) = u^2/2. \)

2 Solutions of the scalar conservation laws

In this section we briefly describe the main features of the solutions of general scalar conservation laws. We first focus in the simplest advection equation.

2.1 The linear advection equation

Here we consider the linear equation
\[ \partial_t u + a \partial_x u = 0, \quad x \in \mathbb{R}, \quad t > 0 \] \hspace{1cm} (16)
where \( a \) is a given constant. For a given initial datum
\[ u(0, x) = u^0(x), \quad x \in \mathbb{R}, \]
the Cauchy problem is well-defined and the solution is simply
\[ u(t, x) = u^0(x - at), \quad t \geq 0. \]

The solution \( u \) at time \( t = t_0 \) is a pure translation of the initial datum \( u^0 \). In fact, if we define the characteristic lines of (17) as
\[ x'(t) = a, \quad x(0) = x_0 \in \mathbb{R}, \]
the solution \( u \) satisfies
\[ \frac{d}{dt} u(t, x(t)) = 0, \]
i.e., it is constant along each characteristic line.

A similar situation occurs for the linear advection equation with a smooth variable coefficient \( a(x) \),
\[ \partial_t u + \partial_x (a(x) u) = 0, \quad x \in \mathbb{R}, \quad t > 0. \] \hspace{1cm} (17)
If we define the characteristics lines by
\[ x'(t) = a(x(t)), \quad x(0) = x_0, \]
then the solution $u$ can be obtained solving an ordinary differential equation along these characteristics, namely

$$\frac{d}{dt}u(t, x(t)) = -a'(x(t))u(t, x(t)).$$

It is important to remark two important properties

1. Finite speed of propagation. The solution $u(t, x)$ depends on the value of the initial data $u^0$ at the point $x(0)$, where, $x(t)$ is the characteristic line passing through $(t, x)$. Thus, solutions in a bounded region $(t, x) \in B \times (0, T)$ will not be affected by perturbations of the initial datum away from a sufficiently large zone.

2. Solutions of the Cauchy problem with discontinuous initial data can be interpreted using the characteristics. This allows to give a natural notion of weak solution for such cases.

### 2.2 The inviscid Burgers equation

Here we focus in a the simplest non-linear conservation law,

$$\begin{cases}
\partial_t u + \partial_x \left( \frac{u^2}{2} \right) = 0, & \text{in } \mathbb{R} \times (0, T), \\
u(x, 0) = u^0(x),
\end{cases}$$

where

\[ \left\{ \begin{array}{l}
\partial_t u + \partial_x \left( \frac{u^2}{2} \right) = 0, & \text{in } \mathbb{R} \times (0, T), \\
u(x, 0) = u^0(x),
\end{array} \right. \tag{18} \]

Our main objective in this section is to study the analytical properties of the solutions of this problem.

### 2.3 Characteristics

Let $u(x, t)$ be a smooth solution of the Burgers equation (18). Then $u$ is also a solution of

$$\partial_t u + u \partial_x u = 0.$$

We introduce the characteristics as the solutions $x(t)$ of the differential equation

$$\frac{dx}{dt} = u(x, t).$$

Along these characteristic lines the solution is constant since

$$\begin{align*}
\frac{d}{dt}u(x(t), t) &= \partial_t u(x(t), t) + \partial_x (u(x(t), t)) \frac{dx}{dt} \\
&= \partial_t u(x(t), t) + \partial_x (u(x(t), t)) u(x(t), t) = 0.
\end{align*}$$
Thus,
\[
\frac{dx}{dt} = u(x, t) = u^0(x(0), 0).
\]

Note that the characteristics are easily obtained in this case. In fact, they are straight lines whose slopes depend on the value of the initial datum at the intersection of the characteristic line with \( t = 0 \).

Taking into account that the solution \( u \) of (18) is constant along characteristics we can easily solve the Cauchy problem (18) for \( t > 0 \).

Note however that, for some initial data (even for smooth ones) two different characteristics lines may possibly meet at some time \( t = t_0 \) (see figure 2). In this case, the solution cannot be continuous for \( t > t_0 \) and classical solutions of (18) will not exist. In the next subsection we introduce a new definition of solution that allows discontinuities.

2.4 Weak solutions

Let \( u \) a smooth solution of (18) and let \( \varphi \in C^1_0(\mathbb{R} \times [0, T]) \) be a test function. Multiplying the equation of \( u \) by \( \varphi \) and integrating by parts we easily obtain

\[
0 = -\int_0^T \int_{\mathbb{R}} \left( u \partial_t \varphi + \frac{u^2}{2} \partial_x \varphi \right) ~dt ~dx - \int_{\mathbb{R}} u(x, 0)\varphi(x, 0), \forall \varphi \in C^1_0(\mathbb{R} \times [0, T])
\]
Figure 3: Characteristics lines for the Burgers equation.

We adopt identity (19) as the definition of weak solution for (18). Note that a weak solution can be a discontinuous function.

When the solution $u$ is smooth with a single discontinuity on a regular curve $\Sigma \subset (0,T) \times \mathbb{R}$ the following characterization of the solutions is easily proved: $u$ is a weak solution if and only if the following holds:

1. $u$ is a classical solution of (18) in the regions where it is smooth ($C^1$).
2. $u$ satisfies the so-called Rankine-Hugoniot conditions

$$[u]_{\Sigma} n_t + [u^2/2]_{\Sigma} n_x = 0, \text{ on } \Sigma,$$

along the discontinuity $\Sigma$. Here, $[u]_{\Sigma}$ represents the jump of $u$ on $\Sigma$ and $(n_t, n_x)$ the normal vector to $\Sigma$.

If we parametrize the discontinuity $\Sigma$ with a function $s(t)$ by

$$\Sigma = \{(t, s(t)), \ t \in (0, T)\}$$

then $s(t)$ must satisfy

$$s'(t) = \frac{[u^2/2]_{(t, s(t))}}{[u]_{(t, s(t))}}.$$

Weak solutions allows us to determine the physically relevant solution when characteristics intersect. However, this definition does not provide unicity for some initial data.
Assume that the characteristics lines starting at \( t = 0 \) do not fill the whole domain \( (t, x) \in (0, T) \times \mathbb{R} \). Then the characteristics approach is unable to determine a solution (see figure 4).

In this case, the choice \( u(x, t) = x/t \) in the region where characteristics are not defined provides a classical solution.

In general, the physical relevant solution is obtained by defining a new class of solutions introduced by Kruskov, known as entropy solutions, for which unicity holds. Entropy solutions can also be characterized as limits, as \( \varepsilon \to 0 \), of solutions of the Burgers equations with viscosity:

\[
\begin{cases}
\partial_t u + \partial_x \left( \frac{u^2}{2} \right) = \varepsilon \partial_{xx} u, & \text{in } \mathbb{R} \times (0, T), \\
u(x, 0) = u^0(x),
\end{cases}
\]

3 Numerical approximation for scalar conservation laws

In this section we briefly describe some basic concepts on explicit finite differences schemes to solve the scalar one-dimensional conservation law

\[
\begin{align*}
\partial_t u + \partial_x f(u) &= 0, \quad x \in \mathbb{R}, \quad t > 0, \\
u(x, 0) &= u^0(x).
\end{align*}
\]
We assume that \( f \) is a \( C^2 \) function, \( u^0 \in L^\infty(\mathbb{R}) \) and we set
\[
a(u) = f'(u).
\]

We consider a suitable discretization of the domain by considering a uniform spatial grid \( \Delta \) with increment \( \Delta x \) and a time-step \( \Delta t \). We set
\[
\lambda = \frac{\Delta t}{\Delta x}.
\]

We introduce a general 3-point explicit difference scheme of the form
\[
v_{j+1}^n = H(v_j^n, v_{j+1}^n), \quad \forall n \geq 0, j \in \mathbb{Z},
\]  
where \( H : \mathbb{R}^3 \to \mathbb{R} \) is a continuous function and \( v_i^n \) denotes an approximation of the exact solution \( u \) at the grid point \((x_j = j\Delta x, t_n = n\Delta t)\).

### 3.1 Conservative schemes

Now we define a particular class of difference schemes, known as conservative schemes, for which the approximations obtained converge to weak solutions of the continuous equation.

**Definition 1** The difference scheme (22) can be put in conservation form if there exists a continuous function \( g : \mathbb{R}^2 \to \mathbb{R} \) such that
\[
H(v_{-1}, v_0, v_1) = v_0 - \lambda [g(v_{-1}, v_0) - g(v_0, v_1)].
\]  
(23)

The function \( g \) is called the numerical flux.

If we define
\[
g_{j+1/2}^n = g(v_j^n, v_{j+1}^n)
\]
then, the numerical scheme (22) reads
\[
v_{j+1}^n = v_j^n - \lambda(g_{j+1/2}^n - g_{j-1/2}^n).
\]  
(24)

**Definition 2** The difference scheme (24) is consistent with equation (20) if
\[
g(v, v) = f(v), \quad \forall v \in \mathbb{R}.
\]  
(25)
Concerning the initial datum (21) we will consider any suitable discretization. A common choice is to take \[ v_{j,0} = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} u^0(x) \, dx, \] (26) where \( x_{j+1/2} = (x_j + x_{j+1})/2 \).

Finally the approximation by a conservative scheme of (20)-(21) is

\[ v_{j}^{n+1} = v_{j}^{n} - \lambda (g_{j+1/2}^{n} - g_{j-1/2}^{n}), \quad j \in \mathbb{Z}, \; n \geq 0 \] \[ v_{j}^{0} = v_{j,0}. \]   (27) \[ (28) \]

### 3.2 The Lax-Wendroff theorem

In order to study the convergence of the numerical approximations obtained with this discrete system we introduce a suitable representation of the discrete data as functions. Thus, for a given sequence \((v_{j}^{n})\) we introduce the piecewise constant function \(v_{\Delta}\) defined in \((0, \infty) \times \mathbb{R}\) by

\[ v_{\Delta}(t, x) = v_{j}^{n}, \quad t \in [t_{n}, t_{n+1}), \; x \in (x_{j-1/2}, x_{j+1/2}). \] \[ (29) \]

#### Theorem 3 (Lax-Wendroff)

Assume that the difference scheme (24) is consistent with (20) and let \(v^{0} = (v_{j,0})\) be given by (26). Assume that there exists a sequence \(\Delta x \to 0\) such that if \(\Delta t = \lambda \Delta x\) (with \(\lambda\) constant)

\[ \|v_{\Delta}\|_{L^\infty((0, \infty) \times \mathbb{R})} \leq C, \]

\[ v_{\Delta} \text{ converges in } L^{1}_{loc}((0, \infty) \times \mathbb{R}) \text{ and a.e. to a function } u \]

Then \(u\) is a weak solution of (20)-(21).

The above theorem tell us that a difference scheme in conservation form which converges always converges to a weak solution.

The main questions now are:

- Find sufficient conditions to convergence.
- Find criteria which ensure that the limit is the unique entropy solution.
- Determine the order of accuracy of the difference scheme.
3.3 Order of accuracy

The order of accuracy of a difference scheme (22) is the largest number $p \geq 1$ such that any smooth solution $u$ of (20) and for $\lambda$ constant

$$u(t + \Delta t, x) - H(u(x - \Delta x, t), u(x, t), u(x + \Delta x, t)) = O(\Delta t^{p+1}), \quad \text{as } \Delta t \to 0.$$ 

A consistent difference scheme in conservation form is at least order one accurate.

3.4 Stability

We focus on the linear advection equation

$$\partial_t u + a \partial_x u = 0,$$

for which the von Neumann analysis allows to establish a criterium for stability.

Assume that we have a 3-points linear difference scheme of the form

$$v_{j}^{n+1} = c_{-1}v_{j-1}^{n} + c_{0}v_{j}^{n} + c_{1}v_{j+1}^{n}, \quad n \geq 0. \quad (30)$$

Define the $\ell^2$-norm of a sequence $v = (v_j)$ as

$$\|v\|_2 = (\Delta x \sum_j v_j^2)^{1/2}.$$ 

Then, the difference scheme is $L^2$-stable if there exists a constant $C > 0$, independent of $\Delta t$ such that

$$\|v^n\|_2 \leq C\|v^0\|_2, \quad \forall n \geq 0.$$ 

It can be shown that the linear difference scheme (30) can be put in conservation form if and only if

$$c_{-1} + c_{0} + c_{1} = 1.$$ 

The numerical flux is then given by

$$g(u, v) = \frac{c_{-1}u - c_{1}v}{\lambda},$$
and the consistency condition \(25\) reads

\[ c_{-1} - c_1 = \lambda a. \]

Thus, setting

\[ q = 1 - c_0, \]

the conservative and consistent schemes can be written in viscous form as

\[ v_j^{n+1} = v_j^n - \lambda a(v_{j+1}^n - v_{j-1}^n)/2 + q(v_{j+1}^n - 2v_j^n + v_{j-1}^n)/2 \]

The coefficient \( \mu = \lambda a \) is called the Courant number and it can be shown that the differential scheme above is \( L^2 \)-stable if \( q \) satisfies

\[ (\lambda a)^2 \leq q \leq 1. \]

In the particular case \( q = (\lambda a)^2 \) we obtain a second order accurate differential scheme known as Lax-Wendroff scheme which is \( L^2 \)-stable under the condition

\[ \lambda |a| \leq 1. \]

This condition can be interpreted geometrically in terms of the domain of dependence of the numerical difference scheme. This interpretation is known as the Courant-Friedrichs-Levy (CFL) condition.

