CONVERGENCE RATES OF SYMPLECTIC PONTRYAGIN APPROXIMATIONS IN OPTIMAL CONTROL THEORY

MATTIAS SANDBERG AND ANDERS SZEPESSY

ABSTRACT. Many inverse problems for differential equations can be formulated as optimal control problems. It is well known that inverse problems often need to be regularized to obtain good approximations. This work presents a systematic method to regularize and to establish error estimates for approximations to some control problems in high dimension, based on symplectic approximation of the Hamiltonian system for the control problem. In particular the work derives error estimates and constructs regularizations for numerical approximations to optimally controlled ordinary differential equations in $\mathbb{R}^d$, with non smooth control. Though optimal controls in general become non smooth, viscosity solutions to the corresponding Hamilton-Jacobi-Bellman equation provide good theoretical foundation, but poor computational efficiency in high dimensions. The computational method here uses the adjoint variable and works efficiently also for high dimensional problems with $d \gg 1$. Controls can be discontinuous due to a lack of regularity in the Hamiltonian or due to colliding backward paths, i.e. shocks. The error analysis, for both these cases, is based on consistency with the Hamilton-Jacobi-Bellman equation, in the viscosity solution sense, and a discrete Pontryagin principle: the bi-characteristic Hamiltonian ODE system is solved with a $\mathcal{C}^2$ approximate Hamiltonian. The error analysis leads to estimates useful also in high dimensions since the bounds depend on the Lipschitz norms of the Hamiltonian and the gradient of the value function but not on d explicitly. Applications to inverse implied volatility estimation, in mathematical finance, and to an optimal design problem are presented. An advantage with the Pontryagin based method is that the Newton method can be applied to efficiently solve the discrete nonlinear Hamiltonian system, with a sparse Jacobian that can be calculated explicitly.

CONTENTS

1991 *Mathematics Subject Classification.* 49L25, 65M25.

Key words and phrases. optimal control, Hamilton-Jacobi, Hamiltonian system, Pontryagin principle, optimal design, implied volatility.

Support by the Swedish Research Council grants 2002-6285 and 2002-4961 and the European network HYKE, funded by the EC as contract HPRN-CT-2002-00282, is acknowledged.

1. The Optimal Control Problem

Many inverse problems for differential equations can be formulated as optimal control problems. It is well known that inverse problems often need to be regularized to obtain good approximations, cf. [12], [28]. This work presents a systematic method to regularize and to establish error estimates for approximations to some control problems in high dimension, based on symplectic approximation of the Hamiltonian system for the control problem. We derive error estimates and construct regularizations for numerical approximations of the optimal control problem to solve an ordinary differential equation

$$\begin{aligned} X'(s) &= f\big(X(s),\alpha(s)\big), \quad 0<s<T \\ X(0) &= X_0. \end{aligned} \tag{1.1}$$

with the *state* variable $X:[0,T]\to\mathbb{R}^d$ and the *control* $\alpha:[0,T]\to B\subset\mathbb{R}^m$, which is adjusted to minimize the functional

$$g\big(X(T)\big)+\int_0^T h\big(X(s),\alpha(s)\big)ds, \tag{1.2}$$

for given data $X_0\in\mathbb{R}^d, f:\mathbb{R}^d\times B\to\mathbb{R}^d, h:\mathbb{R}^d\times B\to\mathbb{R}$ and $g:\mathbb{R}^d\to\mathbb{R}$. Our study focuses on non smooth controls and approximation methods with error analysis useful also for high dimensional problems, $d\gg 1$, e.g. discretizations of controlled partial differential equations. We assume $\alpha\in\mathcal{A}$, where

$$\mathcal{A}=\{\alpha:[0,T]\to B \mid B\subset\mathbb{R}^m,\ \alpha(\cdot) \text{ is measurable}\}$$

is the set of *admissible controls* and B is compact.

The alternative to solve the minimum of (1.2), with the constraint (1.1), is to define the *value function*,

$$u(x,t):=\inf_{X(t)=x,\ \alpha\in\mathcal{A}}\left(g\big(X(T)\big)+\int_t^T h\big(X(s),\alpha(s)\big)ds\right), \tag{1.3}$$

and use dynamic programming: the value function, $u:\mathbb{R}^d\times[0,T]\to\mathbb{R}$, is the bounded uniformly continuous viscosity solution of the *Hamilton–Jacobi–Bellman equation*, cf. [13],

$$\begin{aligned} u_t(x,t)+H(u_x(x,t),x)&=0, \quad (x,t)\in\mathbb{R}^d\times(0,T), \\ u(x,T)&=g(x), \quad x\in\mathbb{R}^d, \end{aligned} \tag{1.4}$$

where u_x denotes the gradient of u with respect to the spatial variable $x\in\mathbb{R}^d$, $u_t=\partial u/\partial t$ is the time derivative and the *Hamiltonian function* $H:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ is defined by

$$H(\lambda,x)=\min_{\alpha\in B}\big(\lambda\cdot f(x,\alpha)+h(x,\alpha)\big). \tag{1.5}$$

Assume temporarily that the Hamiltonian is differentiable. If $u_x(X(t),t) \in \mathcal{C}^1$ along an optimal path $X(t)$, differentiation of (1.4) shows that the pair $\lambda(t) \equiv u_x\big(X(t),t\big)$ and $X(t)$ satisfies the bi-characteristic *Hamiltonian* boundary value *system* (cf. [14])

$$\begin{aligned} X'(s) &= H_\lambda\big(\lambda(s), X(s)\big), \quad X(0) = X_0, \quad 0 < s < T,\\ -\lambda'(s) &= H_x\big(\lambda(s), X(s)\big), \quad \lambda(T) = g_x\big(X(T)\big), \quad 0 < s < T, \end{aligned} \tag{1.6}$$

where H_λ and H_x denote the gradients of H with respect to the λ-variable and x-variable, respectively. A control $\hat{\alpha}\big(\lambda(t), X(t)\big)$ determined by the *Pontryagin principle*

$$\hat{\alpha} \in \operatorname*{argmin}_{a\in B}\big(\lambda(t)\cdot f(X(t),a) + h(X(t),a)\big) \tag{1.7}$$

satisfies for differentiable H, f and h

$$\begin{aligned} f\big(x,\alpha(\lambda,x)\big) &= H_\lambda(\lambda,x),\\ h\big(x,\alpha(\lambda,x)\big) &= H(\lambda,x) - \lambda H_\lambda(\lambda,x),\\ \lambda f_x\big(x,\alpha(\lambda,x)\big) + h_x\big(x,\alpha(\lambda,x)\big) &= H_x(\lambda,x), \end{aligned} \tag{1.8}$$

see [5] and Remark 3.1, which justifies to denote (1.6) the *Pontryagin method*, cf. [22]. The work [3] verifies that the Pontryagin principle holds for (1.6), based on (1.8), by using that the value function is a viscosity solution of the Hamilton-Jacobi equation (1.4).

We consider numerical approximations with the time steps

$$t_n = \frac{n}{N}T, \quad n = 0,1,2,\dots,N.$$

The most basic approximation is based on the minimization

$$\min_{\bar{\alpha}\in B^N}\big(g(\bar{X}_N) + \sum_{n=0}^{N-1} h(\bar{X}_n,\bar{\alpha}_n)\Delta t\big), \tag{1.9}$$

where $\Delta t = t_{n+1} - t_n$, $\bar{X}_0 = X_0$ and $\bar{X}_n \equiv \bar{X}(t_n)$, for $1 \le n \le N$, satisfy the *forward Euler* constraint

$$\bar{X}_{n+1} = \bar{X}_n + \Delta t\, f(\bar{X}_n,\bar{\alpha}_n). \tag{1.10}$$

The existence of at least one minimum of (1.9) is clear since it is a minimization of a continuous function in the compact set B^N. We will focus on an alternative method based on the *discrete Pontryagin principle*

$$\begin{aligned} \bar{X}_{n+1} &= \bar{X}_n + \Delta t H_\lambda\big(\bar{\lambda}_{n+1},\bar{X}_n\big), \quad \bar{X}_0 = X_0,\\ \bar{\lambda}_n &= \bar{\lambda}_{n+1} + \Delta t H_x\big(\bar{\lambda}_{n+1},\bar{X}_n\big), \quad \bar{\lambda}_N = g_x(\bar{X}_N). \end{aligned} \tag{1.11}$$

This method is called the symplectic Euler method for (1.6), cf. [17]. Section 4.4 presents simple generalizations to general higher order accurate symplectic approximations.

The theory of viscosity solutions for Hamilton-Jacobi-Bellman partial differential equations provides good theoretical foundation also for non smooth controls. In particular this mathematical theory removes one of Pontryagin's two reasons[1],

[1] citation from chapter one in [22] "This equation of Bellman's yields an approach to the solution of the optimal control problem which is closely connected with, but different from, the approach described in this book (see Chapter 9). It is worth mentioning that the assumption regarding the continuous differentiability of the functional (9.8) [(1.3) in this paper] is not fulfilled

but not the other, to favor the ODE approach (1.6) and (1.8): the mathematical theory of viscosity solutions handles elegantly the inherent non smoothness in control problems; analogous theoretical convergence results for an ODE approach was developed later based on the so called minmax solutions, see [26]. Finite difference or finite element approximations of the nonlinear partial differential equation (1.4) has convergence proofs with minimal assumptions using the viscosity solution theory, cf. [9] [18] [25] [27], and [2] for second order equations. On the other hand, such finite difference or finite element computations only work efficiently in low dimension, since the problem of finding the control of the ordinary differential equation (1.1) in $\mathbb{R}^d$ has been transformed to solving a partial differential equation in $\mathbb{R}^d \times [0, T]$, while the Pontryagin approach based on the solution of the two ordinary differential equations (1.15) with (1.8), for $\bar{X}$ and the adjoint variable $\bar{\lambda}$, is computationally feasible even in very high dimension, $d \gg 1$, with application to distributed control of partial differential equations cf. [15, 16], optimal shape [21] and inverse problems [28, 12], see Section 2. This work uses the theory of viscosity solutions to Hamilton-Jacobi equations to construct regularized flux and cost functions and to prove convergence for the Pontryagin method for non smooth controls.

Our analysis is a kind of backward error analysis. The standard backward error analysis for Hamiltonian systems (1.6) uses an analytic Hamiltonian and shows that symplectic one step schemes generate approximate paths that solve a modified Hamiltonian system, with the perturbed Hamiltonian given by a series expansion cf. [17]. Our backward error analysis is different and more related to the standard finite element analysis. We first extend the approximate Euler solution to a continuous piecewise linear function in time and define a *discrete value function*, $\bar{u} : \mathbb{R}^d \times [0, T] \to \mathbb{R}$. This value function satisfies a perturbed Hamilton-Jacobi partial differential equation, with a small residual error. A special case of our analysis shows that if the optimal $\hat{\alpha}$ in (1.7) is a differentiable function of x and λ and if the optimal backward paths, $\bar{X}(s)$ for $s < T$, do not collide, more about this later, the discrete value functions, $\bar{u}$, for the Pontryagin method (1.11) satisfies a Hamilton-Jacobi equation:

$$\bar{u}_t + H(\bar{u}_x, \cdot) = \mathcal{O}(\Delta t), \qquad \text{as } \Delta t \to 0+, \tag{1.12}$$

where

$$\bar{u}(x, t_m) \equiv \min_{\bar{\alpha} \in B^N} \left(g(\bar{X}_N) + \sum_{n=m}^{N-1} h(\bar{X}_n, \bar{\alpha}_n) \Delta t \right) \tag{1.13}$$

for solutions $\bar{X}$ to with $\bar{X}(t_m) \equiv \bar{X}_m = x$. The minimum in (1.13) is taken over the solutions to the discrete Pontryagin principle (1.6). The maximum principle for solutions of Hamilton–Jacobi equations and a comparison between the two equations (1.4) and (1.12) show that

$$\|u - \bar{u}\|_{\mathcal{C}} = \mathcal{O}(\Delta t). \tag{1.14}$$

---

in even the simplest cases, so that Bellman's arguments yield a good heuristic method rather than a mathematical solution of the problem. The maximum principle, apart from its sound mathematical basis, also has the advantage that it leads to a system of ordinary differential equations, whereas Bellman's approach requires the solution of a partial differential equation."