3.5 Some examples

The Lax-Friedrichs scheme is given by

\[
\begin{align*}
\frac{v_j^{n+1} - v_{j-1}^{n+1} + v_{j+1}^n}{\Delta t} + \frac{f(v_{j+1}^n) - f(v_{j-1}^n)}{2\Delta x} &= 0, \\
v_j^0 &= v_{0,j},
\end{align*}
\]

which can be put in conservation form with the numerical flux

\[ g(u,v) = \frac{f(u) + f(v)}{2} - \frac{v - u}{2\lambda}. \]

In the linear case, \( q = 1 \) and this scheme is \( L^2 \)-stable under the CFL condition.

The upwind scheme is given by

\[
v^n + 1_j = \begin{cases} 
  v_j^n - \lambda(f(v_j^n) - f(v_{j-1}^n)), & \text{if } f' > 0 \\
  v_j^n - \lambda(f(v_{j+1}^n) - f(v_j^n)), & \text{if } f' < 0
\end{cases}
\]
In the linear case, \( q = |a\lambda| = |\nu| \) and the scheme is \( L^2 \)-stable under the CFL condition.

The Godunov scheme is based on the exact solution of local Riemann problems. The numerical flux is given by

\[
g(u, v) = \begin{cases} \min_{w \in [u,v]} f(w), & \text{if } u \leq v \\ \max_{w \in [u,v]} f(w), & \text{if } v \leq u \end{cases}
\]

In the linear case, it coincides with the upwind difference scheme.

## 4 Some optimal control problems for fluids

In this section we state some different optimization problems and the main difficulties to approximate them numerically.

### 4.1 An optimal design problem for Euler equations

Let \( \Omega \subset \mathbb{R}^2 \) be an exterior domain with smooth boundary \( S \) (see Figure 5). We consider the two-dimensional Euler system

\[
\partial_t U + \nabla \cdot F = \partial_t U + \partial_x F_x + \partial_y F_y = 0, \text{ en } \Omega
\]

with \( U, F \) defined in (11)-(12).

The velocity vector field is denoted by \( \mathbf{v} \),

\[
\mathbf{v} = \begin{pmatrix} u \\ v \end{pmatrix}.
\]

We consider the following boundary condition on \( S \):

\[
\mathbf{v} \cdot \mathbf{n}_S = 0, \text{ on } S,
\]

where \( \mathbf{n}_S \) is the exterior unitary normal vector.

System (32) must be completed with some initial conditions

\[
U(x, 0) = U_0(x).
\]

Finally we introduce the functional

\[
J(S) = \int_0^T \int_S g(P) ds \, dt,
\]

15
where $g(s)$ is a given smooth function and the set of admissible designs $U_{ad}$.

The optimal design problem is then: Find $S^{\text{min}} \in U_{ad}$ such that

$$J(S^{\text{min}}) = \min_{S \in U_{ad}} J(S).$$

In presence of a shock $\Sigma$, i.e. a discontinuity of the variables $U$ in the $(t, x)$ variables, the Euler system above must be completed with the Rankine-Hugoniot conditions on the shock:

$$[U]_{\Sigma} n^t_{\Sigma} + [(F_x, F_y)]_{\Sigma} \cdot n^x_{\Sigma} = 0, \quad \text{on } \Sigma,$$

where $(n^t_{\Sigma}, n^x_{\Sigma}) = n_{\Sigma} \in R \times R^2$ is a normal vector to $\Sigma$, and $[U]_{\Sigma}$ is the jump across the shock defined as

$$[U]_{\Sigma} = \lim_{\varepsilon \to 0} U((x, t) + \varepsilon n_{\Sigma}) - \lim_{\varepsilon \to 0} U((x, t) - \varepsilon n_{\Sigma}).$$

Note that shock may possibly meet the boundary $S$ (see Figure 6).

4.2 The nozzle problem for stationary flows

We consider a nozzle of variable area $A(x)$ in an interval $(0, x_e)$. The simplified stationary Euler system is then given by

$$\partial_x (AF) = \frac{dA}{dx} P, \quad x \in (0, x_e),$$

where $U, F$ and $P$ satisfy (14)-(15).

If the solution is regular (no shocks) then the flow is isentropic and there are three main quantities which are conserved along the duct:
Figure 6: The shock $\Sigma$ meets the boundary $S$ on a point at each time $t \in (0, T)$.

\[
A \rho u = Cte, \ x \in (0, x_e), \\
H = Cte, \ x \in (0, x_e), \\
p_0 = Cte, \ x \in (0, x_e).
\]

If the flow is transonic with a shock at a point $x = x_s \in (0, x_e)$, the following Rankine-Hugoniot conditions hold: $[F] = 0$, which is equivalent to

\[
[A \rho u] = 0, \ [\rho u^2 + p] = 0, \ [\rho u H] = 0,
\]

where $[f]$ denotes the jump of $f$ at $x = x_s$ defined as $[f] = f_+ - f_-$, and $f_+$, $f_-$ are the values of $f$ to the left and right of the point $x = x_s$.

In this case, the following quantities are conserved

\[
A \rho u = Cte, \ x \in (0, x_e), \\
H = Cte, \ x \in (0, x_e), \\
p_0 = \begin{cases} 
p^-_0 = Cte, & x \in (0, x_s) \\
p^+_0 = Cte, & x \in (x_s, x_e) \end{cases}.
\]
We assume that the area satisfies some geometric constraints. More precisely, we take \( A \in U_{ad} \) given by

\[
U_{ad} = \left\{ A(x) \in C^1(0, x_e), \ A(0) = A_0, \ A(x_e) = A_e, \ A \text{ convex function with a minimum at } x = x_t \in (0, x_e) \right\}
\]

The point \( x = x_t \) is usually referred as the throat.

If we assume a transonic flow with a shock at an interior point \( x = x_s \). The Mach number at the throat is \( M = 1 \). The boundary conditions in this case are:

\[
H \text{ and } p_0 \text{ at the inlet } x = 0, \quad p \text{ at the outlet } x = x_e.
\]

We choose as objective function the integral of pressure along the duct

\[
J(A) = \int_0^{x_e} g(p) \, dx, \quad A \in U_{ad}.
\]

The optimization problem is then: Find \( A_{\min} \) such that

\[
J(A_{\min}) = \min_{A \in U_{ad}} J(A).
\]

### 4.3 An optimal control problem for Burgers equation

In this section we present a simpler optimization problem for the Burgers equation. This problem has the advantage that contains some of the difficulties of the previous ones and, at the same time, the analysis of the existence of minimizers and its numerical approximation can be made mathematically less formal.

We consider the following inviscid Burgers equation:

\[
\begin{cases}
\partial_t u + \partial_x (u^2) = 0, & \text{in } \mathbb{R} \times (0, T), \\
u(x, 0) = u^0(x), & x \in \mathbb{R}
\end{cases}
\]

(37)

Given a target \( u^d \in L^2(\mathbb{R}) \) we consider the cost functional to be minimized \( J : L^1(\mathbb{R}) \to \mathbb{R} \), defined by

\[
J(u^0) = \int_{\mathbb{R}} |u(x, T) - u^d(x)|^2 \, dx,
\]

(38)
where \( u(x, t) \) is the unique entropy solution of (37).

We also introduce the set of admissible initial data \( \mathcal{U}_{ad} \subset L^1(\mathbb{R}) \).

We consider the optimization problem: Find \( u^{0,\min} \in \mathcal{U}_{ad} \) such that

\[
J(u^{0,\min}) = \min_{u^0 \in \mathcal{U}_{ad}} J(u^0).
\]

(39)

This is one of the model optimization problems that is often addressed in the context of optimal aerodynamic design, the so-called inverse design problem.

### 4.4 General remarks

Existence of minimizers is easily established (at least for the Burgers optimization problem) under some natural assumptions on the class of admissible data \( \mathcal{U}_{ad} \). However, uniqueness is false, in general, due, in particular, to the possible presence of discontinuities in the solutions.

In practical applications and in order to perform numerical computations and simulations one has to replace the continuous optimization problem above by a discrete approximation. It is then natural to consider a discretization of system of equations under consideration and the functional \( J \). If this is done in an appropriate way, the discrete optimization problem has minimizers that are often taken, for small enough mesh-sizes, as approximations of the continuous minimizers. There are however few results in the context of hyperbolic conservation laws proving rigorously the convergence of the discrete optimal controls towards the continuous ones, as the mesh-size goes to zero.

We refer to [9] for such a convergence result for the above optimization problem for Burgers equation. More precisely, we denote by \( u^\Delta \) the approximation of \( u \) obtained by a suitable discretization of system (37) with mesh-sizes \( \Delta x \) and \( \Delta t \) for space-time discretizations. We also denote by \( J^\Delta \) a discretization of \( J \) and by \( \mathcal{U}_{ad,\Delta} \) a discrete version of the set of admissible controls \( \mathcal{U}_{ad} \), and consider the approximate discrete minimization problem

\[
J^\Delta(u^0_{\min}) = \min_{u^0_{\Delta} \in \mathcal{U}_{ad,\Delta}} J^\Delta(u^0_{\Delta}).
\]

For fixed values of the mesh-size \( \Delta \), the existence of minimizers for this discrete problem is often easy to prove. But, even in that case, their convergence as \( \Delta \to 0 \) is harder to show.
From a practical point of view it is however more important to be able to develop efficient algorithms for computing accurate approximations of the discrete minimizers. This is often not an easy matter due to the high number of the parameters involved, the lack of convexity of the functional under consideration, etc. We discuss this point in the following section.

### 4.5 Existence of minimizers

In this section we prove that, under certain conditions on the set of admissible initial data $\mathcal{U}_{ad}$, there exists at least one minimizer of $J$ given in (38).

To simplify the presentation we consider the class of admissible initial data $\mathcal{U}_{ad}$:

$$\mathcal{U}_{ad} = \{ f \in L^\infty(\mathbb{R}), \text{supp}(f) \subset K, ||f||_\infty \leq C \},$$

where $K \subset \mathbb{R}$ be a bounded interval and $C > 0$ a constant. Note however that the same theoretical results and descent strategies can be applied in a much wider class of admissible sets.

**Theorem 4** (see [9]) Assume that $u^d \in L^2(\mathbb{R})$ and that $\mathcal{U}_{ad}$ is as above. Then the minimization problem,

$$\min_{u^0 \in \mathcal{U}_{ad}} J(u^0), \quad (40)$$

has at least one minimizer $u^{0,\min} \in \mathcal{U}_{ad}$.

Uniqueness is in general false for this optimization problem.

**Proof.** We first prove existence. Let $u^0_n \in \mathcal{U}_{ad}$ be a minimizing sequence of $J$. Then $u^0_n$ is bounded in $L^\infty$ and there exists a subsequence, still denoted by $u^0_n$, such that $u^0_n \rightharpoonup u^0_*$ weakly-* in $L^\infty$. Moreover, $u^0_0 \in \mathcal{U}_{ad}$.

Let $u_n(x, t)$ and $u_*(x, t)$ be the entropy solutions of (37) with initial data $u^0_n$ and $u^0_0$ respectively, and assume that

$$u_n(\cdot, T) \rightharpoonup u_*(\cdot, T), \quad \text{in } L^2(\mathbb{R}). \quad (41)$$

Then, clearly

$$\inf_{u^0 \in \mathcal{U}_{ad}} J(u^0) = \lim_{n \to \infty} J(u^0_n) = J(u^0_*),$$

and we deduce that $u^0_*$ is a minimizer of $J$.