However, in general the optimal controls $\bar\alpha$ and α in (1.10) and (1.1) are discontinuous functions of x, and $\bar\lambda$ or u_x, respectively, and the backward paths *do* collide. There are two different reasons for discontinuous controls:

- The Hamiltonian is in general only Lipschitz continuous, even if f and h are smooth.
- The optimal backward paths may collide.

The standard error analysis for ordinary differential equations is directly applicable to control problems when the time derivative of the control function is integrable. But general control problems with discontinuous controls require alternative analysis. The first step in our error analysis is to construct regularizations of the functions f and h, based on (1.8) applied to a $\mathcal{C}^2(\mathbb{R}^d \times \mathbb{R}^d)$ approximate Hamiltonian H^δ which is λ-concave and satisfies

$$\|H^\delta - H\|_{\mathcal{C}} = \mathcal{O}(\delta), \quad \text{as } \delta \to 0^+,$$

and to introduce the regularized paths

$$\begin{aligned} \bar X_{n+1} &= \bar X_n + \Delta t H^\delta_\lambda\big(\bar\lambda_{n+1}, \bar X_n\big), \quad \bar X_0 = X_0, \\ \bar\lambda_n &= \bar\lambda_{n+1} + \Delta t H^\delta_x\big(\bar\lambda_{n+1}, \bar X_n\big), \quad \bar\lambda_N = g_x(\bar X_N). \end{aligned} \tag{1.15}$$

We will sometimes use the notation $f^\delta \equiv H^\delta_\lambda$ and $h^\delta \equiv H^\delta - \lambda H^\delta_\lambda$.

The next step is to estimate the residual of the discrete value function in the Hamilton-Jacobi-Bellman equation (1.4). The maximum principle for viscosity solutions and the residual estimate imply then an estimate for the error in the value function. An approximation of the form (1.15) may be viewed as a general symplectic one step method for the Hamiltonian system (1.6), see Section 4.4.

There is a second reason to use Hamiltonians with smooth flux: in practice the nonlinear boundary value problem (1.15) has to be solved by iterations. If the flux is not continuous it seems difficult to construct a convergent iterative method, in any case iterations perform better with smoother solutions. An advantage with the Pontryagin based method is that the Newton method can be applied to efficiently solve the discrete nonlinear Hamiltonian system with a sparse Jacobian that often can be calculated explicitly.

If the optimal discrete backward paths $\bar X(t)$ in (1.15) collide on a codimension one surface Γ in $\mathbb{R}^d \times [0,T]$, the dual variable $\bar\lambda = \bar u_x$ may have a discontinuity at Γ, as a function of x. Theorems 4.1 and 4.5 prove, for $\bar u$ based on the Pontryagin method, that in the viscosity solution sense

$$\bar u_t + H(\bar u_x, \cdot) = \mathcal{O}(\Delta t + \delta + \frac{(\Delta t)^2}{\delta}), \tag{1.16}$$

where the discrete value function, $\bar u$, in (1.13) has been modified to

$$\bar u(x, t_m) = \min_{\bar X_m = x} \Big(g(\bar X_N) + \sum_{n=m}^{N-1} h^\delta(\bar X_n, \bar\lambda_{n+1}) \Delta t \Big). \tag{1.17}$$

The regularizations make the right hand side in (1.16) a Lipschitz continuous function of $\big(\bar\lambda(t), \bar X(t), t\big)$, bounded by $C(\Delta t + \delta + \frac{(\Delta t)^2}{\delta})$ where C depends only on the Lipschitz constants of f, h and $\bar\lambda$. Therefore the maximum principle can be used to prove $\|u - \bar u\|_{\mathcal{C}} = \mathcal{O}(\Delta t)$, for $\delta = \Delta t$. Without the regularization, the corresponding error term to in (1.16) is not well defined, even if $\bar u_x$ is smooth. A similar proof applies to the minimization method for smooth Hamiltonians, see

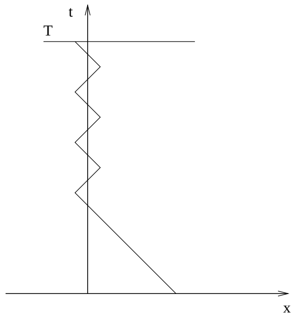

FIGURE 2.1. Example (1a) where the continuous problem has no minimizer among the measurable functions.

[24]. It is important to note that for non smooth control the solution paths $\bar{X}$ may not converge although the value function converges as Δt and δ tend to zero. Therefore our backward error analysis uses consistency with the Hamilton-Jacobi partial differential equation and not with the Hamiltonian system. Convergence if the approximate path $(\bar{X}, \bar{\lambda})$ typically requires Lipschitz continuous flux (H_λ, H_x), which we do not assume in this work.

Outline. Section 2 presents some simple examples how discontinuous control arise and applications to implied volatility estimation and optimal design. The error analysis in this work assumes that the Pontryagin minimization (1.15) has been solved exactly. Section 3 recall standard iterative methods to solve these problems approximately. Section 4 states and proves the error estimates with a non smooth Hamiltonian and colliding paths, for the Pontryagin method. Section 4.4 considers general symplectic one step methods. The remaining sections consider colliding shocks, a maximum principle for Hamilton-Jacobi equations and how to obtain the controls.

2. Optimization Examples

We give some examples when the Hamiltonian, H, is not a differentiable function, and difficulties associated with this.

1a: Let $B = \{-1, 1\}$, $f = \alpha$, $h = x^2/2$ and $g = 0$. Here the continuous minimization problem (1.3) has no minimizer among the measurable functions. A solution in discrete time using a nonregularized Pontryagin method or discrete dynamic programming will behave as in Figure 2.1. First the solution approaches the time axis, and then it oscillates back and forth. As Δt becomes smaller these oscillations do so as well. The infimum for the continuous problem corresponds to a solution $X(t)$ that approaches the time-axis, and then remains on it. However, this corresponds to $\alpha = 0$, which is not in B, and hence the infimum is not attained. A cure to always have an attained minimizing path for the continuous problem is to use controls which are Young measures, see [29] and [19]. We note that the Hamiltonian, $H(\lambda, x) = -|\lambda| + x^2/2$, in this example is not differentiable.

1b: Let $B = [-1, 1]$, $f = \alpha$, $h = x^2/2$ and $g = 0$, which is similar to the previous example but now the set of admissible controls, B, has been changed slightly. Since $0 \in B$, the infimum in (1.3) is now obtained. However, the

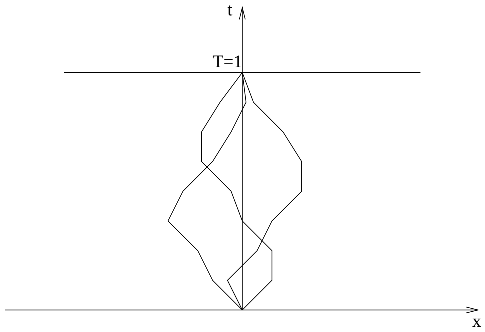

FIGURE 2.2. Example (1c) with $g(x) = x^2$ gives infinitely many minimizing paths through the same starting point.

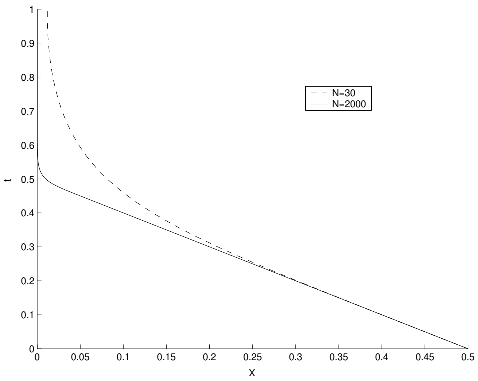

FIGURE 2.3. Solution of the discrete optimization problem (1.15) in Example (1a) and (1b) for $\delta = \Delta t = 1/N$ and $H^\delta_\lambda(\lambda, x) = -\tanh(\lambda/\delta)$, using the Newton method with five iterations.

Hamiltonian remains unchanged compared to the previous example, and a solution to the discrete Pontryagin principle would still be oscillating as in Figure 2.1.

1c: Let $B = [-1, 1]$, $f = \alpha$, $h = 0$ and $g = x^2$. The Hamiltonian is nondifferentiable: $H = -|\lambda|$. If $T = 1$ there are infinitely many solutions to the continuous minimization, the discrete minimization and the unregularized discrete Pontryagin principle, when $X_0 \in (-1, 1)$, as depicted in Figure 2.2.

The problems occurring in the previous examples are all cured by regularizing the Hamiltonian and using the scheme (1.15). That is, the solution to (1.15) in the first two examples is a smooth curve that obtains a increasingly sharp kink near the time-axis as the regularizing parameter, δ, decreases, see Figure 2.3. In the last of the previous examples we, in contrast to the other methods, obtain a unique solution to (1.15).

Another problem that has not to do with nondifferentiability of the Hamiltonian is shown in the following example:

2: Let $B = [-1, 1]$, $f = \alpha$, $h = 0$ and $g = -|x|$. Although H is discontinuous here, this is not what causes problem. The problem is that optimal paths

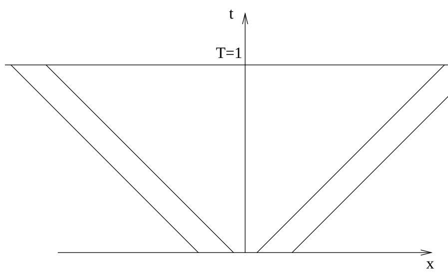

FIGURE 2.4. Solution of the optimization problem in Example (2), where $g(x) = -|x|$, $f = \alpha$, $h = 0$ and $B = [-1, 1]$, for four different starting points.

collide backwards, see Figure 2.4. When $X_0 = 0$ there are two solutions, one going to the left, and one to the right. The left solution has $\lambda = 1$ and the right solution has $\lambda = -1$, so on the time-axis λ is discontinuous. For these values of λ, the Hamiltonian is differentiable, therefore the nonsmoothness of the Hamiltonian is not the issue here. It is rather the global properties of the problem that play a role. This problem is difficult to regularize, and it will not be done here. However, we still can show convergence of the scheme (1.15). This is done in Section 4.3.

When using (1.15) to solve the minimization problem (1.3) it is assumed that the Hamiltonian is exactly known. Is this an unrealistic assumption in practice? In the following two examples we indicate that there exist interesting examples where we know the Hamiltonian. The first has to do with volatility estimation in finance, and the latter with optimization of an electric contact.

2.1. **Implied Volatility.** Black-Scholes equation for pricing general options uses the volatility of the underlying asset. This parameter, however, is difficult to estimate. One way of estimation is to use measured market values of options on the considered asset for standard European contracts. This way of implicitly determining the volatility is called implied volatility.

We assume that the financial asset obeys the following Ito stochastic differential equation,

$$dS(t) = \mu S(t)dt + \sigma\big(t, S(t)\big)S(t)dW(t), \tag{2.1}$$

where $S(t)$ is the price of the asset at time t, μ is a drift term, σ is the volatility and $W : \mathbb{R}_+ \to \mathbb{R}$ is the Wiener process. If the volatility is a sufficiently regular function of S, t, the strike level K and the maturity date T, the so called Dupire equation holds for the option price $C(T, K)$ as a function of T and K, with the present time $t = 0$ and stock price $S(0) = S$ fixed,

$$\begin{aligned} &C_T - \tilde{\sigma} C_{KK} = 0, \quad T \in (0, \infty), K > 0, \\ &C(0, K) = \max\{S - K, 0\} \quad K > 0, \end{aligned} \tag{2.2}$$

where

$$\tilde{\sigma}(T, K) \equiv \frac{\sigma^2(T, K)K^2}{2}.$$

Here the contract is an european call option with payoff function $\max\{S(T)-K,0\}$. We have for simplicity assumed the bank rate to be zero. The derivation of Dupire's equation (2.2) is given in [11] and [20].