Thus, the key point is to prove the strong convergence result (41). Two main steps are necessary to do it: a) The relative compactness of $u_n(\cdot, T)$
in $L^2$. Taking the structure of $\mathcal{U}_{ad}$ into account and using the maximum principle and the finite velocity of propagation that entropy solutions fulfill, it is easy to see that the support of all solutions at time $t = T$, is uniformly included in the same compact set of $\mathbb{R}$. Thus, it is sufficient to prove compactness in $L^2_{loc}$. This is obtained from Oleinik’s one-sided Lipschitz condition,

$$\frac{u(x,t) - u(y,t)}{x-y} \leq \frac{1}{t},$$

which guarantees in fact a uniform bound of the $BV$-norm of $u_n(\cdot, T)$, locally in space (see [6]). The needed compactness property is then a consequence of the compactness of embedding $BV(I) \subset L^2(I)$, for all bounded interval $I$.

b) The identification of the limit as the solution of (37) with initial datum $u^0$. This can be proved using similar compactness arguments passing to the limit in the variational formulation of (37). We refer to [11] for a detailed description of this limit process in the more delicate case where the initial data converge to a Dirac delta.

This completes the prove of the existence of minimizers.

We now prove that the uniqueness of the minimizer is in general false for this type of optimization problems. In fact, we prove that there are target functions $u^d$ for which there exist two different minimizers $u_1^0$ and $u_2^0$ such that the corresponding solutions $u_j$, $j = 1, 2$ satisfy $u_j(T) = u^d$, $j = 1, 2$ in such a way that the minimal value of $J$ vanishes. This is always possible as soon as we deal with solutions having shocks. For example,

$$u_1(x, t) = \begin{cases} 
1 & \text{if } x \leq t/2, \\
0 & \text{if } x > t/2,
\end{cases} \quad u_2(x, t) = \begin{cases} 
1 & \text{if } x \leq t - 1/2, \\
x + (1/2 - x)t & \text{if } t - 1/2 \leq x \leq 1/2, \\
0 & \text{if } x > 1/2,
\end{cases}$$

are two different entropy solutions for which $u_1(x, T) = u_2(x, T)$ at $T = 1$. Thus if we take $T = 1$ and $u^d(x) = u_1(x, 1)$ then there exist two different initial data $u_1^0(x, 0)$ and $u_2^0(x, 0)$ for which $J$ attains its minimum. Note that this is impossible within the class of smooth solutions by backwards uniqueness.

Note that $u^d$ as above does not belong to $L^2(\mathbb{R})$ but, the same argument is valid is $u^d$ is truncated to take zero values at infinity. ■

**Remark 5** The above proof is in fact quite general and it can be adapted to other optimization problems with different functionals and admissible sets.
In particular, using Oleinik’s one-sided Lipschitz condition (42), one can also consider admissible sets of the form: $\mathcal{U}_{ad} = \{ f \in L^1(\mathbb{R}), \text{supp}(f) \subset K, ||f||_1 \leq C \}$.

4.6 The discrete minimization problem

The purpose of this section is to show that discrete minimizers obtained through a numerical scheme to approximate (37) satisfying the so-called OSLC property, converge to a minimizer of the continuous problem as the mesh-size tends to zero. This justifies the usual engineering practice of replacing the continuous functional and model by discrete ones to compute an approximation of the continuous minimizer.

Let us introduce a mesh in $\mathbb{R} \times [0, T]$ given by $(x_j, t^n) = (j\Delta x, n\Delta t)$ ($j = -\infty, ..., \infty$; $n = 0, ..., N + 1$ so that $(N + 1)\Delta t = T$), and let $u^n_j$ be a numerical approximation of $u(x_j, t^n)$ obtained as solution of a suitable discretization of the equation (37).

Let us consider the following approximation of the functional $J$ in (38):

$$J^\Delta(u_0^\Delta) = \frac{\Delta x}{2} \sum_{j=-\infty}^{\infty} (u_j^{N+1} - u_j^d)^2,$$  \hspace{1cm} (43)

where $u_0^\Delta = \{u_j^0\}$ is the discrete initial datum and $u^d_\Delta = \{u_j^d\}$ is the discretization of the target $u^d$ at $x_j$, respectively. A common choice is to take,

$$u_j^d = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} u^d(x) dx,$$  \hspace{1cm} (44)

where $x_{j \pm 1/2} = x_j \pm \Delta x/2$.

We also introduce an approximation of the class of admissible initial data $\mathcal{U}_{ad}$ denoted by $\mathcal{U}_{ad}^\Delta$ and constituted by sequences $\varphi_\Delta = \{\varphi_j\}_{j \in \mathbb{Z}}$ for which the associated piecewise constant interpolation function, that we still denote by $\varphi_\Delta$, defined by

$$\varphi_\Delta(x) = \varphi_j, \quad x_{j-1/2} < x < x_{j+1/2},$$

satisfies $\varphi_\Delta \in \mathcal{U}_{ad}$. Obviously, $\mathcal{U}_{ad}^\Delta$ coincides with the class of discrete vectors with support on those indices $j$ such that $x_j \in K$ and for which the discrete $L^\infty$-norm is bounded above by the same constant $C$. 

22
Finally we introduce a 3-point conservative numerical approximation scheme for (37):

\[ u_{j}^{n+1} = u_{j}^{n} - \lambda \left( g_{j+1/2}^{n} - g_{j-1/2}^{n} \right) = 0, \quad \lambda = \frac{\Delta t}{\Delta x}, \quad j \in \mathbb{Z}, \quad n = 0, ..., N, \]

(45)

where,

\[ g_{j+1/2}^{n} = g(u_{j}^{n}, u_{j+1}^{n}), \]

and \( g \) is the numerical flux. These schemes are consistent with (37) when \( g(u, u) = u^2/2 \).

When the function \( H(u, v, w) = v - \lambda (g(u, v) - g(v, w)) \) is a monotone increasing function in each argument the scheme is said to be monotone. These are particularly interesting schemes since the discrete solutions obtained with them converge to weak entropy solutions of the continuous conservation law, as the discretization parameters tend to zero, under a suitable CFL condition (see [16], Chp.3, Th. 4.2).

For each \( h > 0 \) it is easy to see that the discrete analogue of Theorem 4 holds. In fact this is automatic in the present setting since \( U_{ad}^{\Delta t} \) only involves a finite number of mesh-points. But passing to the limit as \( h \to 0 \) requires a more careful treatment. In fact, for that to be done, one needs to assume that the scheme under consideration satisfies the so-called one-sided Lipschitz condition (OSLC), which is a discrete version of Oleinik’s condition above:

\[ \frac{u_{j+1}^{n} - u_{j}^{n}}{\Delta x} \leq \frac{1}{n \Delta t}. \]

(46)

We then consider the following discrete minimization problem: Find \( u_{\Delta}^{0,\text{min}} \) such that

\[ J^{\Delta}(u_{\Delta}^{0,\text{min}}) = \min_{u_{\Delta}^{0} \in U_{ad}^{\Delta t}} J^{\Delta}(u_{\Delta}^{0}). \]

(47)

It is by now well known that Godunov’s, Lax-Friedrichs and Engquist-Osher schemes satisfy this OSLC condition. We refer to [6] for a discussion of this issue and also for the construction of a second order MUSCL method satisfying OSLC.

**Theorem 6** (see [2]) Assume that \( u_{\Delta}^{n} \) is obtained by a conservative monotone numerical scheme consistent with (37) and satisfying the OSLC condition.

Then:
For all $\Delta x, \Delta t > 0$, the discrete minimization problem (47) has at least one solution $u^{0,\min}_\Delta \in \mathcal{U}_{ad}$.

Any accumulation point of $u^{0,\min}_\Delta$ with respect to the weak-$*$ topology in $L^\infty$, as $\Delta x, \Delta t \to 0$ (with $\Delta t/\Delta x = \lambda$ fixed and under a suitable CFL condition), is a minimizer of the continuous problem (40).

The existence of discrete minimizers in the first statement of the Theorem is obvious in this case since we are dealing with a finite dimensional problem. Actually, at this point the OSLC property is not necessary. However, in more general situations (for other classes of discrete admissible controls) we could apply the same argument as in the proof of Theorem 4 based on the OSLC property and the $BV$ estimates this yields (see [6]).

The second statement is less trivial. It requires, definitely, of the OSLC property to guarantee the compactness of numerical solutions as $\Delta x, \Delta t \to 0$.

Remark 7 The most frequently used conservative 3-point numerical schemes derived to approximate (37) satisfy the hypotheses of Theorem 6. This is in particular the case of the Lax-Friedrichs, Engquist-Osher or Godunov schemes whose numerical fluxes for Burgers equation are given, respectively, by

\begin{align*}
g^{LF}(u,v) &= \frac{u^2 + v^2}{4} - \frac{v - u}{2\lambda}, \quad (48) \\
g^{EO}(u,v) &= \frac{u(u + |u|)}{4} + \frac{v(v - |v|)}{4}, \quad (49) \\
g^{G}(u,v) &= \begin{cases} \\
\min_{w \in [u,v]} w^2/2, & \text{if } u \leq v, \\
\max_{w \in [u,v]} w^2/2, & \text{if } u \geq v.
\end{cases} \quad (50)
\end{align*}

Proof. (of Theorem 6) The case where $\Delta x$ and $\Delta t$ are fixed being trivial, we address the limit problem $\Delta \to 0$.

We follow an standard $\Gamma$-convergence argument. The key ingredient is the following continuity property: Assume that $u^0_\Delta \in \mathcal{U}_{ad}^{\Delta x}$ satisfies $u^0_\Delta \rightharpoonup u^0$ in $L^\infty(\mathbb{R})$ with respect to the weak-$*$ topology, then

$$J^{\Delta}(u^0_\Delta) \to J(u^0). \quad (51)$$

This is due to the fact that the OSLC condition guarantees uniform local $BV$ bounds on the discrete solutions and the compactness of the embedding and
the finite velocity of propagation of solutions in view of the class of initial data \( U_{ad} \) we are considering (see [6]) which guarantees that \( u_{\Delta}^{N+1} \to u(\cdot, T) \) in \( L^2(\mathbb{R}) \), under the CFL condition that guarantees the convergence of the numerical schemes, and its monotonicity.

Now, let \( \hat{u}^0 \in U_{ad} \) be an accumulation point of \( u_{\Delta}^{0,\text{min}} \) with respect to the weak--* topology of \( L^\infty \). To simplify the notation we still denote by \( u_{\Delta}^{0,\text{min}} \) the subsequence for which \( u_{\Delta}^{0,\text{min}} \to \hat{u}^0 \), weakly--* in \( L^\infty(\mathbb{R}) \), as \( \Delta x \to 0 \). Let \( v^0 \in U_{ad} \) be any other function. We are going to prove that

\[
J(\hat{u}^0) \leq J(v^0). \tag{52}
\]

To do this we construct a sequence \( v_{\Delta}^0 \in U_{ad}^{\Delta} \) such that \( v_{\Delta}^0 \to v^0 \), in \( L^1(\mathbb{R}) \), as \( \Delta x, \Delta t \to 0 \) (we can consider in particular the approximation in (44)).

Taking into account the continuity property (51), we have

\[
J(v^0) = \lim_{\Delta \to 0} J(\Delta)(v_{\Delta}^0) \geq \lim_{\Delta \to 0} J(\Delta)(u_{\Delta}^{0,\text{min}}) = J(\hat{u}^0),
\]

which proves (52).

**Remark 8** Theorem 6 concerns global minima. However, both the continuous and discrete functionals may possibly have local minima as well. Extending this kind of \( \Gamma \)-convergence result for local minima requires important further developments.

### 5 Gradient calculus

The most efficient methods to approximate minimizers are the gradient methods (steepest descent, conjugate gradient, etc.) although they hardly distinguish local or global minimizers. This is an added difficulty in problems with many local minima, a fact that cannot be excluded in our optimization problems, due to the nonlinear dependence of the state on the initial datum. However we will not address this problem here. We shall rather focus on building efficient descent algorithms.

Descent algorithms are iterative processes. In each step of the iteration the descent direction is built by means of the gradient of the functional with respect to the controls.
5.1 Unconstrained optimization

Let $H$ be a Hilbert space and $J : H \to \mathbb{R}$ a continuous functional. We are going to characterize the minima of $J$ in $H$ and to briefly describe the gradient type methods to approximate them.

**Definition 9** The directional derivative of $J$ at $u \in H$ in the direction $v \in H$ is defined by

$$
\lim_{\lambda \to 0, \lambda > 0} \frac{J(u + \lambda v) - J(u)}{\lambda}.
$$

(53)

If this limit exists for any $v \in H$, and there exists a linear map $J'(u) \in H'$ (dual space of $H$) such that

$$
<J'(u), v>_{H', H} = \lim_{\lambda \to 0, \lambda > 0} \frac{J(u + \lambda v) - J(u)}{\lambda},
$$

(54)

where $<\cdot, \cdot>_{H', H}$ denotes the duality product in $H$, then we say that $J$ is Gateaux differentiable at $u$ and we refer as $J'(u)$ the Gateaux derivative at $u$.

A functional $J$ is said to be Frechet differentiable at $u \in H$ if there exists a linear map $DJ(u) \in H'$ such that

$$
\lim_{\|v\|_H \to 0} \frac{|J(u + v) - J(u) - <DJ(u), v>_{H', H}|}{\|v\|_H} = 0.
$$

(55)

If $J$ is Gateaux (Frechet) differentiable for any $u \in H$ then we say that $J$ is Gateaux (Frechet) differentiable.

**Definition 10** Assume that $J$ is Gateaux differentiable at $u$. Then, there exists $\text{Grad } J(u) \in H$ such that

$$
<J'(u), v>_{H', H} = \langle \text{Grad } J(u), v \rangle, \quad \forall v \in H,
$$

(56)

where $(\cdot, \cdot)$ represents the scalar product in $H$. $\text{grad } J(u)$ is the gradient of $J$ at $u$.

**Remark 11** If $J$ is Frechet differentiable at $u$ then $J$ is Gateaux differentiable at $u$.

**Remark 12** If the Frechet derivative $DJ(u)$ is continuous from $H$ to $H'$, then we will write $J \in C^1$ and the Gateaux and Frechet derivatives coincide.
**Theorem 13** Assume that there exists \( u_0 \in H \) such that

\[
J(u_0) = \inf_{u \in H} J(u)
\]  
(57)

and that \( J \) is Gateaux differentiable. Then

\[
J'(u_0) = 0.
\]  
(58)

**Theorem 14** Assume that \( J \) is Gateaux differentiable, convex and that there exists \( u_0 \in H \) such that

\[
J'(u_0) = 0.
\]  
(59)

Then

\[
J(u_0) = \inf_{u \in H} J(u).
\]  
(60)

Once we have identified the minimizers of the functional \( J \) as their critical points we describe the gradient and Newton type methods.

### 5.2 Gradient Methods

Let \( J : H \to R \) be a Gateaux differentiable functional. Then

\[
J(u + \lambda v) = J(u) + \lambda (\text{Grad } J(u), v) + o(\lambda), \quad \forall v \in H, \lambda \in R.
\]  
(61)

Gradient methods are based on the following remark: if we take \( v = -\text{Grad } J(u) \) and \( 0 < \lambda << 1 \), in (61) we have

\[
J(u - \lambda \text{Grad } J(u)) - J(u) = -\lambda \|\text{Grad } J(u)\|^2 + o(\lambda)
\]  
(62)

that must be negative if \( \lambda \) is sufficiently small. Therefore, if \( \lambda \) is small,

\[
J(u - \lambda \text{Grad } J(u)) < J(u)
\]  
(63)

and the sequence

\[
u^{n+1} = u^n - \lambda \text{Grad } J(u^n), \quad n = 1, 2, ...
\]  
(64)

is such that \( J(u^n) \) decreases.

**Theorem 15** If \( J \) is continuously differentiable, bounded from below and coercive, then any accumulation point \( u^* \) of \( u^n \) generated by (64) satisfies

\[
\text{Grad } J(u^*) = 0.
\]  
(65)
When we consider the best step size $\lambda^n$ in the descent direction $-\text{Grad} J(u^n)$,

$$
\lambda^n = \arg \min_{\lambda} \{J(u^n - \lambda \text{Grad} J(u^n))\}
$$

we obtain the so-called **steepest descent method with optimal step**.

The choice of the optimal step size requires a one-dimensional optimization algorithm which is not necessarily easy. However, there exist other methods to find efficient step sizes. However we will not pursue in that direction. The following algorithm consider an almost constant step size and it is easy to implement numerically.

---

**5.3 Newton method**

Newton method is usually applied to approximate the roots of an equation or system of equations. In our case we apply it to compute the critical points

$$
J'(u) = 0.
$$

As the gradient method it also consists in an iterative method in which the new approximation is obtained from the previous one by means of the formula

$$
u^{n+1} = u^n - v^n
$$

where $v^n$ is the solution of the system

$$
J''(u^n) \cdot v^n = \text{Grad} J(u^n).
$$
Newton method converges very fast but it requires to compute the second derivative of the functional \( J \). It is possible to circumvent this difficulty by computing an approximation of this second derivative \( J''(u^n) \). This produce the well-known quasi-Newton methods.

A possible implementation of the Newton method is as follows:

\[
\begin{align*}
\text{Elección de } u, \varepsilon > 0 \\
\text{Cálculo de } \text{Grad } J(u) \\
\text{WHILE } \|\text{Grad } J(u)\| > \varepsilon \\
\text{Cálculo de } J''(u) \\
\text{Solve } J''(u) \cdot v = \text{Grad } J(u) \\
u = u - v \\
\text{Compute } \text{Grad } J(u) \\
\end{align*}
\]

END

5.4 Adjoint based methods to compute the gradient

We explain this method in a simple example. Let \( \Omega \subset \mathbb{R}^n \) be a \( C^2 \) domain and \( \omega \subset \Omega \) a nonempty open subset. Consider the following system:

\[
\begin{align*}
-\Delta u(x) &= g(x) \chi_\omega(x), \text{ en } \Omega \\
u &= 0, \text{ en } \partial\Omega,
\end{align*}
\]

(70)

where \( \chi_\omega(x) \) is the characteristic function of the subset \( \omega \).

We consider the following optimal control problem:

Problem 1: Given \( A \subset \Omega \) and \( h \in L^2(A) \), find \( g \in L^2(\Omega) \) such that \( u(x)\chi_A(x) = h(x) \) on the subset \( A \subset \Omega \).

In general, this control problem has no solution. For example, if \( A \) is a subset of \( \omega \), a necessary condition to have solution is that \( h \) must be harmonic on \( A \setminus \omega \). Therefore, this type of problems are usually stated as optimal control problems:

Problem 2: Given \( A \subset \Omega \) and \( h \in L^2(A) \), find \( g \in L^2(\Omega) \) such that \( u(x)\chi_A(x) \) is as close as possible as \( h(x) \) on the subset \( A \).

To solve this problem we introduce the functional \( J_k : L^2(\omega) \to \mathbb{R} \) defined by

\[
J_k(g) = \frac{1}{2} \int_A |u - h|^2 dx + \frac{k}{2} \int_\omega |g|^2 dx
\]

and we look for

\[
\min_{g \in L^2(\omega)} J_k(g), \quad k \text{ sufficiently large.}
\]
We observe that $J_k$ is continuous, coercive and convex functional and therefore it has a minimizer $g_k$. It is easy to prove that, if Problem 1 has solution, then

$$g_k \to g \quad \text{in } L^2(\omega).$$

The minimization of $J_k$ requires to compute the gradient $\text{grad } J_k(g)$. It is easy to check that

$$\text{grad } J_k(g) = p \big|_\omega + kg$$

where $p(x)$ is the solution of the adjoint problem

$$\begin{cases}
-\Delta p = (u - h)|_A, & x \in \Omega \\
p = 0, & x \in \partial \Omega,
\end{cases}$$

where $u$ is the solution of (70).

6 Gradient calculus for conservation laws

In this section we follow the analysis in [9]. Descent algorithms are iterative processes. In each step of the iteration the descent direction is built by means of the gradient of the functional with respect to the controls. For example, in the case of the optimization problem for the Burgers equation above, the sensibility of the discrete cost functional $J^\Delta$ with respect to $u^\Delta_0$ depends on the sensitivity of the solution of the numerical scheme, used to discretize (37), with respect to $u^\Delta_0$, which in fact involves an infinite number of parameters, one for each mesh-point. Thus, in practice, computing the sensibility of the cost functional requires to differentiate this numerical scheme with respect to the initial datum.

When the numerical scheme under consideration is differentiable this is easy to do and the classical adjoint state method provides a significant shortcut when computing the derivatives with respect to all control parameters (mesh-points in this case). We illustrate this below.

But for large complex systems, as Euler equations in higher dimensions, the existing most efficient numerical schemes (upwind, Godunov, Roe, etc.) are not differentiable (see for example [18] or [20]). In this case, the gradient of the functional is not well defined and there is not a natural and systematic way to compute its variations.

In face of this difficulty, it would be natural to explore the possible use of non-smooth optimization techniques. But this does not seem to had been
done and is out of the scope of this notes. By the contrary, the following two other approaches have been developped: The first one, is based on the use of automatic differentiation, which basically consists in differentiating the numerical method (even if it is not differentiable!), differentiating each line in the code (see for instance [25]). This approach often produces oscillations which are due precisely to the lack of differentiability. The second one, the so-called continuous approach, consists in proceeding in two steps as follows: One first linearizes the continuous system (37) to obtain a descent direction of the continuous functional $J$ and then takes a numerical approximation of this descent direction with the discrete values provided by the numerical scheme. Of course the validity of this approximation as a descent direction for the discrete problem is not at all assured either.

But the continuous approach has to face another major drawback, when solutions develop shock discontinuities, as it is the case in the context of the hyperbolic conservation laws we are considering here. Indeed, the formal differentiation of the continuous state equation (37) yields

$$\partial_t \delta u + \partial_x (u \delta u) = 0.$$ 

But this is only justified when the state $u$ on which the variations are being computed, is smooth enough. In particular, it is not justified when the solutions are discontinuous since singular terms may appear on the linearization over the shock location. Accordingly in optimal control applications one also needs to take into account the sensitivity for the shock location. This has been studied by different authors with different approaches (see [28], [15] or [8]). Roughly speaking, the main conclusion of that analysis is that the classical linearized system for the variations of the solutions must be complemented with some new equations for the sensitivity of the shock position.

These issues had been the object of intensive research, as indicated above, but there is not a systematic receipt about how to use this new notions of linearizations to implement efficient descent methods. This is due, to some extent, to the fact that two related but different issues have been treated simultaneously without sufficiently distinguishing one from another: a) The lack of regularity of the solutions of the continuous state equation that makes the formal linearization above difficult to justify and that adds some unexpected terms to the classical derivative of the functional, to take into account the possible contribution of jump discontinuities, and b) the numerical schemes being non-differentiable.
In presence of shocks, the existing results allow deriving the correct linearization of the system. We then derive the adjoint system which contains an internal boundary condition along the shock which has been referred in the literature as an internal boundary condition for the adjoint system (see [13] where the quasi one dimensional stationary Euler equations are considered). From the numerical point of view, the use of this adjoint system makes methods more efficient, since it takes into account explicitly the sensibility of the solution with respect to shock variations. But if applied directly with the aid of the notion of generalized tangent vectors ([7], [8]) the descent method, in each step of the iteration, adds new discontinuities to the state, thus yielding solutions with increasing complexity.

The detailed analysis of the continuous linearized equations in the presence of shocks is only well-understood in a certain number of situations: 1−d scalar conservation laws ([3], [4]) with the aid of notions of duality and reversible measure valued solutions, multi-dimensional scalar conservation was subject to one-sided Lipschitz conditions ([5]), and also when the shock is a priori known to be located on a single regular curve, or a regular manifold in higher dimensions (see [7], [8] and [28] for 1−d problems, and in the multi-dimensional case [23] and [24] where general systems of conservation laws in higher dimensions are considered). We also refer to [15] for an analysis of the linearization for multi-dimensional perturbations of 1−d scalar conservation laws.

6.1 The continuous approach without shocks

In this subsection we give an expression for the sensitivity of the functional \( J \) with respect to the initial datum based on a classical adjoint calculus for smooth solutions. First we present a formal calculus and then we show how to justify it when dealing with a classical solution for (37), i.e. when there are no discontinuities.

Let \( C^1_0(\mathbb{R}) \) be the set of \( C^1 \) functions with compact support and let \( u^0 \in C^1_0(\mathbb{R}) \) be a given datum for which there exists a classical solution \( u(x, t) \) of (37) in \( (x, t) \in \mathbb{R} \times [0, T] \), which can be extended to a classical solution in \( t \in [0, T+\tau] \) for some \( \tau > 0 \). Note that this imposes some restrictions on \( u^0 \) other than being smooth. More precisely, we must have \( T + \tau > \max_x [1/u'_0(x)] \) to guarantee that two different characteristics do not meet in the time interval \([0, T + \tau]\). Let \( \delta u^0 \in C^1_0(\mathbb{R}) \) be any possible variation of the initial datum \( u^0 \). Due to the finite speed of propagation, this perturbation will only affect the
solution in a bounded set of \((x,t) \in \mathbb{R} \times [0,T]\). This simplifies the argument below that applies in a much more general setting provided solutions are smooth enough.