The optimization problem now consists of finding $\sigma(T,K)$ such that

$$\int_0^{\hat{T}} \int_{\mathbb{R}_+} (C-\hat{C})^2(T,K)w(T,K)dKdT \tag{2.3}$$

is minimized, where $\hat{C}$ are the measured market values on option prices for different strike prices and strike times and w is a non negative weight function. In practice, $\hat{C}$ is not known everywhere, but for the sake of simplicity, we assume it is and set $w \equiv 1$, i.e. there exists a future time $\hat{T}$ such that $\hat{C}$ is defined in $\mathbb{R}_+ \times [0,\hat{T}]$. If the geometric Brownian motion would be a perfect model for the evolution of the price of the asset, the function $\sigma(T,K)$ would be constant, but as this is not the case, the σ that minimizes (2.3) (if a minimizer exists) varies with T and K.

It is possible to use (1.6) and (1.15) to perform the minimization of (2.3) over the solutions to a finite difference discretization of (2.2)

$$\begin{aligned} &\min_{\tilde{\sigma}} \int_0^{\hat{T}} \Delta K \sum_i (C-\hat{C})_i^2 dT \\ &\text{subject to} \quad \frac{\partial C_i(T)}{\partial T} = \tilde{\sigma} D^2 C_i(T), \\ &\qquad\qquad\quad C_i(0) = \max(S - i\Delta K, 0), \end{aligned} \tag{2.4}$$

where we now let $C_i(T) \approx C(T, i\Delta K)$ denote the discretized prize function, for strike time T and strike price $i\Delta K$, and D^2 is the standard three point difference approximation of the second order partial derivative in K. In order to have a finite dimensional problem we restrict to a compact interval $(0, M\Delta K)$ in K with the boundary conditions

$$C_0 = S, \quad C_M = 0.$$

The Hamiltonian for this problem is

$$\begin{aligned} H(\lambda, C) &= \Delta K \min_{\tilde{\sigma}} \sum_{i=1}^{M-1} \left(\lambda_i \tilde{\sigma}_i (D^2C)_i + (C-\hat{C})_i^2 \right) \\ &= \Delta K \sum_{i=1}^{M-1} \left(\min_{\tilde{\sigma}_i} \lambda_i \tilde{\sigma} (D^2C)_i + (C-\hat{C})_i^2 \right) \end{aligned}$$

where λ is the adjoint associated to the constraint (2.4). We have used that the components of the flux, f, in this problem is $\tilde{\sigma}_i(D^2C)_i$, that the running cost, h, is $\Delta K \sum_i (C-\hat{C})_i^2$, and further that each $\tilde{\sigma}_i$ minimizes $\lambda_i \tilde{\sigma}_i (D^2C)_i$ separately, so that the minimum can be moved inside the sum. If we make the simplifying assumption that $0 \le \sigma_- \le \tilde{\sigma} \le \sigma_+ < \infty$ we may introduce a function $s : \mathbb{R} \to \mathbb{R}$ as

$$s(y) \equiv \min_{\tilde{\sigma}} y\, \tilde{\sigma} = \begin{cases} y\sigma_-, & y > 0 \\ y\sigma_+, & y < 0. \end{cases}$$

Using s, it is possible to write the Hamiltonian as

$$H(\lambda, C) = \Delta K \sum_{i=1}^{M-1} \left(s\big(\lambda_i (D^2C)_i\big) + \big(C-\hat{C}\big)_i^2 \right).$$

Since s is nondifferentiable, so is H. However, s may easily be regularized, and it is possible to obtain the regularization in closed form, e.g. as in Example 1. Using a regularized version s_δ of s, the regularized Hamiltonian becomes

$$H^\delta(\lambda, C) = \Delta K \sum_{i=1}^{M-1} \Big(s_\delta\big(\lambda_i (D^2 C)_i\big) + \big(C - \hat{C}\big)_i^2 \Big),$$

which using Gateaux derivatives gives the Hamiltonian system

$$\begin{aligned} \frac{\partial C_i(T)}{\partial T} &= s'_\delta\big(\lambda_i (D^2 C)_i\big) D^2 C_i(T), \quad C_0 = S \quad C_M = 0, \\ -\frac{\partial \lambda_i(T)}{\partial T} &= D^2\Big(s'_\delta\big(\lambda_i (D^2 C)_i\big)\lambda\Big) + 2(C - \hat{C})_i, \\ \lambda_0 &= \lambda_M = 0, \end{aligned} \tag{2.5}$$

with data

$$C_i(0) = \max(S - i\Delta K, 0), \qquad \lambda(\hat{T}) = 0.$$

The corresponding Hamilton-Jacobi equation for the value function

$$u(C, \tau) = \int_\tau^{\hat{T}} \sum_{i=1}^{M-1} (C - \hat{C})_i^2 \Delta K \, dT$$

is

$$\begin{aligned} &u_T + H(u_C, C) = 0, \quad T < \hat{T}, \\ &u(\hat{T}, \cdot) = 0, \end{aligned}$$

where u_C is the Gateaux derivative with respect to C in the scalar product $(x, y) \equiv \Delta K \sum_i x_i, y_i$. With this scalar product the Hamiltonian system (2.5) takes the form

$$\begin{aligned} (C_T, v) &= (H^\delta_\lambda, v), \quad \forall v \in \mathbb{R}^{M-1} \\ (\lambda_T, v) &= -(H^\delta_C, v), \quad \forall v \in \mathbb{R}^{M-1} \end{aligned}$$

where H^δ_λ and H^δ_C are the Gateaux derivatives.

A choice of the regularization parameter δ, depending also on data error, can be obtained e.g. by the discrepancy principle, cf. [28], [12]. The Newton method described in Section 3 works well to solve the discrete equations for $d = 10$. Higher dimensions may require other regularization. The results of one trial volatility estimation is given in Figure 2.5.

2.2. **Optimal Design of Electric Conduction.** The problem is to place a given amount of conducting material in a given domain $\Omega \subset \mathbb{R}^d$ in order to minimize the power loss for a given surface current q, satisfying $\int_{\partial\Omega} q ds = 0$: let $\eta \in \mathbb{R}$ be a given constant, associated to the given amount of material, and find an optimal conduction distribution $\sigma : \Omega \to \{\sigma_-, \sigma_+\}$, where $\sigma_\pm > 0$, such that

$$\begin{aligned} &\mathrm{div}\big(\sigma \nabla \varphi(x)\big) = 0, \ x \in \Omega, \qquad \sigma \frac{\partial \varphi}{\partial n}\Big|_{\partial\Omega} = q \\ &\min_\sigma \Big(\int_{\partial\Omega} q\varphi ds + \eta \int_\Omega \sigma dx \Big), \end{aligned} \tag{2.6}$$

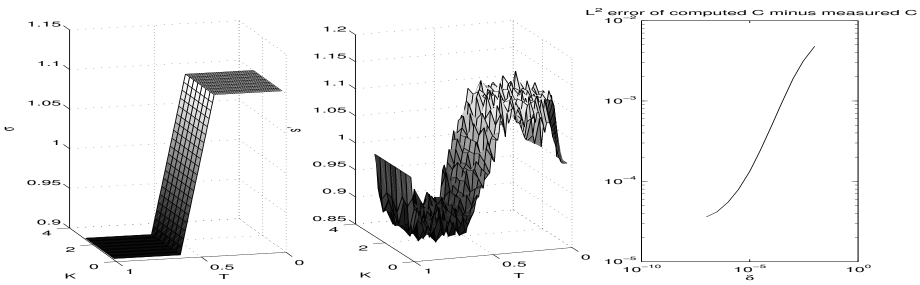

FIGURE 2.5. Results of a computer experiment where the volatility σ in the picture to the left is used to obtain the "measured" $\hat{C}$. Uniform noise of amplitude 10^{-4} is also added to $\hat{C}$. The error $\|C-\hat{C}\|_{L^2}$ is plotted versus δ in the picture to the right. In the middle picture the approximate volatility, s'_δ is shown for the value of δ $(= 3\cdot 10^{-6})$ that minimizes $\|s'_\delta-\sigma\|_{L^2}$. In this experiment, $M=9$ and $N=100$.

where $\partial/\partial n$ denotes the normal derivative and ds is the surface measure on $\partial\Omega$. Note that (2.6) implies that the power loss satisfies

$$\begin{aligned}\int_{\partial\Omega} q\varphi ds &= -\int_\Omega \operatorname{div}(\sigma\nabla\varphi)\varphi dx + \int_{\partial\Omega}\sigma\frac{\partial\varphi}{\partial n}\varphi ds\\ &= \int_\Omega \sigma\nabla\varphi\cdot\nabla\varphi dx.\end{aligned}$$

The Lagrangian takes the form

$$\int_{\partial\Omega} q(\varphi+\lambda)ds + \int_\Omega \sigma\,\underbrace{(\eta-\nabla\varphi\cdot\nabla\lambda)}_{v}\,dx$$

and the Hamiltonian becomes

$$H(\lambda,\varphi)=\min_\sigma\int_\Omega\sigma v dx+\int_{\partial\Omega}q(\varphi+\lambda)ds=\int_\Omega\underbrace{\min_\sigma\sigma v}_{s(v)}\,dx+\int_{\partial\Omega}q(\varphi+\lambda)ds$$

with the regularization

$$H^\delta(\lambda,\varphi)=\int_\Omega s_\delta(\eta-\nabla\varphi\cdot\nabla\lambda)dx+\int_{\partial\Omega}q(\varphi+\lambda)ds,$$

depending on the concave regularization $s_\delta\in\mathcal{C}^2(\mathbb{R})$ as in Section 2.1. The value function

$$u(\varphi,\tau)=\int_\tau^T(\int_{\partial\Omega}q\varphi ds+\eta\int_\Omega\sigma dx)dt$$

for the parabolic variant of (2.6), i.e.

$$\varphi_t=\operatorname{div}\big(\sigma\nabla\varphi(x)\big),$$

yields the infinite dimensional Hamilton-Jacobi equation

$$\partial_t u+H(\partial_\varphi u,\varphi)=0\quad t<T,\quad u(\cdot,T)=0,$$

using the Gateaux derivative $\partial_\varphi u = \lambda$ of the functional $u(\varphi,t)$ in $L^2(\Omega)$. The regularized Hamiltonian generates the following parabolic Hamiltonian system for φ and λ

$$\int_\Omega \big(\partial_t\varphi w + s'(\eta - \nabla\varphi\cdot\nabla\lambda)\nabla\varphi\cdot\nabla w\big)dx = \int_{\partial\Omega} qwds$$
$$\int_\Omega \big(-\partial_t\lambda v + s'(\eta - \nabla\varphi\cdot\nabla\lambda)\nabla\lambda\cdot\nabla v\big)dx = \int_{\partial\Omega} qvds$$

for all test functions $v, w \in V \equiv \{v \in H^1(\Omega) \mid \int_\Omega vdx = 0\}$. Time independent solutions satisfy $\lambda = \varphi$ by symmetry. Therefore the electric potential satisfies the nonlinear elliptic partial differential equation

$$\operatorname{div}\big(s'_\delta(\eta - |\nabla\varphi|^2)\nabla\varphi(x)\big) = 0 \quad x \in \Omega, \quad s'_\delta\frac{\partial\varphi}{\partial n}|_{\partial\Omega} = q, \tag{2.7}$$

which can be formulated as the convex minimization problem: $\varphi \in V$ is the unique minimizer (up to a constant) of

$$-\Big(\int_\Omega s_\delta(\eta - |\nabla\varphi(x)|^2)dx + 2\int_{\partial\Omega} q\varphi ds\Big). \tag{2.8}$$

In [7] we study convergence of

$$\lim_{T\to\infty} \frac{u(\varphi,t) - \bar u(\varphi,t)}{T},$$

where $\bar u$ is the value function associated to finite element approximations of the minimization (2.8).

The Newton method in Section 3 works well to solve the finite element version of (2.7) by successively decreasing δ, also for large d, see [7], where also the corresponding inverse problem to use measured approximations of φ to determine the domain where $\sigma = \sigma_-$ and $\sigma = \sigma_+$ is studied. A numerical solution to (2.7) can be seen in Figure 2.6.

3. Solution of the Discrete Problem

We assume in the theorems that the Pontryagin minimization (1.15) has been solved exactly. In practice (1.15) can only be solved approximately by iterations. The simplest iteration method to solve the boundary value problem (1.15) is the shooting method: start with an initial guess of $\bar\lambda[0]$ and compute, for all time steps n, the iterates

$$\begin{aligned} \bar X_{n+1} &= \bar X_n + \Delta t H^\delta_\lambda\big(\bar\lambda_{n+1}[i], \bar X_n\big), \quad n = 0,\dots,N-1, \quad \bar X_0 = X_0\\ \bar\lambda_n[i+1] &= \bar\lambda_{n+1}[i] + \Delta t H^\delta_x\big(\bar\lambda_{n+1}[i], \bar X_n\big),\ n = N-1,\dots,0,\ \bar\lambda_N = g_x(\bar X_N). \end{aligned} \tag{3.1}$$

An alternative method, better suited for many boundary value problems, is to use Newton iterations for the nonlinear system $F(\bar X, \bar\lambda) = 0$ where $F : \mathbb{R}^{Nd}\times\mathbb{R}^{Nd} \to \mathbb{R}^{2Nd}$ and

$$\begin{aligned} F(\bar X,\bar\lambda)_{2n} &= \bar X_{n+1} - \bar X_n - \Delta t H^\delta_\lambda\big(\bar\lambda_{n+1}, \bar X_n\big),\\ F(\bar X,\bar\lambda)_{2n+1} &= \bar\lambda_n - \bar\lambda_{n+1} - \Delta t H^\delta_x\big(\bar\lambda_{n+1}, \bar X_n\big). \end{aligned} \tag{3.2}$$

An advantage with the Pontryagin based method (3.2) is that the Jacobian of F can be calculated explicitly and it is sparse. The Newton method can be used to solve the volatility and optimal design examples in Section 2, where the parameter δ is successively decreasing as the nonlinear equation (3.2) is solved more accurately.

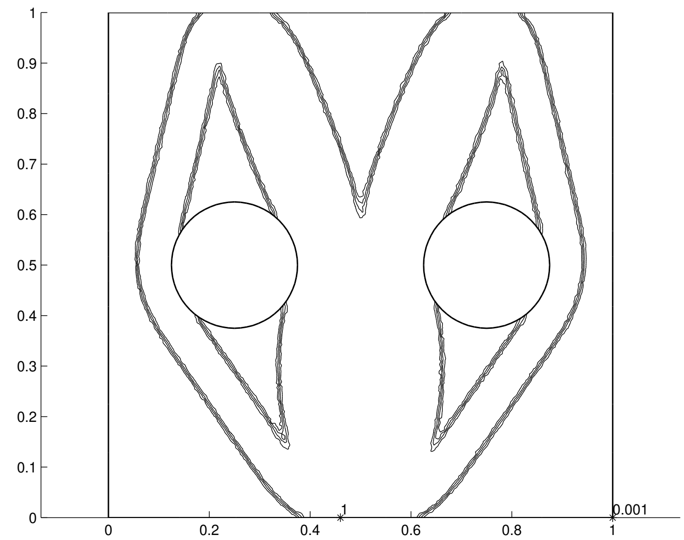

FIGURE 2.6. Contour plot of s'_δ as an approximation of the conductivity σ. As seen, Ω is in this example a square with two circles cut out. Electrical current enters Ω at two positions on the top of the square and leaves at one position on the bottom. The contours represent the levels 0.2, 0.4, 0.6 and 0.8. A piecewise linear FEM was used with 31440 elements, maximum element diameter 0.01, $\sigma_- = 0.001$, $\sigma_+ = 1$, $\eta = 0.15$ and $\delta = 10^{-5}$.

Let us use dynamic programming to show that the system (1.15) has a solution in the case that $\bar\lambda$ is a Lipschitz continuous function of (x,t), with Lipschitz norm independent of Δt, and $\delta > C\Delta t$. One step

$$x = y + \Delta t H_\lambda^\delta\big(\lambda(x), y\big) \tag{3.3}$$

for fixed $y \in \mathbb{R}^d$ has a solution $x(y)$ since the iterations

$$x[i+1] = y + \Delta t H_\lambda^\delta\big(\lambda(x[i]), y\big)$$

yield a contraction for the error $e[i] = x[i+m] - x[i]$

$$e[i+1] = \Delta t\Big(H_\lambda^\delta\big(\lambda(x[i+m]), y\big) - H_\lambda^\delta\big(\lambda(x[i]), y\big)\Big) = \Delta t\overline{H_{\lambda\lambda}^\delta \lambda_x} e[i].$$

Conversely, for all $x \in \mathbb{R}^d$ equation (3.3) has a solution $y(x)$ for each step since the iterations

$$y[i+1] = x - \Delta t H_\lambda^\delta\big(\lambda(x), y[i]\big)$$

generate a contraction for the error. The dynamic programming principle then shows that there are unique paths through all points $\bar X_{n+1}$ leading to all $\bar X_n$ for all n.

Remark 3.1 (Proof of (1.8) following [5, 14]). *A control $\hat\alpha$ such that*

$$\lambda \cdot f(x,\hat\alpha) + h(x,\hat\alpha) = H(\lambda, x)$$

satisfies for all small perturbations $w, v \in \mathbb{R}^d$

$$\begin{aligned}
&H(\lambda + w, x + v) - H(\lambda, x)\\
&\le (\lambda + w)\cdot f(x+v,\hat\alpha) + h(x+v,\hat\alpha) - \big(\lambda\cdot f(x,\hat\alpha) + h(x,\hat\alpha)\big)\\
&= f(x,\hat\alpha)\cdot w + \sum_i \big(\lambda\cdot f_{x_i}(x,\hat\alpha) + h_{x_i}(x,\hat\alpha)\big)v_i + o(|v|+|w|),
\end{aligned}$$

which implies (1.8).

4. Convergence of Euler Pontryagin Approximations

Theorem 4.1. *Assume that the Hamiltonian H, defined in* (1.5)*, is Lipschitz continuous on $\mathbb{R}^d \times \mathbb{R}^d$ and that (1.15) has a solution $(\bar X, \bar\lambda)$, where $\bar\lambda_{n+1}$ has uniformly bounded first variation with respect to $\bar X_n$ for all n and all Δt, i.e. there is a constant K such that*

$$|\partial_{\bar X_n}\bar\lambda_{n+1}| \le K. \tag{4.1}$$

Then the optimal solution, $(\bar X, \bar\lambda)$, of the Pontryagin method (1.15) satisfies the error estimate

$$\begin{aligned}
\Big|\inf_{\alpha\in\mathcal{A}}\Big(g\big(X(T)\big) + \int_0^T h\big(X(s),\alpha(s)\big)ds\Big) - \Big(g(\bar X_N) + \Delta t\sum_{n=0}^{N-1} h^\delta(\bar X_n, \bar\lambda_{n+1})\Big)\Big|\\
= \mathcal{O}(\Delta t + \delta + \frac{(\Delta t)^2}{\delta})\\
= \mathcal{O}(\Delta t), \quad \textit{for } \delta = \Delta t.
\end{aligned} \tag{4.2}$$

The bound $\mathcal{O}(\Delta t)$ in (4.2) depends on the dimension d through the Lipschitz norms of the Hamiltonian H and the constant K in (4.1).

Section 4.3 presents a convergence result for the case when backward paths $\bar X(t)$ collide on a $\mathcal{C}^1$ codimension one surface in $\mathbb{R}^d \times [0,T]$. The next subsections give a construction of a regularization H^δ and the proof of Theorem 4.1.

4.1. Construction of a Regularization.

A possible regularization of H is to let H^δ be a standard convolution mollification of H

$$H^\delta(\lambda, x) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} H(z,y)\omega^\delta(z-\lambda)\omega^\delta(y-x)dzdy,$$

with $\omega^\delta : \mathbb{R}^d \to \mathbb{R}_+$ a $\mathcal{C}^2$ function compactly supported in the ball $\{y \in \mathbb{R}^d : |y| \le \delta\}$ and with integral one $\int_{\mathbb{R}^d}\omega^\delta(y)dy = 1$. This regularization remains concave in λ. Our analysis is not dependent of this specific regularization, but uses that

$$\|H - H^\delta\|_C + \delta\|H^\delta\|_{C^1} + \delta^2\|H^\delta\|_{C^2} = \mathcal{O}(\delta),$$

and that H^δ remains a concave function of λ.

4.2. Convergence without Shocks and Colliding Paths.

The proof of the theorem is based on four lemmas. In all of those we suppose that the assumptions of Theorem 4.1 are valid.

Lemma 4.2. *The discrete dual function is the gradient of the value function, i.e.*

$$\bar u_x(\bar X_n, \bar t_n) = \bar\lambda_n. \tag{4.3}$$

Proof. The relation (4.3) holds for $t_N = T$. Use the induction assumption that (4.3) holds for $t_N, t_{N-1}, \ldots, t_{n+1}$. Then the definitions of f^δ and h^δ imply

$$\begin{aligned}
\frac{\partial \bar u}{\partial \bar X_n}(\bar X_n, t_n) &= \partial_{\bar X_n}\big(\bar u(\bar X_{n+1}, t_{n+1}) + \Delta t h^\delta(\bar\lambda_{n+1}, \bar X_n)\big)\\
&= \partial_{\bar X_n}\bar X_{n+1}\frac{\partial \bar u}{\partial \bar X_{n+1}}(\bar X_{n+1}, t_{n+1}) + \Delta t\partial_{\bar X_n} h^\delta(\bar\lambda_{n+1}, \bar X_n)\\
&= \big(I + \Delta t\partial_{\bar X_n} H^\delta_\lambda(\bar\lambda_{n+1}, \bar X_n)\big)\bar\lambda_{n+1} + \Delta t\partial_{\bar X_n} h^\delta(\bar\lambda_{n+1}, \bar X_n)\\
&= \bar\lambda_{n+1} + \Delta t\partial_{\bar X_n}(H^\delta_\lambda\lambda + h^\delta)(\bar\lambda_{n+1}, \bar X_n)\\
&\quad - \Delta t H^\delta_\lambda(\bar\lambda_{n+1}, \bar X_n)\partial_{\bar X_n}\bar\lambda_{n+1}\\
&= \bar\lambda_{n+1} + \Delta t H^\delta_x(\bar\lambda_{n+1}, \bar X_n)\\
&= \bar\lambda_n.
\end{aligned}$$

□

Section 4.4 shows that (4.3) holds precisely for symplectic methods.

We now extend $\bar u$ to be a function defined for all t. First extend the solution $\bar X$ to all time as a continuous piecewise linear function

$$\bar X(t) = \frac{t_{n+1} - t}{\Delta t}\bar X_n + \frac{t - t_n}{\Delta t}\bar X_{n+1}, \quad \text{for } t_n \le t < t_{n+1}, \tag{4.4}$$

so that

$$\bar X'(t) = H^\delta_\lambda(\bar\lambda_{n+1}, \bar X_n). \tag{4.5}$$

The following lemma shows that two different solutions can not collide for suitable small Δt.