Then for \(\varepsilon > 0\) sufficiently small, the solution \(u^\varepsilon(x,t)\) corresponding to the initial datum

\[
u^\varepsilon_0(x) = u^0(x) + \varepsilon \delta u^0(x),
\]

is also a classical solution in \((x,t) \in \mathbb{R} \times (0,T)\) and \(u^\varepsilon \in C^1(\mathbb{R} \times [0,T])\) can be written as

\[
u^\varepsilon = u + \varepsilon \delta u + o(\varepsilon),
\]

where \(\delta u\) is the solution of the linearized equation,

\[
\begin{cases}
    \partial_t \delta u + \partial_x (u \delta u) = 0, \\
    \delta u(x,0) = \delta u^0(x).
\end{cases}
\]

Let \(\delta J\) be the Gateaux derivative of \(J\) at \(u^0\) in the direction \(\delta u^0\). We have

\[
\delta J = \int_{\mathbb{R}} (u(x,T) - u^d(x)) \delta u(x,T) \, dx,
\]

where \(\delta u\) solves the linearized system above (73). Now, we introduce the adjoint system

\[
\begin{cases}
    -\partial_t p - u \partial_x p = 0, \\
    p(x,T) = p^T(x),
\end{cases}
\]

where \(p^T = u(x,T) - u^d(x)\). Multiplying the equations satisfied by \(\delta u\) by \(p\), integrating by parts, and taking into account that \(p\) satisfies (75), we easily obtain

\[
\int_{\mathbb{R}} (u(x,T) - u^d(x)) \delta u(x,T) \, dx = \int_{\mathbb{R}} p(x,0) \delta u^0 \, dx.
\]

Thus, \(\delta J\) in (74) can be written as,

\[
\delta J = \int_{\mathbb{R}} p(x,0) \delta u^0(x) \, dx.
\]

This expression provides an easy way to compute a descent direction for the continuous functional \(J\), once we have computed the adjoint state. We just take

\[
\delta u^0 = -p(x,0).
\]
Under the assumptions above on \( u^0, u, \delta u, \) and \( p \) can be obtained from their data \( u^0(x), \delta u^0(x) \) and \( p^T(x) \) by using the characteristic curves associated to (37). For the sake of completeness we briefly explain this below.

The characteristic curves associated to (37) are defined by

\[
x'(t) = u(t, x(t)), \quad t \in (0, T); \quad x(0) = x_0.
\]

They are straight lines whose slopes depend on the initial data:

\[
x(t) = x_0 + tu^0(x_0), \quad t \in (0, T).
\]

As we are dealing with classical solutions, \( u \) is constant along such curves and, by assumption, two different characteristic curves do not meet each other in \( \mathbb{R} \times [0, T + \tau] \). This allows to define \( u \) in \( \mathbb{R} \times [0, T + \tau] \) in a unique way from the initial data.

For \( \varepsilon > 0 \) sufficiently small, the solution \( u^\varepsilon(x, t) \) corresponding to the initial datum (71) has similar characteristics to those of \( u \). This allows guaranteeing that two different characteristic lines do not intersect for \( 0 \leq t \leq T \) if \( \varepsilon > 0 \) is small enough. Note that \( u^\varepsilon \) may possibly be discontinuous for \( t \in (T, T + \tau] \) if \( u^0 \) generates a discontinuity at \( t = T + \tau \) but this is irrelevant for the analysis in \( [0, T] \) we are carrying out. Therefore \( u^\varepsilon(x, t) \) is also a classical solution in \( (x, t) \in \mathbb{R} \times [0, T] \) and it is easy to see that the solution \( u^\varepsilon \) can be written as (72) where \( \delta u \) satisfies (73).

System (73) can be solved again by the method of characteristics. In fact, as \( u \) is a regular function, the first equation in (73) can be written as

\[
\frac{d}{dt} \delta u(x(t), t) = -\partial_x u(x(t), t) \delta u,
\]

i.e.

\[
\frac{d}{dt} \delta u(x(t), t) = -\partial_x u(x(t), t) \delta u,
\]

where \( x(t) \) are the characteristics curves defined by (79). Thus, the solution \( \delta u \) along a characteristic line can be obtained from \( \delta u^0 \) by solving this differential equation, i.e.

\[
\delta u(x(t), t) = \delta u^0(x_0) \exp \left( -\int_0^t \partial_x u(x(s), s) ds \right).
\]

Finally, the adjoint system (75) is also solved by characteristics, i.e.

\[
p(x(t), t) = p^T(x(T)).
\]
This yields the steepest descent direction in (78) for the continuous functional.

**Remark 16** Note that for classical solutions the Gateaux derivative of $J$ at $u^0$ is given by (77) and this provides an obvious descent direction for $J$ at $u^0$, given by $\delta u^0 = -p(x,0) \in C^1_0(\mathbb{R})$. However this is not very useful in practice since, even when we initialize the iterative descent algorithm with a smooth $u^0$ we cannot guarantee that the solution will remain classical along the iteration.

### 6.2 The continuous approach in presence of shocks

In this section we collect some existing results on the sensitivity of the solution of conservation laws in the presence of shocks. We follow the analysis in [7] but similar results in different forms and degrees of generality can be found in [1], [3], [4], [28] or [15], for example.

We focus on the particular case of solutions having a single shock, but the analysis can be extended to consider more general one-dimensional systems of conservation laws with a finite number of noninteracting shocks (see [7]). The theory of duality and reversible solutions developed in [3], [4] is the one leading to more general results.

We introduce the following hypothesis:

**$(H)$** Assume that $u(x,t)$ is a weak entropy solution of (37) with a discontinuity along a regular curve $\Sigma = \{(t, \varphi(t)), t \in [0,T]\}$, which is Lipschitz continuous outside $\Sigma$. In particular, it satisfies the Rankine-Hugoniot condition on $\Sigma$

$$\varphi'(t)[u]_{\varphi(t)} = [u^2/2]_{\varphi(t)}.$$  

(82)

Here we have used the notation: $[v]_{x_0} = v(x_0^+) - v(x_0^-)$ for the jump at $x_0$ of any piecewise continuous function $v$ with a discontinuity at $x = x_0$.

Note that $\Sigma$ divides $\mathbb{R} \times (0,T)$ in two parts: $Q^-$ and $Q^+$, the subdomains of $\mathbb{R} \times (0,T)$ to the left and to the right of $\Sigma$ respectively.

As we will see, in the presence of shocks, for correctly dealing with optimal control and design problems, the state of the system has to be viewed as being a pair $(u, \varphi)$ combining the solution of (37) and the shock location $\varphi$. This is relevant in the analysis of sensitivity of functions below and when applying descent algorithms.
Figure 7: Subdomains $Q^-$ and $Q^+$.

Then the pair $(u, \varphi)$ satisfies the system

$$
\begin{aligned}
\partial_t u + \partial_x (u^2 \frac{u^2}{2}) &= 0, & \text{in } Q^- \cup Q^+, \\
\varphi'(t)[u]_{\varphi(t)} &= [u^2/2]_{\varphi(t)}, & t \in (0, T), \\
\varphi(0) &= \varphi^0, \\
u(x, 0) &= u^0(x), & \text{in } \{x < \varphi^0\} \cup \{x > \varphi^0\}.
\end{aligned}
$$

(83)

We now analyze the sensitivity of $(u, \varphi)$ with respect to perturbations of the initial datum, in particular, with respect to variations $\delta u^0$ of the initial profile $u^0$ and $\delta \varphi^0$ of the shock location $\varphi^0$. To be precise, we adopt the functional framework based on the generalized tangent vectors introduced in [7].

**Definition 17** ([7]) Let $v : \mathbb{R} \to \mathbb{R}$ be a piecewise Lipschitz continuous function with a single discontinuity at $y \in \mathbb{R}$. We define $\Sigma_v$ as the family of all continuous paths $\gamma : [0, \varepsilon_0] \to L^1(\mathbb{R})$ with

1. $\gamma(0) = v$ and $\varepsilon_0 > 0$ possibly depending on $\gamma$.

2. For any $\varepsilon \in [0, \varepsilon_0]$ the functions $u^\varepsilon = \gamma(\varepsilon)$ are piecewise Lipschitz with a single discontinuity at $x = y^\varepsilon$ depending continuously on $\varepsilon$ and there exists a constant $L$ independent of $\varepsilon \in [0, \varepsilon_0]$ such that

$$
|u^\varepsilon(x) - u^\varepsilon(x')| \leq L|x - x'|,
$$

whenever $y^\varepsilon \notin [x, x']$. 

36
Furthermore, we define the set $T_v$ of generalized tangent vectors of $v$ as the space of $(\delta v, \delta y) \in L^1 \times \mathbb{R}$ for which the path $\gamma_{(\delta v, \delta y)}$ given by

$$\gamma_{(\delta v, \delta y)}(\varepsilon) = \begin{cases} v + \varepsilon \delta v + [v]_y \chi_{[y+\varepsilon \delta y, y]} & \text{if } \delta y < 0, \\ v + \varepsilon \delta v - [v]_y \chi_{[y,y+\varepsilon \delta y]} & \text{if } \delta y > 0, \end{cases}$$

satisfies $\gamma_{(\delta v, \delta y)} \in \Sigma_v$.

Finally, we define the equivalence relation $\sim$ defined on $\Sigma_v$ by

$$\gamma \sim \gamma' \text{ if and only if } \lim_{\varepsilon \to 0} \|\gamma(\varepsilon) - \gamma'(\varepsilon)\|_{L^1} = 0,$$

and we say that a path $\gamma \in \Sigma_v$ generates the generalized tangent vector $(\delta v, \delta y) \in T_v$ if $\gamma$ is equivalent to $\gamma_{(\delta v, \delta y)}$ as in (84).

**Remark 18** The path $\gamma_{(\delta v, \delta y)} \in \Sigma_v$ in (84) represents, at first order, the variation of a function $v$ by adding a perturbation function $\varepsilon \delta v$ and by shifting the discontinuity by $\varepsilon \delta y$.

Note that, for a given $v$ (piecewise Lipschitz continuous function with a single discontinuity at $y \in \mathbb{R}$) the associated generalized tangent vectors $(\delta v, \delta y) \in T_v$ are those pairs for which $\delta v$ is Lipschitz continuous with a single discontinuity at $x = y$.

Let $u^0$ be the initial datum in (83) that we assume to be Lipschitz continuous to both sides of a single discontinuity located at $x = \varphi^0$, and consider a generalized tangent vector $(\delta u^0, \delta \varphi^0) \in L^1(\mathbb{R}) \times \mathbb{R}$ for all $0 \leq t \leq T$. Let $u^{0,\varepsilon} \in \Sigma_{u^0}$ be a path which generates $(\delta u^0, \delta \varphi^0)$. For $\varepsilon$ sufficiently small the solution $u^\varepsilon(\cdot, t)$ of (83) is Lipschitz continuous with a single discontinuity at $x = \varphi^\varepsilon(t)$, for all $t \in [0, T]$. Thus $u^\varepsilon(\cdot, t)$ generates a generalized tangent vector $(\delta u^\varepsilon(\cdot, t), \delta \varphi(t)) \in L^1 \times \mathbb{R}$. Moreover, in [8] it is proved that it satisfies the following linearized system:

$$\begin{cases} \partial_t \delta u + \partial_x (u \delta u) = 0, & \text{in } Q^- \cup Q^+, \\
\delta \varphi'(t)[u]_{\varphi(t)} + \delta \varphi(t) \left( \varphi'(t)[u_x]_{\varphi(t)} - [u_x u]_{\varphi(t)} \right) + \varphi'(t) [u \delta u]_{\varphi(t)} - [u \delta u]_{\varphi(t)} = 0, & \text{in } (0, T), \\
\delta u(x, 0) = \delta u^0, & \text{in } \{x < \varphi^0\} \cup \{x > \varphi^0\}, \\
\delta \varphi(0) = \delta \varphi^0, \end{cases}$$

with the initial data $(\delta u^0, \delta \varphi^0)$. 37
Remark 19 The linearized system (85) can be obtained, at least formally, by a perturbation argument in two steps: first we make the change of variables $\hat{x} = x - \varphi(t)$ which transforms system (85) in a new coupled system but in a fixed domain $\{\hat{x} < 0\} \cup \{\hat{x} > 0\}$, and where the variable $\varphi$ enters in the equations satisfied by $u$ to both sides of $\hat{x} = 0$. Then, we introduce a perturbation of the data $(u_0^\varepsilon, \varphi_0^\varepsilon) = (u_0, \varphi_0) + \varepsilon(\delta u_0, \delta \varphi_0)$ and compute the equations of the first order perturbation of the solution. This is in fact the usual approach in the study of the linearized stability of shocks. We refer to [15] for a detailed description of this method in the scalar case and [23], [24] for more general systems of conservation laws in higher dimensions.