Lemma 4.3. *There is a positive constant c such that if $\Delta t \le c\delta$ two different solutions $(\bar X^1, \bar\lambda^1)$ and $(\bar X^2, \bar\lambda^2)$ of (1.15) do not intersect.*

Proof. Assume there exist two optimal paths $(\bar X^1, \bar\lambda^1)$ and $(\bar X^2, \bar\lambda^2)$ that intersect at time t, where $\bar t_n < t \le \bar t_{n+1}$, then

$$\bar X^1_n + (t - \bar t_n)H^\delta_\lambda(\bar\lambda^1_{n+1}, \bar X^1_n) = \bar X^2_n + (t - \bar t_n)H^\delta_\lambda(\bar\lambda^2_{n+1}, \bar X^2_n)$$

which can be written

$$\bar X^1_n - \bar X^2_n = (t - t_n)\big(H^\delta_\lambda(\bar\lambda^2_{n+1}, \bar X^2_n) - H^\delta_\lambda(\bar\lambda^1_{n+1}), \bar X^1_n)\big). \tag{4.6}$$

To obtain an estimate of the size of the right hand side in (4.6) integrate along the line

$$\bar X(s) = \bar X^1_n + s(\bar X^2_n - \bar X^1_n),$$

with $\bar\lambda^i_{n+1}$ a function of $\bar X^i_n$. The difference in the right hand side of (4.6) is

$$\begin{aligned}
H^\delta_\lambda(\bar\lambda^2_{n+1}, \bar X^2_n) - H^\delta_\lambda(\bar\lambda^1_{n+1}, \bar X^1_n) &= \int_0^1 \frac{dH^\delta_\lambda}{ds}ds\\
= \int_0^1 \big(H^\delta_{\lambda x} + H^\delta_{\lambda\lambda}\partial_{\bar X_n}\bar\lambda_{n+1}\big)ds(\bar X^2_n - \bar X^1_n).
\end{aligned}$$

By assumption it holds that $\|H^\delta_{\lambda x} + H^\delta_{\lambda\lambda}\partial_{\bar X_n}\bar\lambda_{n+1}\|_{\mathcal C} = \mathcal O\big((C_\lambda(1+K)/\delta\big)$. Hence the norm of the right hand side in (4.6) is $\mathcal O(\delta^{-1}\Delta t)\|\bar X^1_n - \bar X^2_n\|$. Therefore there is a positive constant c such that if $\Delta t < c\delta$, the equation (4.6) has only the solution $\bar X^1_n = \bar X^2_n$. □

Since the optimal paths $\bar X$ do not collide, for suitable small Δt, the value function $\bar u$ is uniquely defined along the optimal paths, by (1.17) and

$$\bar u\big(\bar X(t),t\big) = \bar u(\bar X_{n+1},t_{n+1}) + (t_{n+1}-t)h^\delta(\bar X_n,\bar\lambda_{n+1}), \quad t_n < t < t_{n+1} \tag{4.7}$$

and we are ready for the main lemma

Lemma 4.4. *The value function for the Pontryagin method satisfies a Hamilton-Jacobi equation close to (1.4), more precisely there holds*

$$\begin{aligned} \bar u_t + H(\bar u_x,\cdot) &= \mathcal{O}(\delta + \Delta t + \frac{(\Delta t)^2}{\delta}) \quad \text{in } \mathbb{R}^d\times(0,T),\\ \bar u &= g \quad \text{on } \mathbb{R}^d. \end{aligned} \tag{4.8}$$

The error term $\mathcal{O}(\delta+\Delta t+\frac{(\Delta t)^2}{\delta})$ *in (4.8) is a Lipschitz continuous function of* $\bar u_x(x,t)$*,* x *and* t *satisfying*

$$|\mathcal{O}(\delta+\Delta t+\frac{(\Delta t)^2}{\delta})| \le CC_\lambda\Big(\delta + C_x\Delta t + C_xC_\lambda(1+K)\frac{(\Delta t)^2}{\delta}\Big),$$

where C_x *and* C_λ *are the Lipschitz constants of* H *in the* x *and* λ *variable, respectively, and* $C\sim 1$ *does not depend on the data.*

Proof. The proof starts with the observation

$$\begin{aligned} 0 &= \frac{d}{dt}\bar u(\bar X(t),t) + h^\delta(\bar\lambda_{n+1},\bar X_n)\\ &= \bar u_t(\bar X(t),t) + \bar u_x(\bar X(t),t)\cdot f^\delta(\bar\lambda_{n+1},\bar X_n) + h^\delta(\bar\lambda_{n+1},\bar X_n). \end{aligned} \tag{4.9}$$

The idea is now to use that the dual function $\bar\lambda$ is the gradient of $\bar u$ at the time levels t_n, by Lemma 4.2, (and a good approximation at times in between) and that the modified discrete Pontryagin method shows that the right hand side in (4.9) is consistent with the correct Hamiltonian H.

We will first derive an estimate of $|\bar u_x(\bar X(t),t) - \bar\lambda_{n+1}|$ for $t_n < t < t_{n+1}$. We have that

$$\bar u(\bar X(t),t) = \bar u(\bar X_{n+1},\bar t_{n+1}) + (\bar t_{n+1}-t)h^\delta(\bar\lambda_{n+1},\bar X_n)$$

Therefore $\bar u_x(\bar X(t),t)$ can be written as

$$\begin{aligned} \bar u_x(\bar X(t),t) &= \frac{\partial \bar X_n}{\partial \bar X_t}\Big(\frac{\partial \bar X_{n+1}}{\partial \bar X_n}\bar u_x(\bar X_{n+1},t_{n+1}) + (t_{n+1}-t)\partial_{\bar X_n}h^\delta(\bar\lambda_{n+1},\bar X_n)\Big)\\ &= \frac{\partial \bar X_n}{\partial \bar X_t}\Big(\frac{\partial \bar X_{n+1}}{\partial \bar X_n}\bar\lambda_{n+1} + (t_{n+1}-t)\partial_{\bar X_n}h^\delta(\bar\lambda_{n+1},\bar X_n)\Big). \end{aligned}$$

Introduce the notation

$$\begin{aligned} A \equiv \partial_{\bar X_n}H^\delta_\lambda(\bar\lambda_{n+1},\bar X_n) &= H^\delta_{\lambda x}(\bar\lambda_{n+1},\bar X_n) + H^\delta_{\lambda\lambda}(\bar\lambda_{n+1},\bar X_n)\partial_{\bar X_n}\bar\lambda_{n+1}\\ &= \mathcal{O}\big(C_\lambda(1+K)/\delta\big). \end{aligned} \tag{4.10}$$

We have

$$\begin{aligned} \frac{\partial \bar X_{n+1}}{\partial \bar X_n} &= I + \Delta t A = I + (t-t_n)A + (t_{n+1}-t)A\\ \frac{\partial \bar X_n}{\partial \bar X_t} &= \big(I + (t-t_n)A\big)^{-1} \end{aligned}$$

therefore as in Lemma 4.2

$$
\begin{aligned}
&\bar u_x(\bar X(t),t)\\
&= \bar\lambda_{n+1} + (t_{n+1}-t)\big(I+(t-t_n)A\big)^{-1}\big(A\bar\lambda_{n+1} + \partial_{\bar X_n} h^\delta(\bar\lambda_{n+1},\bar X_n)\big)\\
&= \bar\lambda_{n+1} + (t_{n+1}-t)\big(I+(t-t_n)A\big)^{-1} H_x^\delta(\bar\lambda_{n+1},\bar X_n)\\
&= \bar\lambda_{n+1} + \mathcal{O}\big(C_x\Delta t + C_x C_\lambda (K+1)(\Delta t)^2/\delta\big).
\end{aligned}
\tag{4.11}
$$

Introduce the notation $\tilde\lambda \equiv \bar u_x(\bar X(t),t)$ and split the Hamiltonian term in (4.9) into three error parts:

$$
\begin{aligned}
r(\tilde\lambda,\bar X(t),t) &\equiv \tilde\lambda f^\delta(\bar\lambda_{n+1},\bar X_n) + h^\delta(\bar\lambda_{n+1},\bar X_n) - H\big(\tilde\lambda,\bar X(t)\big)\\
&= \tilde\lambda f^\delta(\bar\lambda_{n+1},\bar X_n) + h^\delta(\bar\lambda_{n+1},\bar X_n) - H^\delta(\tilde\lambda,\bar X_n)\\
&\quad + H^\delta(\tilde\lambda,\bar X_n) - H^\delta\big(\tilde\lambda,\bar X(t)\big)\\
&\quad + H^\delta\big(\tilde\lambda,\bar X(t)\big) - H\big(\tilde\lambda,\bar X(t)\big)\\
&\equiv I + II + III.
\end{aligned}
\tag{4.12}
$$

Taylor expansion of H^δ to second order and (4.11) show

$$
\begin{aligned}
|I| &= |H^\delta(\bar\lambda_{n+1},\bar X_n) + (\tilde\lambda-\bar\lambda_{n+1})H_\lambda^\delta(\bar\lambda_{n+1},\bar X_n) - H^\delta(\tilde\lambda,\bar X_n)|\\
&\le \min\big(2C_\lambda|\tilde\lambda-\bar\lambda_{n+1}|, |(\tilde\lambda-\bar\lambda_{n+1})H_{\lambda\lambda}^\delta(\xi,\bar X_n)(\tilde\lambda-\bar\lambda_{n+1})|/2\big)\\
&\le CC_\lambda\big(C_x\Delta t + C_xC_\lambda(K+1)(\Delta t)^2/\delta\big);
\end{aligned}
$$

the Lipschitz continuity of H^δ implies

$$|II| \le |H_x^\delta||\bar X(t)-\bar X_n| \le |H_x^\delta||H_\lambda^\delta|\Delta t;$$

and the approximation H^δ satisfies

$$|III| \le CC_\lambda\delta.$$

The combination of these three estimates proves (4.8).

To finish the proof of the lemma we show that the error function r can be extended to a Lipschitz function in $\mathbb{R}^d\times\mathbb{R}^d\times[0,T]$. We note that by (4.1), (4.4) and (4.11) $\tilde\lambda$ is a Lipschitz function of X_t and t, and $r(\tilde\lambda(X_t,t),X_t,t)$ is Lipschitz in X_t and t. By

$$r(\lambda,X,t)\equiv r(\tilde\lambda(X,t),X,t)$$

we obtain a Lipschitz function r in $\mathbb{R}^d\times\mathbb{R}^d\times[0,T]$.

□

The results in these lemmas finishes the proof of Theorem 4.1: the combination of the residual estimates in Lemma 4.4 and the $\mathcal{C}$-stability estimate of viscosity solutions in Lemma 6.1 proves the theorem.

4.3. **Convergence with Shocks and Colliding Paths.** This section uses that the assumptions in Theorem 4.1 hold with the exception that the set of backward optimal paths $\{(\bar X(t),t) : t<T\}$, solving (1.15) and (4.4), may collide into a codimension one surface Γ in space-time $\mathbb{R}^d\times[0,T]$. In this section we assume that

$$\bar u(x,t_m) = \min_{\bar X_m = x}\Big(g(\bar X_N) + \sum_{n=m}^{N-1} h^\delta(\bar X_n,\bar\lambda_{n+1})\Delta t\Big) \tag{4.13}$$

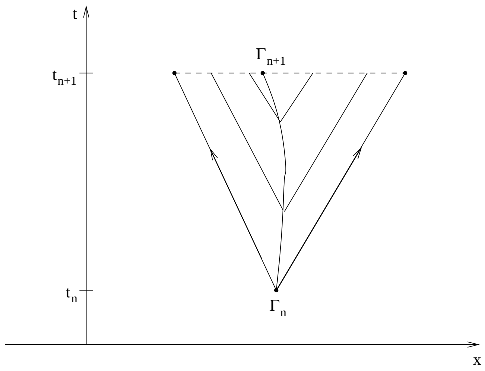

FIGURE 4.1. The cone in which backward paths collide at the surface Γ.

is attained by precisely one path for $(x,t_n) \in \mathbb{R}^d \times [0,T] \setminus \Gamma$ and that the minimum is attained by precisely two paths at $(x,t_n) \in \Gamma$; in Section 5 we consider the minimum attained by several paths. Colliding backward paths (or characteristics) X in general lead to a discontinuity in the gradient of the value function, $\lambda = u_x$, on the surface of collision, which means that the surface is a shock wave for the multidimensional system of conservation laws

$$\partial_t \lambda^i(x,t) + \frac{d}{dx_i} H\big(\lambda(x,t),x\big) = 0 \quad (x,t) \in \mathbb{R}^d \times [0,T],\ i = 1,\dots,d.$$

The properties concave Hamiltonian $H(\cdot,x)$; uniformly differentiable f with respect to x; Lipschitz continuous and semiconcave g; imply that the value function, u, is Lipschitz continuous and semiconcave, cf. [5], [1], [4], so that u is twice differentiable almost everywhere and the derivative has bounded variation. When the Hamiltonian $H(\cdot,x)$ is strictly concave the second derivative of u becomes a measure whose singular part is supported on the closure of the set of points where u is not differentiable, see [6].

Theorem 4.5. *Suppose that the assumptions in Theorem 4.1 hold with the exception that there is a codimension one Lipschitz surface Γ in $\mathbb{R}^d \times [0,T]$, where $\bar\lambda$ may be discontinuous, and the minimum in* (4.13) *is attained by precisely two paths, $\bar X^+$ and $\bar X^-$, for $(x,t) \in \Gamma$ and one path for $(x,t) \in \mathbb{R}^d \times [0,T] \setminus \Gamma$. Assume also that*

$$|\partial_{\bar X_n} \bar\lambda_{n+1}|\big|_{\{\bar X_n \in \mathbb{R}^d \setminus \Gamma_n\}} \le K, \quad \text{for all } n,$$

and

$$\frac{d\bar X^+}{dt} \cdot n \ge 0, \quad \frac{d\bar X^-}{dt} \cdot n \le 0,$$

where n is the normal to Γ at (x,t). Then the error estimate of Theorem 4.1 remains true, provided $\delta > C\Delta t$ for a certain positive constant C.

Let us first explain some assumptions for a construction of the surface Γ. Assume that at time t_n there is a codimension one surface Γ_n, in $\mathbb{R}^d$, where $\bar\lambda_n$ may be discontinuous and the minimum in (4.13) is attained by precisely two paths for $x \in \Gamma_n$ and one path for $x \in \mathbb{R}^d \setminus \Gamma_n$. The paths (4.4) starting on either side of Γ_n with $\bar u$ defined by (1.17) and (4.7) define $\bar u$ outside a cone in $\mathbb{R}^d \times [t_n, t_{n+1}]$, see Figure 4.1. In the union of these cones in $\mathbb{R}^d \times [0,T]$, we assume that a codimension one surface Γ, intersecting Γ_n for all n, can be constructed as follows. A point (x, t), where $t_n < t < t_{n+1}$, lies on Γ if there are two solutions starting in (x, t) to (1.15) where the first time step is $t_{n+1} - t$, and the corresponding values of $\bar{u}$ are equal. With this artificial time node at Γ, the backward paths are stopped at $(x, t) = (\bar{X}(t), t) \in \Gamma$ and defined for $t_n \leq t < s < t_{n+1}$ by

$$\bar{X}(s) = x + (s - t) H^\delta_\lambda(\bar{\lambda}_{n+1}, x),$$

with the value function

$$\bar{u}(\bar{X}(s), s) = g(\bar{X}_N) + h^\delta(\bar{X}(t), \bar{\lambda}_{n+1})(t_{n+1} - s) + \sum_{m=n+1}^{N-1} h^\delta(\bar{X}_m, \bar{\lambda}_{m+1})\Delta t.$$

Since by assumption there are no collisions outside Γ the value function $\bar{u}(x, t)$ is well defined everywhere as before. Denote the jump, for fixed t, of a function w at Γ by $[w]$. To have two colliding paths at a point on Γ requires that $\bar{\lambda}$ has a jump $[\bar{\lambda}] \neq 0$ there, since $[\bar{\lambda}] = 0$ yields only one path. The implicit function theorem shows that for fixed t any compact subset of the set $\Gamma(t) \equiv \Gamma \cap (\mathbb{R}^d \times \{t\})$ is a $\mathcal{C}^1$ surface: the surface $\Gamma(t)$ is defined by the value functions, $\bar{u}^1$ and $\bar{u}^2$ for the two paths colliding on Γ, being equal on Γ and there are directions $\hat{n} \in \mathbb{R}^d$ so that the Jacobian determinant $\hat{n} \cdot \nabla(\bar{u}^1 - \bar{u}^2) = \hat{n} \cdot [\bar{\lambda}] \neq 0$. Therefore compact subsets of the surface $\Gamma(t)$ has a well defined unit normal n. We assume that $\Gamma(t)$ has a normal everywhere.

Proof. The proof, to show that the estimate (4.8) in Lemma 4.4 holds in viscosity solution sense, is divided into the two steps: establishing that the subdifferential of $\bar{u}$ is empty on Γ; and verifying that the superdifferential of $\bar{u}$ is a subsolution to the Hamilton-Jacobi equation (4.8). We state the definition of the unique viscosity solution, based on sub and superdifferentials, in Definition 4.6 below.

The first step is to verify the claim $[\bar{\lambda}] \cdot n \leq 0$, which implies that the subdifferential of $\bar{u}$ is empty. By assumption

$$H^\delta_\lambda(\bar{\lambda}^+_{n+1}, x) \cdot n \geq 0, \quad H^\delta_\lambda(\bar{\lambda}^-_{n+1}, x) \cdot n \leq 0,$$

which yields

$$\overline{H^\delta_{\lambda\lambda}}(\bar{\lambda}^+_{n+1} - \bar{\lambda}^-_{n+1}) \cdot n \geq 0 \tag{4.14}$$

where $\overline{H^\delta_{\lambda\lambda}} \equiv \int_0^1 H^\delta_{\lambda\lambda}(\hat{\lambda}(s), x) ds$ and $\hat{\lambda}(s) \equiv s\bar{\lambda}^+_{n+1} + (1 - s)\bar{\lambda}^-_{n+1}$. The equation (1.15) implies

$$[\bar{\lambda}] = \big(I + (t_{n+1} - t)\overline{H^\delta_{x\lambda}}\big)(\bar{\lambda}^+_{n+1} - \bar{\lambda}^-_{n+1})$$

and consequently

$$\bar{\lambda}^+_{n+1} - \bar{\lambda}^-_{n+1} = \big(I + (t_{n+1} - t)\overline{H^\delta_{x\lambda}}\big)^{-1}[\bar{\lambda}]. \tag{4.15}$$

We know that $\bar{u}$ is continuous also around Γ, therefore the jump of the gradient, $[\bar{u}_x]$, has to be parallel to the normal, n, of the surface Γ. Lemma 4.2 shows that $[\bar{u}_x] = [\bar{\lambda}]$ and we conclude that this jump $[\bar{\lambda}]$ is parallel to n so that $[\bar{\lambda}] = [\bar{\lambda} \cdot n]n$, which combined with (4.14) and (4.15) shows that

$$0 \leq [\bar{\lambda} \cdot n]\overline{H^\delta_{\lambda\lambda}}\big(I + (t_{n+1} - t)\overline{H^\delta_{x\lambda}}\big)^{-1} n \cdot n.$$

The facts that the matrix $\overline{H^\delta_{\lambda\lambda}}$ is symmetric negative semi definite and that the matrix $\big(I + (t_{n+1} - t)\overline{H^\delta_{x\lambda}}\big)^{-1}$ is positive definite for $\delta > C\Delta t$ imply that

$$\overline{H^\delta_{\lambda\lambda}}\big(I + (t_{n+1} - t)\overline{H^\delta_{x\lambda}}\big)^{-1} n \cdot n \leq 0, \tag{4.16}$$

which proves the claim $[\bar\lambda] \cdot n \leq 0$, if we can exclude equality in (4.16). Equality in (4.16) means that $\overline{H^\delta_{\lambda\lambda}} n = 0$ and implies $H^\delta_\lambda(\bar\lambda^+(t), x) = H^\delta_\lambda(\bar\lambda^-(t), x)$ which is not compatible with two outgoing backward paths. Hence equality in (4.16) is ruled out.

From the first step we know that the subdifferential of $\bar u$ is empty, if $\bar u$ is not differentiable. In order to show that the error estimate from Theorem 4.1 holds it is sufficient to prove that $\bar u$ is a viscosity solution to

$$u_t + H(u_x, x) - q(u_x, x, t) = 0,$$

where q is a Lipschitz function such that $q(\bar\lambda(x,t), x, t) = r(\bar\lambda(x,t), x, t)$ for all (x,t), and $\sup|q| = \sup|r|$, where r is the remainder function $\mathcal{O}(\delta + \Delta t + \Delta t^2/\delta)$ derived in (4.8). Let $D^+\bar u$ denote the superdifferential of $\bar u$ at Γ. We know that $\bar u$ is a classical solution in $\mathbb{R}^d \times [0,T] \setminus \Gamma$. To prove that $\bar u$ is a viscosity solution means therefore to verify that

$$p_t + H(p_x, x) - q(p_x, x, t) \geq 0, \tag{4.17}$$

for all $(p_x, p_t) \in D^+\bar u(x,t)$ and all $(x,t) \in \Gamma$. We know that $D^+\bar u$ is the convex set $\{sp^+ + (1-s)p^- : s \in [0,1]\}$, where $p^\pm$ denotes the two right and left limits of the differential of $\bar u$ at Γ satisfying

$$p^\pm_t \partial_t \bar u + H(p^\pm_x, x) - r(p^\pm_x, x, t) = 0.$$

Denote by $\bar\lambda^\pm$ the boundary values of $\bar\lambda$ at Γ. Let $q(\lambda, x, t)$ satisfy

$$q(\bar\lambda(x,t), x, t) = r(\bar\lambda(x,t), x, t), \quad \text{for } (x,t) \in \mathbb{R}^d \times [0,T] \setminus \Gamma$$

and for $(x,t) \in \Gamma$,

$$q(\lambda, x, t) = s r(\bar\lambda^+, x, t) + (1-s) r(\bar\lambda^-, x, t), \quad \text{if } 0 \leq s \leq 1,$$

where

$$s = \frac{(\lambda - \bar\lambda^-) \cdot (\bar\lambda^+ - \bar\lambda^-)}{|\bar\lambda^+ - \bar\lambda^-|^2}.$$

With this choice q becomes a linear function of the variable λ, for fixed x, t, and since the Hamiltonian $H(\cdot, x)$ is a concave function it is clear that equation (4.17) holds.