In this way, we can obtain formally the expansion

$$(u_\varepsilon, \varphi_\varepsilon) = (u, \varphi) + \varepsilon(\delta u, \delta \varphi) + O(\varepsilon^2).$$

However, this expansion is only justified for general scalar one-dimensional conservation laws of the form

$$\partial_t u + \partial_x (f(u)) = 0,$$

when the function $f \in C^1$ is convex. In this case, it is possible to establish a differentiability result of the solution $u$ with respect to small perturbations of the initial data $u_0$ and the discontinuity position $\varphi_0$ (see [7]). For more general situations this differentiability has been proved only in a weak sense, as in [7] for systems of conservation laws, or [28], for scalar equations in several space dimensions.

Remark 20 The linearized system (85) has a unique solution which can be computed in two steps. The method of characteristics determines $\delta u$ in $Q^- \cup Q^+$, i.e. outside $\Sigma$, from the initial data $\delta u_0$, by the method of characteristics (note that system (85) has the same characteristics as (83)). This yields the value of $u$ and $u_x$ to both sides of the shock $\Sigma$ and allows determining the coefficients of the ODE that $\delta \varphi$ satisfies. Then, we solve the ordinary differential equation to obtain $\delta \varphi$.

In this section we have assumed that the discontinuity of the solution of the Burgers equation $u$ is present in the whole time interval $t \in [0, T]$. It is interesting to note that discontinuities may appear at time $\tau \in (0, T)$ for some regular initial data. In this case we can, at least formally, obtain a generalization of (85). Let us show this in a particular situation. Assume
that \( u^0 \) is a regular initial datum for which the weak entropy solution \( u \) of the Burgers equation has a discontinuity at \( x = \varphi(t) \) with \( t \in [\tau, T] \). Assume that we consider variations \( \delta u^0 \) for which the corresponding solution of \( (37) \) has also a discontinuity in the same time interval \( t \in [\tau, T] \). Then, the linearization can be done separately in \( t \in [0, \tau) \) and in \( t \in [\tau, T] \). The linearized equations in the last interval are similar to the ones obtained in \( (85) \).

Concerning the interval \( [0, \tau) \) the solution is regular and the linearization is obviously given by \( (73) \). The only question is then how to compute the value of \( \delta \varphi(\tau) \) from the initial datum \( \delta u^0 \). This can be obtained by linearizing the weak formulation of the Burgers equation in \( t \in (0, \tau) \).

The weak solutions of the Burgers equation satisfy

\[
- \int_0^\tau \int_\mathbb{R} u \psi_t \, dx \, dt - \int_0^\tau \int_\mathbb{R} u^2 \psi_x \, dx \, dt + \int_\mathbb{R} u(x, \tau) \psi(x, \tau) \, dx \\
- \int_\mathbb{R} u^0(x) \psi(x, 0) \, dx = 0, \quad \forall \psi \in C^1_0(\mathbb{R} \times [0, \tau]).
\]

The linearized weak formulation is given by

\[
- \int_0^\tau \int_\mathbb{R} \delta u \psi_t \, dx \, dt - \int_0^\tau \int_\mathbb{R} u \delta u \psi_x \, dx \, dt + \int_\mathbb{R} \delta u(x, \tau) \psi(x, \tau) \, dx - [u^0]_{\varphi(\tau)} \delta \varphi(\tau) \psi(\varphi(\tau), \tau) \\
- \int_\mathbb{R} \delta u^0(x) \psi(x, 0) \, dx = 0, \quad \forall \psi \in C^1_0(\mathbb{R} \times [0, \tau]).
\]

Taking into account that \( \delta u \) is constant along the characteristic lines of the Burgers equation we easily obtain

\[
- \int_D \delta u \psi_t \, dx \, dt - \int_D u \delta u \psi_x \, dx \, dt - [u^0]_{\varphi(\tau)} \delta \varphi(\tau) \psi(\varphi(\tau), \tau) - \int_{D_0} \delta u^0(x) \psi(x, 0) \, dx = 0, \quad \forall \psi \in C^1_0(\mathbb{R} \times [0, \tau]).
\]

where \( D \) is the triangular region \( D \in \mathbb{R} \times [0, \tau] \) occupied by the characteristics that meet at \( (x, t) = (\varphi(\tau), \tau) \) and \( D_0 \) is \( D \cap \{ t = 0 \} \). Taking, in particular, \( \psi(x, t) = 1 \) in \( (x, t) \in D \) we obtain

\[
\int_{D_0} \delta u^0(x) \, dx = -[u^0]_{\varphi(\tau)} \delta \varphi(\tau).
\]
Thus, the linearized system in this case reads,

$$\begin{cases}
\partial_t \delta u + \partial_x (u \delta u) = 0, & \text{in } Q^- \cup Q^+,
\delta \varphi'(t)[u]_{\varphi(t)} + \delta \varphi(t) \left( \varphi'(t)[u_x]_{\varphi(t)} - [u_x u]_{\varphi(t)} \right) + \varphi'(t)[\delta u]_{\varphi(t)} - [u \delta u]_{\varphi(t)} = 0, & \text{in } (\tau, T),
\delta \varphi(\tau) = -\frac{1}{[u^0]_{\varphi(\tau)}} \int_{D_0} \delta u^0,
\delta u(x, 0) = \delta u^0, & \text{in } x \in \mathbb{R}.
\end{cases} \quad (86)$$

### 6.3 Sensitivity of the functional

In this section we study the sensitivity of the functional $J$ with respect to variations associated with the generalized tangent vectors defined in the previous section. We first define an appropriate generalization of the Gateaux derivative.

**Definition 21** ([7]) Let $J : L^1(\mathbb{R}) \to \mathbb{R}$ be a functional and $u^0 \in L^1(\mathbb{R})$ be Lipschitz continuous with a discontinuity at $x = \varphi^0$, an initial datum for which the solution of (77) satisfies hypothesis (H). We say that $J$ is Gateaux differentiable at $u^0$ in a generalized sense if for any generalized tangent vector $(\delta u^0, \delta \varphi^0)$ and any family $u^{0,\epsilon} \in \Sigma_{u^0}$ associated to $(\delta u^0, \delta \varphi^0)$ the following limit exists

$$\delta J = \lim_{\epsilon \to 0} \frac{J(u^{0,\epsilon}) - J(u^0)}{\epsilon},$$

and it depends only on $(u^0, \varphi^0)$ and $(\delta u^0, \delta \varphi^0)$, i.e. it does not depend on the particular family $u^{0,\epsilon}$ which generates $(\delta u^0, \delta \varphi^0)$.

The limit $\delta J$ is the generalized Gateaux derivative of $J$ in the direction $(\delta u^0, \delta \varphi^0)$.

The following result provides an easy characterization of the generalized Gateaux derivative of $J$ in terms of the solution of the associated adjoint system. A similar result is formally obtained in [13] in the context of the one-dimensional Euler system. In [8] it is shown how this generalization of the Gateaux derivative can be used to obtain some optimality conditions in a similar optimization problem but, as far as we know, it has not been used to develop a complete descent algorithm as we do here.

**Proposition 22** (see [9]) The Gateaux derivative of $J$ can be written as

$$\delta J = \int_{\{x < \varphi^0\} \cup \{x > \varphi^0\}} p(x, 0) \delta u^0(x) \, dx + q(0)[u^0]_{\varphi^0} \delta \varphi^0, \quad (87)$$
where the adjoint state pair \((p, q)\) satisfies the system

\[
\begin{cases}
-\partial_t p - u \partial_x p = 0, & \text{in } Q^- \cup Q^+, \\
[p]_{\Sigma} = 0, \\
q(t) = p(\varphi(t), t), & \text{in } t \in (0, T) \\
q'(t) = 0, & \text{in } t \in (0, T) \\
p(x, T) = u(x, T) - u^d, & \text{in } \{x < \varphi(T)\} \cup \{x > \varphi(T)\} \\
q(T) = \frac{1}{2}[(u(x, T) - u^d)^2]_{\varphi(T)}. 
\end{cases}
\] (88)

**Remark 23** System (90) has a unique solution. In fact, to solve the backwards system (90) we first define the solution \(q\) on the shock \(\Sigma\) from the condition \(q' = 0\), with the final value \(q(T)\) given in (90). This determines the value of \(p\) along the shock. We then propagate this information, together with the datum of \(p\) at time \(t = T\) to both sides of \(\varphi(T)\), by characteristics. As both systems (37) and (90) have the same characteristics, any point \((x, t) \in \mathbb{R} \times (0, T)\) is reached backwards in time by an unique characteristic coming either from the shock \(\Sigma\) or the final data at \((x, T)\) (see Figure 8 where we illustrate this construction in the case of a shock located along a stright line, as it happens to the Riemann problem). The solution obtained this way coincies with the reversible solutios introduced in [3] and [4].

**Remark 24** Note that the second third and fourth equations in (90) come, by duality, from the linearization of the Rankine-Hugoniot condition (82). Besides, they are in fact the conditions that allow us to obtain a unique solution in (90). They are needed to determine the value of \(p\) on \(\Sigma\).

Solutions of (90) can be also obtained as limit of solutions of the transport equation with artificial viscosity depending of a small parameter \(\varepsilon \to 0\),

\[
\begin{cases}
-\partial_t p - u \partial_x p = \varepsilon \partial_x p, & \text{in } x \in \mathbb{R}, t \in (0, T), \\
p(x, T) = p^T_n(x), & \text{in } x \in \mathbb{R}, 
\end{cases}
\] (89)

and a suitable choice of the initial data \(p^T_n(x)\), depending on \(n \to \infty\). To be more precise, let \(p^T_n(x)\) be any sequence of Lipschitz continuous functions, uniformly bounded in \(BV_{loc}(\mathbb{R})\), such that

\[
p^T_n(x, T) \to p^T(x) = u(x, T) - u^d(x), \quad \text{in } L^1_{loc}(\mathbb{R}),
\]
Figure 8: Characteristic lines entering on a shock and how they may be used to build the solution of the adjoint system both away from the shock and on its region of influence.

\[ p_T^T(\varphi(T), T) = \frac{1}{2} \left[ \left( u(x, T) - u^d \right)^2 \right]_{\varphi(T)}. \]

We first take the limit of the solutions \( p_{\varepsilon,n} \) of (89) as \( \varepsilon \to 0 \), to obtain the solution \( p_n \) of

\[
\begin{cases}
  -\partial_t p - u \partial_x p = 0, & \text{in } x \in \mathbb{R}, t \in (0, T), \\
  p(x, T) = p_T^T(x), & \text{in } x \in \mathbb{R},
\end{cases}
\]

the so called reversible solution (see [3]). These solutions can be characterized by the fact that they take the value \( p_n(\varphi(T), T) \) in the whole region occupied by the characteristics that meet the shock (see [3], Th. 4.1.12). Thus, in particular they satisfy the 2nd, 3rd, 4th and 6th equations in (90). Moreover, \( p_n \to p \) as \( n \to \infty \), and \( p \) takes a constant value in the region occupied by the characteristics that meet the shock. Note that, by construction, this constant is the same value for all \( p_n \) in this region. Thus, this limit solution \( p \) coincides with the solution of (90) constructed above. This allows in fact extending the notion of reversible solutions in [3] to data that, on the point \( x(T) \) are completely disconnected with the values of \( p \) to both sides of it. This is precisely due to the point of view we have adopted in which the linearized
state has two different components \((\delta u, \delta \varphi)\) so that the adjoint state has also two components \((p, q)\) with different initial data at time \(t = T\).

**Remark 25** In the expression \((87)\) for the derivative of \(J\) the shock of the initial datum \(u^0\) appears. When \(u^0\) does not present a shock, obviously, this term cancels in \(\delta J\). It is however important to note that, this is compatible with the possible appearance of shocks for times \(\tau \in (0, T)\). In that case this singular term does not affect the value of \(\delta J\) apparently but in practice it does. Indeed, in this case the adjoint system has to be written in the form \((90)\) for \(\tau < t < T\) and later extended to the time interval \((0, \tau)\) as the classical adjoint system \((75)\). Thus, the presence of the shock does affect the value of the adjoint state \(p\) at the initial time \(t = 0\) and consequently, also, the value of \(\delta J\).