We now give the construction of q, which by no means is unique. For $(x,t) \in \mathbb{R}^d \times [0,T]$ with $dist\big((x,t), \Gamma\big) > 1$ we let $q(\lambda, x, t) = r\big(\lambda(x,t), x, t\big)$, for all λ. What is left are two Lipschitz domains with given values on the boundaries. We may for instance let q be the harmonic function in these domains with the given boundary values, see Figure 4.2. The function q will be a Lipschitz function since for $(x,t) \in \Gamma$, $|r(\bar\lambda^+, x, t) - r(\bar\lambda^-, x, t)| = \mathcal{O}(|\bar\lambda^+ - \bar\lambda^-|)$. The domain is made compact by the dashed lines, but as these can be put arbitrarily far away it is clear that q can be extended to infinity without becoming greater than $\sup\|r\|$. ☐

Definition 4.6. *The subdifferential of $v : \mathbb{R}^d \times [0,T] \to \mathbb{R}$ is the set*

$$D^-v(y) \equiv \{p \in \mathbb{R}^{d+1} \ \big| \ \liminf_{z \to 0} \frac{v(y+z) - v(y) - p \cdot z}{|z|} \geq 0\}$$

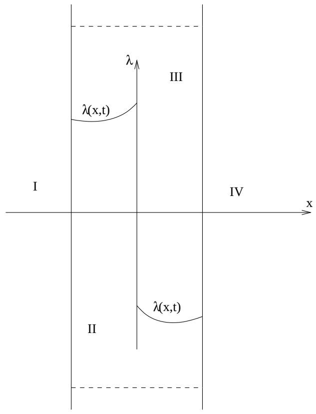

FIGURE 4.2. Construction of the function q. The shock surface is assumed to be located at $x = 0$. In the regions I and IV, q has the same value for all λ. In regions II and III, q is the solution to the Dirichlet problem with given values on $\lambda(x,t)$, the λ axis between $\bar\lambda^+$ and $\bar\lambda^-$, the surfaces $\{(x,t):\ dist\big((x,t),\Gamma\big) = 1\}$ and the dashed lines where λ is given by a linear interpolation.

and, similarly, the superdifferential of $v : \mathbb{R}^d \times [0,T] \to \mathbb{R}$ *is the set*

$$D^+v(y) \equiv \{p \in \mathbb{R}^{d+1} \mid \limsup_{z\to 0} \frac{v(y+z) - v(y) - p\cdot z}{|z|} \le 0\}.$$

The unique bounded and uniformly continuous viscosity solution, u*, to the Hamilton-Jacobi equation* (1.4)*, established in* [10]*, is defined by* $u(\cdot,T) = g$ *and that for all* $(x,t) \in \mathbb{R}^d \times (0,T)$

$$p_t + H(p_x, x) \le 0 \quad \forall (p_x,p_t) \in D^-u(x,t) \tag{4.18}$$

and

$$p_t + H(p_x, x) \ge 0 \quad \forall (p_x,p_t) \in D^+u(x,t). \tag{4.19}$$

Similarly a bounded uniformly continuous u *is called a supersolution (respectively subsolution) of* (1.4) *if* (4.18) *(respectively* (4.19)*) holds.*

4.4. **General Approximations.** The essential property of the symplectic Euler method we have used is that $\bar u_x(\bar X_n, t_n) = \bar\lambda_n$. This relation is in the following sense equivalent to symplectic approximation (cf. Remark 4.8):

Theorem 4.7. *Consider a general one step method*

$$\begin{aligned} \bar X_{n+1} &= A(\bar\lambda_{n+1}, \bar X_n)\\ \bar\lambda_n &= C(\bar\lambda_{n+1}, \bar X_n) \end{aligned} \tag{4.20}$$

with

$$\bar u(\bar X_n, t_n) = g(\bar X_N) + \sum_{m=n}^{N-1} B(\bar\lambda_{n+1}, \bar X_n).$$

(1) *Then* $\bar u_x(\bar X_n, t_n) = \bar\lambda_n$*, for all* n*, implies that the mapping* $\phi : (\bar X_n, \bar\lambda_n) \mapsto (\bar X_{n+1}, \bar\lambda_{n+1})$ *is symplectic.*

(2) *Conversely, if* ϕ *is symplectic,* (4.20) *can be written as* (4.21) *with a Hamiltonian* $\tilde H$*, and the canonical choice* $B = \Delta t(\tilde H - \lambda \tilde H)$ *implies that* $\bar u_x(\bar X_n, t_n) = \bar\lambda_n$*, for all* n*.*

Proof. As in Lemma 4.2 we have

$$\bar u_x(\bar X_n, t_n) = \frac{dA\big(\bar X_n, \bar\lambda_{n+1}(\bar X_n)\big)}{d\bar X_n}\bar u_x(\bar X_{n+1}, t_{n+1}) + \frac{dB\big(\bar X_n, \bar\lambda_{n+1}(\bar X_n)\big)}{d\bar X_n}.$$

Therefore the relation

$$\bar u_x(\bar X_n, t_n) = \bar\lambda_n$$

holds if and only if $\lambda A_\lambda + B_\lambda = 0$ and $\lambda A_x + B_x = C$.

Let $S \equiv \lambda A + B$. Then $\lambda A_\lambda + B_\lambda = 0$ is equivalent to $S_\lambda = A$, but $S_\lambda = A$ implies $B = S - \lambda S_\lambda$ so that $\lambda A_x + B_x = S_x$. Therefore $\lambda A_\lambda + B_\lambda = 0$ and $\lambda A_x + B_x = C$ is equivalent to $A = S_\lambda$ and $C = S_x$.

Let $S \equiv \bar\lambda_{n+1} \cdot \bar X_n + \Delta t \tilde H(\bar\lambda_{n+1}, \bar X_n)$. Then (4.20), with $A = S_\lambda$ and $C = S_x$, becomes

$$\begin{aligned} \bar X_{n+1} &= \bar X_n + \Delta t \tilde H_\lambda(\bar X_n, \bar\lambda_{n+1}) \\ \bar\lambda_n &= \bar\lambda_{n+1} + \Delta t \tilde H_x(\bar X_n, \bar\lambda_{n+1}), \end{aligned} \tag{4.21}$$

which by Remark 4.8 is equivalent to symplecticity of the mapping $(\bar X_n, \bar\lambda_n) \mapsto (\bar X_{n+1}, \bar\lambda_{n+1})$ and proves (1). Conversely, (2) follows from Lemma 4.2 applied to $\tilde H$ instead of H^δ. ☐

Remark 4.8. *A one step method* (4.20)*, interpreted as*

$$(\bar X_n, \bar\lambda_n) \mapsto (\bar X_{n+1}, \bar\lambda_{n+1}),$$

is called symplectic if there exists a function $\tilde H(\bar\lambda_{n+1}, \bar X_n)$ *such that* (4.21) *holds, see Theorem 5.1, Lemma 5.2 and (5.5) in Chapter VI of* [17]*, where a thorough study on symplectic methods can be found.*

To generalize the error estimate of Theorems 4.1 and 4.5 to general symplectic one step approximations (4.21), e.g. the second order symplectic Runge-Kutta method

$$\tilde H = \frac{1}{2}\Big(H(\bar\lambda_{n+1}, \bar X_n) + H\big(\bar\lambda_{n+1} + \Delta t H_x(\bar\lambda_{n+1}, \bar X_n), \bar X_n + \Delta t H_\lambda(\bar\lambda_{n+1}, \bar X_n)\big)\Big)$$

requires first an extension of $\bar X_n$ and $\bar u$ to all time, by approximations $(\bar f, \bar h)$ of (f^δ, h^δ) with

$$\frac{d\bar X}{dt} = \bar f \quad \text{and} \quad \frac{d\bar u}{dt} = -\bar h,$$

and then an estimate of the residual error r as in (4.12). In practice we need more regularity of H^δ to take advantage of higher order methods. Since we only have Lipschitz bounds of H the estimate of r is not smaller than the error $h^\delta - \bar h$, which is $\mathcal{O}(\|H^\delta\|_{\mathcal{C}^p})(\Delta t)^p = \mathcal{O}((\Delta t)^p/\delta^{p-1})$ for a pth order accurate method. Consequently the residual error is not smaller than $\mathcal{O}(\delta + (\Delta t)^p/\delta^{p-1}) = \mathcal{O}(\Delta t)$ for $\delta \simeq \Delta t$, so that our error estimate does not improve for higher order schemes, without additional assumptions. On the other hand by extending $\bar X$ as a piecewise linear function, as before, the only change of the analysis in Sections 4.2 and 4.3 to other symplectic methods (4.21) is to replace $H^\delta(\bar\lambda_{n+1}, \bar X_n)$ by $\tilde H(\bar\lambda_{n+1}, \bar X_n)$ and since

$$\|H^\delta - \tilde H\|_{\mathcal{C}} + \delta\|H^\delta - \tilde H\|_{\mathcal{C}^1} + \delta^2\|H^\delta - \tilde H\|_{\mathcal{C}^2} = \mathcal{O}(\Delta t)$$

the estimate (4.8) holds for all symplectic methods which are at least first order accurate.

Similarly, by considering $(\bar X_{n+1}, \bar\lambda_n)$, instead of $(\bar X_n, \bar\lambda_{n+1})$, as independent variables the scheme

$$\begin{aligned}\bar X_n &= A(\bar X_{n+1}, \bar\lambda_n)\\ \bar\lambda_{n+1} &= C(\bar X_{n+1}, \bar\lambda_n),\end{aligned}$$

is symplectic if and only if

$$\begin{aligned}\bar X_n &= \bar X_{n+1} - \Delta t \hat H_\lambda(\bar X_{n+1}, \bar\lambda_n)\\ \bar\lambda_{n+1} &= \bar\lambda_n - \Delta t \hat H_x(\bar X_{n+1}, \bar\lambda_n),\end{aligned} \tag{4.22}$$

and the error analysis of the methods (4.21) applies with

$$\tilde H(\bar X_n, \bar\lambda_{n+1}) = (\bar X_{n+1} - \bar X_n)\cdot(\bar\lambda_{n+1} - \bar\lambda_n) + \hat H(\bar X_{n+1}, \bar\lambda_n).$$

An example of a method (4.22) is the Euler method $\hat H = H$, which is backward Euler for $\bar X$ forwards in time and backward Euler for $\bar\lambda$ backwards in time, in contrast to (1.15) which is forward Euler for $\bar X$ forwards in time and forward Euler for $\bar\lambda$ backwards in time.

5. Colliding Shocks

This section describes when the minimum of $\bar u$ in (1.17) is attained by up to $d+1$ paths in $\mathbb{R}^d$. We perform the discussion with three paths, $\bar X^+$, $\bar X^-$ and $\bar X^0$, which yields a collision of three shock waves. The generalization to $d+1$ paths is then straightforward.

The codimension one surface Γ in space-time consists now of three parts which intersect where the minimum is attained by the three paths. The three parts are generated, as in Section 4.3, by the minimum attained by two paths $\bar X^+\&\bar X^-$, $\bar X^+\&\bar X^0$, and $\bar X^-\&\bar X^0$, respectively, see Figure 5.1. Section 4.3 showed that the subdifferential of $\bar u$ is empty on the three parts where the minimum is attained by two paths. The subdifferential is then also empty on the subset of Γ where the minimum is attained by three paths and the superdifferential becomes the convex combination

$$p_x = s_+\bar\lambda^+ + s_-\bar\lambda^- + s_0\bar\lambda^0$$

where $0 \le s_+,\ s_-, s_0 \le 1$ and

$$s_+ + s_- + s_0 = 1.$$

Here $\bar\lambda^+$, $\bar\lambda^-$, $\bar\lambda^0$ are the three limits of $\bar\lambda$ along the paths $\bar X^+$, $\bar X^-$ and $\bar X^0$, respectively. A construction of the remainder q, which is linear in λ on Γ, is still possible and we conclude that $\bar u$ is a viscosity solution to the Hamilton-Jacobi equation (4.8) with the same estimate of the remainder as in Section 4.