The adjoint system in this case is obtained from \((86)\), as in the proof of Proposition 22 below, and it is given by

\[
egin{aligned}
-\partial_t p - u \partial_x p &= 0, \quad \text{in} \quad (x,t) \in \mathbb{R} \times (0, T) \setminus \Sigma, \\
[p]_\Sigma &= 0, \\
q(t) &= p(\varphi(t), t), \quad \text{in} \quad t \in (\tau, T) \\
q'(t) &= 0, \quad \text{in} \quad t \in (\tau, T) \\
p(x, T) &= u(x, T) - u^d, \quad \text{in} \quad \{x < \varphi(T)\} \cup \{x > \varphi(T)\} \\
q(T) &= \frac{\frac{1}{2}[(u(x, T) - u^d)^2]_{\varphi(T)}.}
\end{aligned}
\]

Let us briefly comment the result of Proposition 22 before giving its proof. Formula \((87)\) provides an obvious way to compute a first descent direction of \(J\) at \(u^0\). We just take

\[
(\delta u^0, \delta \varphi^0) = (-p(x, 0), -q(0)[u]_{\varphi^0}).
\]

Here, the value of \(\delta \varphi^0\) must be interpreted as the optimal infinitesimal displacement of the discontinuity of \(u^0\).

However, it is important to underline that this \((\delta u^0, \delta \varphi^0)\) is not a generalized tangent vector in \(T_{u^0}\) since \(p(x, 0)\) is not continuous away from \(x \neq \varphi^0\). In fact, \(p(x, t)\) takes the constant value \(q(T)\) in the whole triangular region occupied by the characteristics of \((37)\) which meet the shock \(\Sigma\). Thus, \(p\) has, in general, two discontinuities at the boundary of this region and so will have \(p(x, 0)\) (see Figure 9).

This is an important drawback in developing a descent algorithm for \(J\). Indeed, according to the Definition 17 if \((\delta u^0, \delta \varphi^0)\) is a descent direction
belonging to \( T_{\varphi^0} \), the new datum \( u^{0,\text{new}} \) should be obtained from \( u^0 \) following a path associated to this descent direction

\[
u^{0,\text{new}} = \begin{cases} u^0 + \varepsilon \delta u^0 + [u^0]_{\varphi^0} \chi_{[\varphi^0+\varepsilon \delta \varphi^0,\varphi^0]} & \text{if } \delta \varphi^0 < 0, \\ u^0 + \varepsilon \delta u^0 - [u^0]_{\varphi^0} \chi_{[\varphi^0,\varphi^0+\varepsilon \delta \varphi^0]} & \text{if } \delta \varphi^0 > 0, \end{cases}
\]  

(92)

for some \( \varepsilon > 0 \) small enough, correctly chosen. Note that, if we take (91) as descent direction \((\delta u^0, \delta \varphi^0)\), which is not a generalized tangent vector as explained above, the new datum \( u^{0,\text{new}} \) will have three discontinuities; the one coming from the displacement of the discontinuity of \( u^0 \) at \( \varphi^0 \) and two more produced by the discontinuities of \( p(x, 0) \). Thus, in an iterative process, the descent algorithm will create more and more discontinuities increasing artificially the complexity of solutions. This motivates the alternating descent method we propose that, based on this notion of generalized gradients, develops a descent algorithm that keeps the complexity of solutions bounded. This will be done in the following Section.

We finish this section with the proof of the Proposition 22.

**Proof.** (Proposition 22) A straightforward computation shows that \( J \) is Gateaux differentiable in the sense of Definition 21 and that, the generalized Gateaux derivative of \( J \) in the direction of the generalized tangent vector

Figure 9: Solution \( u(x, t) \) of the Burgers equation with an initial datum having a discontinuity (left) and adjoint solution which takes a constant value in the region occupied by the characteristics that meet the shock (right).
$(\delta u^0, \delta \phi^0)$ is given by

$$
\delta J = \int_{\{x<\phi(T)\} \cup \{x>\phi(T)\}} (u(x, T) - u^d(x)) \delta u(x, T) - \left[ \frac{(u(x, T) - u^d(x))^2}{2} \right]_{\phi(T)} \delta \phi(T),
$$

where the pair $(\delta u, \delta \phi)$ solves the linearized problem (85) with initial data $(\delta u^0, \delta \phi^0)$.

We now introduce the adjoint system (90). Multiplying the equations of $\delta u$ by $p$ and integrating we obtain

$$
0 = \int_{Q^- \cup Q^+} (\partial_t \delta u + \partial_x (u \delta u)) p \, dx dt - \int_{Q^- \cup Q^+} (\partial_t p + u \partial_x p) \delta u \, dx dt
$$

$$
+ \int_{\{x<\phi(T)\} \cup \{x>\phi(T)\}} \delta u(x, T) p(x, T) \, dx - \int_{\{x<\phi^0\} \cup \{x>\phi^0\}} \delta u^0(x) p(x, 0) \, dx
$$

$$
- \int_{\Sigma} ([u \delta u]_{\Sigma} n_t + [u \delta u]_{\Sigma} n_x) \, d\Sigma,
$$

where $(n_x, n_t)$ are the cartesian components of the normal vector to the curve $\Sigma$.

Therefore,

$$
\delta J = \int_{\{x<\phi(T)\} \cup \{x>\phi(T)\}} \delta u(x, T) (u(x, T) - u^d(x)) \, dx - \left[ \frac{(u(x, T) - u^d(x))^2}{2} \right]_{\phi(T)} \delta \phi(T)
$$

$$
= \int_{\{x<\phi^0\} \cup \{x>\phi^0\}} \delta u^0(x) p(x, 0) \, dx + \int_{\Sigma} ([u \delta u]_{\Sigma} n_t + [u \delta u]_{\Sigma} n_x) \, d\Sigma
$$

$$
- \left[ \frac{(u(x, T) - u^d(x))^2}{2} \right]_{\phi(T)} \delta \phi(T),
$$

(95)

Assume, for the moment, that the following identity holds:

$$
\int_{\Sigma} ([u \delta u]_{\Sigma} n_t + [u \delta u]_{\Sigma} n_x) \, d\Sigma = \int_{\Sigma} [p]_{\Sigma} (\delta u n_t + u \delta u n_x) \, d\Sigma
$$

$$
- \int_{\Sigma} \partial_t \bar{g} (\delta \phi(t)|u|_{\phi(t)}) \, d\Sigma + \bar{p}^T (\phi(T)) \delta \phi(T) [u]_{\phi(T)} - \bar{p}(\phi(0), 0) \delta \phi^0 [u](\Sigma)
$$

Here $\bar{g}$ represents the average of $g$ to both sides of the shock $\Sigma$, i.e.

$$
\bar{g}(x) = \frac{1}{2} \left( \lim_{\varepsilon \to 0} g(x + \varepsilon n_{\Sigma}) + \lim_{\varepsilon \to 0} g(x - \varepsilon n_{\Sigma}) \right), \quad x \in \Sigma.
$$

45
Then, substituting (96) in (95) and taken into account the final condition on \((p, q)\) at \(t = T\) in (90), we obtain

\[
\delta J = \int R \delta u^0(x)p(x, 0) \, dx + \int_{\Sigma} [p]_{\Sigma} \left( \delta u n_t + u \delta u n_x \right) \, d\Sigma \\
- \int_{\Sigma} \partial_t \bar{p} \left( \delta \varphi(t)[u]_{\varphi(t)} \right) \, d\Sigma - \bar{p}(\varphi(0), 0) \delta \varphi^0[u]_{\varphi(0)}.
\]

To obtain formula (87), the second and third terms in this expression must vanish. This is the case if \((p, q)\) satisfies (90). This concludes the proof of Proposition 22.

Let us now prove formula (96). Using the elementary identity

\[
[f g]_{\Sigma} = [f]_{\Sigma} g + [g]_{\Sigma} f, \tag{97}
\]

we get

\[
\int_{\Sigma} ([\delta u]_{\Sigma} n_t + [u \delta u]_{\Sigma} n_x) \, d\Sigma = \int_{\Sigma} [p]_{\Sigma} \left( \delta u n_t + u \delta u n_x \right) \, d\Sigma + \int_{\Sigma} \bar{p} \left( [\delta u]_{\Sigma} n_t + [u \delta u]_{\Sigma} n_x \right) \, d\Sigma,
\]

and we obtain the first term in the identity (96). We now simplify the second term:

\[
\int_{\Sigma} \bar{p} \left( [\delta u]_{\Sigma} n_t + [u \delta u]_{\Sigma} n_x \right).
\]

The cartesian components of the normal vector to \(\Sigma\) are given by

\[
n_t = \frac{-\varphi'(t)}{\sqrt{1 + (\varphi'(t))^2}}, \quad n_x = \frac{1}{\sqrt{1 + (\varphi'(t))^2}}.
\]

Therefore, taking into account the second equation in system (85),

\[
[\delta u]_{\Sigma} n_t + [u \delta u]_{\Sigma} n_x = \frac{-\varphi'(t)[\delta u]_{\Sigma} + [u \delta u]_{\Sigma}}{\sqrt{1 + (\varphi'(t))^2}} \tag{98}
\]

\[
= \frac{\delta \varphi'(t)[u]_{\varphi(t)} + \delta \varphi(t) \left( \varphi'(t)[u]_{\varphi(t)} - \left[ \partial_x (u^2/2) \right]_{\varphi(t)} \right)}{\sqrt{1 + (\varphi'(t))^2}}
\]

\[
= \frac{\delta \varphi'(t)[u]_{\varphi(t)} + \delta \varphi(t) \left[ \frac{\partial}{\partial t} \varphi(t), t \right]_{\varphi(t)}}{\sqrt{1 + (\varphi'(t))^2}} = \partial_t \left( \delta \varphi(t)[u]_{\varphi(t)} \right).
\]
Finally,
\[
\int_{\Sigma} \bar{p} \left( \left[ \delta u \right]_{\Sigma} n_t + \left[ u \delta u \right]_{\Sigma} n_x \right) \, d\Sigma = \int_{\Sigma} \bar{p} \partial_{tg} \left( \delta \varphi(t)[u]_{\varphi(t)} \right) \, d\Sigma
\]
\[
= - \int_{\Sigma} \partial_{tg} \bar{p} \left( \delta \varphi(t)[u]_{\varphi(t)} \right) + \bar{p}^T(\varphi(T))\delta \varphi(T)[u]_{\varphi(T)} \, d\Sigma - \bar{p}(\varphi(0), 0)\delta \varphi^0[u]_{\varphi(0)}.
\]

6.4 The discrete approach: Differentiable numerical schemes

Computing the gradient of the discrete functional $J^\Delta$ requires computing one derivative of $J^\Delta$ with respect to each node of the mesh. This can be done in a cheaper way using the adjoint state. We illustrate it on two different numerical schemes: Lax-Friedrichs and Engquist-Osher. Note that both schemes satisfy the hypothesis of Theorem 6 and therefore the numerical minimizers are good approximations of minimizers of the continuous problem. However, as the discrete functionals $J^\Delta$ are not necessarily convex the gradient methods could possibly provide sequences that do not converge to a global minimizer of $J^\Delta$. But this drawback and difficulty appears in most applications of descent methods in optimal design and control problems. As we will see, in the present context, the approximations obtained by gradient methods are satisfactory, although convergence is slow due to unnecessary oscillations that the descent method introduces.

Computing the gradient of $J^\Delta$, rigorously speaking, requires the numerical scheme (45) under consideration to be differentiable and, often, this is not the case. To be more precise, for the Burgers equation (37) we can choose several efficient methods which are differentiable (as the Lax-Friedrichs and the Engquist-Osher one) but this is not the situation for general systems of conservation laws in higher dimensions, as Euler equations. For such complex systems the efficient methods, as Godunov, Roe, etc., are not differentiable (see, for example [18] or [22]) thus making the approach in this section useless.

We observe that when the 3-point conservative numerical approximation scheme (45) used to approximate the Burgers equation (37) has a differentiable numerical flux function $g$, then the linearization is easy to compute.
We obtain
\[
\begin{align*}
\delta u_{j}^{n+1} &= \delta u_{j}^{n} - \lambda \left( \partial_{1} g_{j+1/2}^{n} \delta u_{j}^{n} + \partial_{2} g_{j+1}^{n} \delta u_{j+1}^{n} - \partial_{1} g_{j-1/2}^{n} \delta u_{j-1}^{n} - \partial_{2} g_{j-1}^{n} \delta u_{j}^{n} \right) = 0, \\
& \quad j \in \mathbb{Z}, \quad n = 0, \ldots, N. \\
\end{align*}
\]  \tag{99}

In view of this, the discrete adjoint system can also be written for any differentiable flux function \( g \):
\[
\begin{align*}
p_{j}^{n} &= p_{j}^{n+1} + \lambda \left( \partial_{1} g_{j+1/2}^{n} (p_{j+1}^{n+1} - p_{j}^{n+1}) + \partial_{2} g_{j+1}^{n} (p_{j}^{n+1} - p_{j-1}^{n+1}) \right), \\
p_{j}^{N+1} &= p_{j}^{T}, \quad j \in \mathbb{Z}, \quad n = 0, \ldots, N. \\
\end{align*}
\]  \tag{100}

In fact, when multiplying the equations in (99) by \( p_{j}^{n+1} \) and adding in \( j \in \mathbb{Z} \) and \( n = 0, \ldots, N \), the following identity is easily obtained,
\[
\Delta x \sum_{j \in \mathbb{Z}} p_{j}^{T} \delta u_{j}^{N+1} = \Delta x \sum_{j \in \mathbb{Z}} p_{0}^{0} \delta u_{j}^{0}. 
\]  \tag{101}

This is the discrete version of formula (76) which allows us to simplify the derivative of the discrete cost functional.

Thus, for any variation \( \delta u_{0}^{\Delta} \in \mathcal{U}^{\Delta} \) of \( u_{0}^{\Delta} \), the Gateaux derivative of the cost functional defined in (43) is given by
\[
\delta J^{\Delta} = \Delta x \sum_{j \in \mathbb{Z}} (u_{j}^{N+1} - u_{j}^{d}) \delta u_{j}^{N+1}, 
\]  \tag{102}

where \( \delta u_{j}^{0} \) solves the linearized system (99). If we consider \( p_{j}^{n} \) the solution of (100) with final datum
\[
p_{j}^{T} = u_{j}^{N+1} - u_{j}^{d}, \quad j \in \mathbb{Z}, 
\]  \tag{103}

then \( \delta J^{\Delta} \) in (102) can be written as,
\[
\delta J^{\Delta} = \Delta x \sum_{j \in \mathbb{Z}} p_{j}^{0} \delta u_{j}^{0}, 
\]  \tag{104}

and this allows to obtain easily the steepest descent direction for \( J^{\Delta} \) by considering
\[
\delta u_{0}^{\Delta} = -p_{0}^{\Delta}. 
\]  \tag{105}
We now present two particular examples. Let us consider first the Lax-Friedrichs scheme:

\[
\begin{align*}
&\left\{ \frac{u_j^{n+1} - u_{j+1}^n}{\Delta t} + \frac{f(u_j^{n+1}) - f(u_{j-1}^n)}{2\Delta x} = 0, \quad n = 0, ..., N, \\ &u_0^j = u_{0,j}, \quad j \in \mathbb{Z}, \right. \\
\end{align*}
\]

(106)

where \( f(s) = s^2 / 2 \). The numerical scheme (106) can be written in conservation form with the numerical flux given in (48). Moreover, it satisfies the hypotheses of Theorem 6, under the Courant-Friedrichs-Levy (CFL) condition \( \lambda \max |u^0| \leq 1 \), and it is differentiable.

For any variation \( \delta u_\Delta^0 \in \mathcal{U}_\Delta \) of \( u_\Delta^0 \), the Gateaux derivative of the cost functional is given by (104) where the values \( p_j^n \) satisfy the adjoint system,

\[
\begin{align*}
&\left\{ \frac{p_j^n - p_{j+1}^{n+1} + p_{j-1}^{n+1}}{\Delta t} + u_j^n \frac{p_{j+1}^{n+1} - p_{j+1}^{n}}{2\Delta x} = 0, \quad n = 0, ..., N, \\ &p_N^{n+1} = p_T^j, \quad j \in \mathbb{Z}, \right. \\
\end{align*}
\]

(107)

with \( p_j^T = (u_j^{N+1} - u_j^d) \in \mathcal{U}_{ad}^\Delta \).

Note that, formally, (107) is in fact the Lax-Friedrichs numerical scheme applied to the continuous adjoint system (75).

The Engquist-Osher scheme can be treated similarly. In this case the numerical flux is given by (49) and we get the adjoint system

\[
\begin{align*}
&\left\{ p_j^n = p_{j+1}^{n+1} + \lambda \left( \frac{u_j^n + |u_j^n|}{2} (p_{j+1}^{n+1} - p_{j+1}^{n}) + \frac{u_j^n - |u_j^n|}{2} (p_{j+1}^{n+1} - p_{j-1}^{n+1}) \right), \quad n = 0, ..., N, \\ &p_N^{n+1} = p_T^j, \quad j \in \mathbb{Z}. \right. \\
\end{align*}
\]

(108)

The derivative \( \delta J^\Delta \) is given again by (104) and the steepest descent direction is (105) where, now, \( p \) solves (108).

We observe that (108) is the upwind method for the continuous adjoint system. Thus, this is another case in which the adjoint of the discretization corresponds to a well-known discretization of the adjoint problem.

**Remark 26** We do not address here the problem of the convergence of these adjoint schemes towards the solutions of the continuous adjoint system. Of course, this is an easy matter when \( u \) is smooth but is is far from being trivial when \( u \) has shock discontinuities. Whether or not these discrete adjoint systems, as \( \Delta \to 0 \), allow reconstructing the complete adjoint system, with the inner Dirichlet condition along the shock, constitutes an interesting problem for future research. We refer to [19] for preliminary work on this direction.
6.5 The discrete approach: Non-differentiable numerical schemes

We describe here the most common method to compute “gradients” of functionals when the underlying numerical scheme used to approximate the flow equations is non-differentiable (see for example [15] where this method is used in the context of linearized stability). To illustrate this method we focus on the Roe scheme which is one of the most popular ones to approximate solutions of conservation laws.

In the particular case of the Burgers equation under consideration Roe’s scheme coincides with Godunov’s one.

However, Roe’s scheme is not monotone for general fluxes \( f \) and it is well-known that this scheme admits entropy violating discontinuities (see [16]). Therefore, convergence of discrete minimizers towards continuous ones can not be guaranteed for more general fluxes.

The scheme can be modified to obtain the conservation of entropy (see [16] for the Harten and Hyman modification) but we will not consider this modification here (which is still non-differentiable) since we are mainly interested in the issue of “linearizing” non-differentiable schemes.

The Roe scheme for a general conservation law

\[
\partial_t u + \partial_x f(u) = 0,
\]

is a 3-point conservative scheme of the form (45) with numerical flux

\[
g^R(u, v) = \frac{1}{2}(f(u) + f(v) - |A(u, v)|(v - u)),
\]

where the matrix \( A(u, v) \) is a Roe linearization which is an approximation of \( f' \) (see, for example, [15]). In the scalar case under consideration \( f(u) = u^2/2 \) and

\[
A(u, v) = \begin{cases}
\frac{f(u) - f(v)}{u - v} = \frac{u + v}{2}, & \text{if } u \neq v, \\
f'(u) = u, & \text{if } u = v.
\end{cases}
\]

Note that the previous scheme is not differentiable, in general, due to the presence of the absolute value of \( A \) in \( g^R \). Thus, we cannot linearize this system and obtain its adjoint, in a rigorous sense.

In [15] the following scheme is proposed for the linearization

\[
\delta u_j^{n+1} = \delta u_j^n - \lambda(h_{j+1/2}^n - h_{j-1/2}^n), \quad j \in \mathbb{Z}, \quad n = 0, ..., N, \quad (109)
\]
where

\[ h_{j+1/2}^n = h(u_j^n, u_{j+1}^n; \delta u_j^n, \delta u_{j+1}^n), \]
\[ h(u, v; w, z) = \frac{1}{2} (A(u, v)(w + z) - |A(u, v)|(z - w)). \]  

Equation (110) is in fact an approximation of the natural choice

\[ h(u, v; w, z) = \frac{\partial g}{\partial u}^R + \frac{\partial g}{\partial v}, \]

where \( \frac{\partial f}{\partial u} \) is approximated by the Roe linearization \( A(u, v) \), and the non-differentiable term \( |A(u, v)| \) in (109) is assumed to have zero derivative. This last assumption could be formally interpreted as a particular choice of the subgradient of the absolute value function \( a(x) = |x| \) at \( x = 0 \).

In this way

\[ h_{j+1/2}^n = \frac{1}{2} (A_{j+1/2}(\delta u_j^n + \delta u_{j+1}^n) - |A_{j+1/2}|(\delta u_{j+1}^n - \delta u_j^n)), \]
\[ A_{j+1/2} = A(u_j^n, u_{j+1}^n). \]

The corresponding adjoint system to the linearized equations (109) is given by

\[
\begin{aligned}
    p_j^n &= p_j^{n+1} + \lambda (\alpha_j^n (p_{j+1}^{n+1} - p_{j+1}^n) + \beta_j^n (p_{j-1}^{n+1} - p_{j-1}^n)) , \\
    p_j^{N+1} &= p_j^T, \quad j \in \mathbb{Z}, \\
    p_j^0 &= p_j^T, \quad j \in \mathbb{Z},
\end{aligned}
\]

where

\[ \alpha_j^n = \frac{1}{2} (A_{j+1/2} + |A_{j+1/2}|), \quad \beta_j^n = \frac{1}{2} (A_{j+1/2} - |A_{j+1/2}|). \]

In fact, multiplying the equations in (109) by \( p_j^{n+1} \) and adding in \( j \in \mathbb{Z} \) and \( n = 0, ..., N \) we obtain:

\[
0 = \sum_{j \in \mathbb{Z}} \sum_{n=0}^N (\delta u_j^{n+1} - \delta u_j^n + \lambda (h_{j+1/2}^n - h_{j-1/2}^n)) p_j^{n+1}
\]
\[
= \sum_{j \in \mathbb{Z}} \sum_{n=0}^N (p_j^n - p_j^{n+1} - \lambda [\alpha_j^n (p_{j+1}^{n+1} - p_{j+1}^n) + \beta_j^n (p_{j-1}^{n+1} - p_{j-1}^n)]) \delta u_j^n
\]
\[
+ \sum_{j \in \mathbb{Z}} \delta u_j^{N+1} p_j^{N+1} - \sum_{j \in \mathbb{Z}} \delta u_j^0 p_j^0 = \sum_{j \in \mathbb{Z}} \delta u_j^{N+1} p_j^{N+1} - \sum_{j \in \mathbb{Z}} \delta u_j^0 p_j^0. \]  

(112)
To obtain (112) we have used the following identity:

\[
\sum_{j \in \mathbb{Z}} h_{j+1/2}^{n+1} p_j^{n+1} = \sum_{j \in \mathbb{Z}} \frac{1}{2} A_{j+1/2} (\delta u_j^n + \delta u_{j+1}^n) p_j^{n+1} \\
- \sum_{j \in \mathbb{Z}} \frac{1}{2} |A_{j+1/2}| (\delta u^n_{j+1} - \delta u^n_j) p_j^{n+1} \\
= \sum_{j \in \mathbb{Z}} \frac{1}{2} (A_{j+1/2} p_j^{n+1} + A_{j-1/2} p_{j-1}^{n+1}) \delta u^n_j \\
- \sum_{j \in \mathbb{Z}} \frac{1}{2} (|A_{j-1/2}| p_{j-1}^{n+1} - |A_{j+1/2}| p_j^{n+1}) \delta u^n_j,
\]

and an analogous one for the term \(\sum_{j \in \mathbb{Z}} h_{j-1/2}^{n+1} p_j^{n+1}\).

Then, as for differentiable schemes, formula (112) allows to simplify the “derivative” \(\delta J^\Delta\) which is formally written as (104). Thus, a tentative descent direction for \(J^\Delta\) is given by (105), where \(p_j^n\) is the solution of the adjoint system (111) with final datum \(p_{N+1}^n = u_{N+1}^n - u_d^n\).

The above computation does not provide the gradient of the discrete functional, which is non-differentiable in this case. But the value obtained through this computation could be used as an alternative “descent direction” in a gradient-type algorithm.

Note that the approach of using “pseudogradients” we have presented here is a common practice in optimal design in aeronautics where efficient solvers are often non-differentiable (see [26]).

**References**


[4] Bouchut F. and James F., Differentiability with respect to initial data for
a scalar conservation law, Proceedings of the 7th Int. Conf. on Hyperbolic

for multi-dimensional transport equations with one-sided Lipschitz coef-

[6] Brenier, Y. and Osher, S. The Discrete One-Sided Lipschitz Condition for
Convex Scalar Conservation Laws, SIAM Journal on Numerical Analysis,

[7] Bressan A. and Marson A., A variational calculus for discontinuous so-
lutions of systems of conservation laws, Commun. in Partial Differential

79-94

the optimal control of the inviscid Burgers equation in the presence of

[10] Dal Maso G., Le Floch Ph. and Murat F., Definition and weak stability

and source-type solutions for a diffusion-convection equation, Arch. Rat.

[12] Garreau, S., Guillaume, Ph. and Masmoudi, M. The topological asympto-
tic for PDE systems: the elasticity case. SIAM J. Control Optim. 39
(6) (2001), 1756–177

[13] Giles M.B. and Pierce N.A., Analytic adjoint solutions for the quasi one-