This derivation generalizes to a minimum attained by several paths, but the case with up to $d+1$ paths in $\mathbb{R}^d$ is special since this configuration is stable in the following sense. Here also the argument is for three paths, with easy generalization. We note that a piecewise smooth exact solution u to the Hamilton-Jacobi equation (1.4) satisfies the Rankine-Hugoniot condition

$$-s[\lambda] + [H]n = 0. \tag{5.1}$$

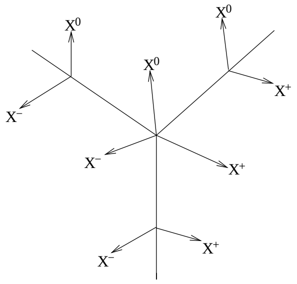

FIGURE 5.1. Collision of three backward paths X^+, X^- and X^0 yields three colliding shocks.

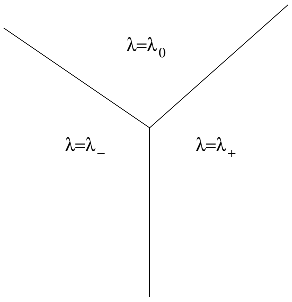

FIGURE 5.2. Collision of three planar shocks

The collision of three planar shock waves, cf. Figure 5.2, given by three states λ^0, λ^- and λ^+ intersecting at $x = 0$ at $t = 0$, is stable since the position $\Delta x \in \mathbb{R}^d$ of the intersection of the shock fronts at time Δt described by the three equations

$$(\Delta x - \Delta t s_i n_i) \cdot n_i = 0, \quad i = 0, -, +$$

can be written

$$\Delta x \cdot n_i = \Delta t s_i$$

which by (5.1) takes the form

$$\begin{aligned}
\frac{\Delta x}{\Delta t} \cdot (\lambda^+ - \lambda^-) &= H^+ - H^-, \\
\frac{\Delta x}{\Delta t} \cdot (\lambda^+ - \lambda^0) &= H^+ - H^0, \\
\frac{\Delta x}{\Delta t} \cdot (\lambda^- - \lambda^0) &= H^- - H^0.
\end{aligned}$$

Note that this system of three equations is in fact given by two linearly independent equations, with the intersection solution of a codimension 2 hyperplane in $\mathbb{R}^d$, so that the picture of three constant states λ_0, λ_+ and λ_- separated by three colliding shock waves remains a solution. The same line of reasoning gives that a collision of up to $d + 1$ shocks in $\mathbb{R}^d$ is stable.

6. $\mathcal{C}$ Stability for Hamilton-Jacobi Equations

The seminal construction of viscosity solutions by Crandall and Lions [8] also includes $\mathcal{C}$ stability results formulated in a general setting. We restate a variant adapted to the convergence results in this paper.

Lemma 6.1. *Suppose $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a Lipschitz continuous Hamiltonian satisfying for a constant C and for all $x, \hat{x}, \lambda, \hat{\lambda} \in \mathbb{R}^d$*

$$|H(\lambda, x) - H(\lambda, \hat{x})| \le C_x |x - \hat{x}|(1 + |\lambda|),$$
$$|H(\lambda, x) - H(\hat{\lambda}, x)| \le C_\lambda |\lambda - \hat{\lambda}|.$$

Suppose also that $e : \mathbb{R}^d \times [0,T] \to \mathbb{R}$ and $g : \mathbb{R}^d \to \mathbb{R}$ are Lipschitz continuous. Then, the bounded uniformly continuous viscosity solutions u and $\hat{u}$ of the Hamilton-Jacobi equations

$$u_t + H(u_x, \cdot) = 0 \quad \text{in } \mathbb{R}^d \times (0,T), \quad u|_{\mathbb{R}^d \times \{T\}} = g, \tag{6.1}$$

$$\hat{u}_t + H(\hat{u}_x, \cdot) = e \quad \text{in } \mathbb{R}^d \times (0,T), \quad \hat{u}|_{\mathbb{R}^d \times \{T\}} = g, \tag{6.2}$$

satisfy the $\mathcal{C}$-stability estimate

$$\|u - \hat{u}\|_{\mathcal{C}(\mathbb{R}^d \times [0,T])} \le T \|e\|_{\mathcal{C}(\mathbb{R}^d \times [0,T])}. \tag{6.3}$$

The comparison principle, cf. [4], for a subsolution, u^-, and a supersolution, u^+, of a Hamilton-Jacobi equation says that

$$\sup_{\mathbb{R}^d \times [0,T]} (u^- - u^+) \le \sup_{\mathbb{R}^d \times \{T\}} (u^- - u^+),$$

which can be proved with the usual doubling of variables technique for the uniqueness, cf. [13]. Since u is a viscosity solution of (6.1), we see by Definition 4.6 that $u^\pm \equiv u \pm (T - t)\|e\|_{\mathcal{C}(\mathbb{R}^d \times [0,T])}$ are super and subsolutions, respectively, of (6.2), and since $\hat{u}$ is both a sub and a supersolution to (6.2), the comparison principle yields

$$u^- \le \hat{u} \le u^+,$$

which implies (6.3).

7. How to obtain the Controls

The optimal control for the exact problem (1.4) is determined by the value function through the Pontryagin principle

$$\alpha(x,t) \in \operatorname*{argmin}_{a \in B} \big(u_x(x,t) \cdot f(x,a) + h(x,a)\big).$$

Assume we have solved a discrete approximating optimal control problem and obtained the approximations $\bar{X}$, $\bar{\lambda}$ and $\bar{u}$. Can they be used to determine an approximation of the control α? Even in the case that the optimal control $S(\lambda, x) \equiv \operatorname{argmin}_a \big(\lambda \cdot f(x,a) + h(x,a)\big)$ is a function, it is in general not continuous as function of x and λ but only piecewise Lipschitz continuous. Therefore the approximate control $S(\bar{\lambda}(t), x)$ cannot be accurate in the maximum norm. However, weaker measures of the control can converge; for instance the value function is accurately approximated in Theorems 4.1 and 4.5. At the points where S is Lipschitz continuous the error in the control is proportional to the error $|\bar{\lambda}(x,t) - u_x(x,t)|$, for fixed x. If we assume that the error $\bar{u}(\cdot,t) - u(\cdot,t)$ is bounded by ϵ in a $\sqrt{\epsilon}$-neighborhood of

x and that $\bar u_{xx}$ and u_{xx} also are bounded there, we obtain, for difference quotients $\Delta u/\Delta x$ and $|\Delta x| = \sqrt{\epsilon}$, the error estimate

$$\bar\lambda - u_x = \bar\lambda - \frac{\Delta \bar u}{\Delta x} + \frac{\Delta \bar u}{\Delta x} - \frac{\Delta u}{\Delta x} + \frac{\Delta u}{\Delta x} - u_x = \mathcal{O}(\Delta x + \epsilon/\Delta x) = \mathcal{O}(\sqrt{\epsilon}).$$

Convergence of the approximate path $(\bar X, \bar\lambda)$ typically requires Lipschitz continuous flux (H_λ, H_x), which we do not assume in this work.

Acknowledgment. The authors thank Jesper Carlsson for the Figure 2.6 and calculations in obtaining it.

References

[1] G. Barles, Solutions de Viscosité des Equations de Hamilton-Jacobi. Mathématiques & Applications (Berlin), 17. Springer-Verlag, Paris, 1994.

[2] G. Barles and E. Jakobsen, On the convergence rate of approximation schemes for Hamilton-Jacobi-Bellman equations. M2AN Math. Model. Numer. Anal. 36 (2002), no. 1, 33–54.

[3] E. Barron and R. Jensen, The Pontryagin maximum principle from dynamic programming and viscosity solutions to first-order partial differential equations. Trans. Amer. Math. Soc. 298 (1986), no. 2, 635–641.

[4] M. Bardi and I. Capuzzo-Dolcetta, Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations, with appendices by M. Falcone and P. Soravia, Systems and Control: Foundations and Applications. Birkhäuser Boston, Inc., Boston, MA, 1997.

[5] P. Cannarsa and H. Frankowska, Some characterizations of the optimal trajectories in control theory, SIAM J. Control Optim., 29 (1991), 1322-1347.

[6] P. Cannarsa, A. Mennucci and C. Sinestrari, Regularity results for solutions of a class of Hamilton-Jacobi equations. Arch. Rational Mech. Anal. 140 (1997), no. 3, 197–223.

[7] J. Carlsson, M. Sandberg and A. Szepessy, Convergence rates for distributed optimal control problems, work in progress.

[8] M. Crandall and P.-L. Lions, Viscosity solutions of Hamilton-Jacobi equations. Trans. Amer. Math. Soc. 277 (1983), no. 1, 1–42.

[9] M. Crandall and P.-L. Lions, Two approximations of solutions of Hamilton-Jacobi equations. Math. Comp. 43 (1984), no. 167, 1–19.

[10] M. Crandall, L.C. Evans and P.-L. Lions, Some properties of viscosity solutions of Hamilton-Jacobi equations. Trans. Amer. Math. Soc. 282 (1984), no. 2, 487–502.

[11] B. Dupire, Pricing with a smile, Risk (1994), 18-20.

[12] H. Engl, M. Hanke and A. Neubauer, Regularization of Inverse Problems. Mathematics and its Applications, 375. Kluwer Academic Publishers Group, Dordrecht, 1996.

[13] L.C. Evans, Partial Differential Equations. Graduate Studies in Mathematics, 19. American Mathematical Society, Providence, RI, 1998.

[14] H. Frankowska, Contigent cones to reachable sets of control systems, SIAM J. Control Optim. 27 (1989), 170-198.

[15] R. Glowinski and J.-L. Lions, Exact and approximate controllability for distributed parameter systems. Acta numerica, 1995, 159–333, Acta Numer., Cambridge Univ. Press, Cambridge, 1995.

[16] R. Glowinski and J.-L. Lions, Exact and approximate controllability for distributed parameter systems. Acta numerica, 1994, 269–378, Acta Numer., Cambridge Univ. Press, Cambridge, 1994.

[17] E. Harrier, C. Lubich and G. Wanner, Geometric Numerical Integrators: Structure Preserving Algorithms for Ordinary Differential Equations, Springer 2002.

[18] M. Falcone and R. Ferretti, Semi-Lagrangian schemes for Hamilton-Jacobi equations, discrete representation formulae and Godunov methods, J. Comput. Phys. 175 (2002), no. 2, 559–575.

[19] P. Pedregal,Optimization, relaxation and Young measures. Bull. Amer. Math. Soc. (N.S.) 36 (1999), no. 1, 27–58.

[20] Y. Achdou and O. Pironneau, Volatility smile by multilevel least square. Int. J. Theor. Appl. Finance 5 (2002), no. 6, 619–643.

[21] B. Mohammadi and O. Pironneau, Applied Shape Optimization for Fluids. Numerical Mathematics and Scientific Computation. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 2001

[22] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze and E.F. Mishchenko, The Mathematical Theory of Optimal Processes, Pergamon Press, 1964.

[23] E. Polak, Optimization, Algorithms and Consistent Approximations, Applied Mathematical Sciences, 124. Springer-Verlag, New York, 1997.

[24] M. Sandberg, Convergence rates for Euler approximation of non smooth optimal control, work in progress.

[25] P. Souganidis, Existence of viscosity solutions of Hamilton-Jacobi equations. J. Differential Equations 56 (1985), no. 3, 345–390.

[26] A. Subbotin, Generalized Solutions of First-Order PDEs. The Dynamical Optimization Perspective. Translated from the Russian. Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA, 1995.

[27] C.-T. Lin and E. Tadmor, L^1-stability and error estimates for approximate Hamilton-Jacobi solutions. Numer. Math. 87 (2001), no. 4, 701–735.

[28] C. Vogel, Computational Methods for Inverse Problems. Frontiers in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002.

[29] L.C. Young, Lectures on the Calculus of Variations and Optimal Control Theory. Saunders Co., Philadelphia-London-Toronto, Ont. 1969

Institutionen för Matematik, Kungl. Tekniska Högskolan,, S-100 44 Stockholm; e-mail: mattiass@math.kth.se & szepessy@kth.se; telephone: +46 8 7906196 & +46 8 7907494