Hamilton-Jacobi equations for sorting and percolation problems

by

Jeffrey William Calder

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Applied and Interdisciplinary Mathematics)
in The University of Michigan
2014

Doctoral Committee:

Professor Selim Esedoḡlu, Co-Chair
Professor Alfred O. Hero III, Co-Chair
Professor Jinho Baik
Professor Joseph G. Conlon
Professor Anna C. Gilbert
Professor Roman Vershynin

ACKNOWLEDGEMENTS

There are many people that have kindly offered their help and support, and without whom, writing this thesis would not be possible.

First, thank you to my advisors Professors Selim Esedoḡlu and Alfred Hero. Thank you for inspiring me and developing my interest in rigorous applied mathematics. You have both played a fundamental role in my training as an applied mathematician and have provided me with incredible opportunities to travel, speak, and work. I will always be grateful for your continued support, enthusiasm, and motivation.

Thank you to Professor Jinho Baik for pointing out the connection between my dissertation work and last passage percolation, and providing guidance for the second part of this dissertation. Thank you to the other committee members, Professors Anna Gilbert, Joseph Conlon, and Roman Vershynin. Though you were not involved in the day to day research activities, you have all provided valuable feedback on my work, and for that I am grateful.

There are many graduate students who have contributed to this dissertation in some form. Thank you to Mark Hsiao for stimulating discussions and collaboration on applications of Pareto methods, and for providing code to manipulate some of the datasets used in this thesis. Thank you to Daniel DeWoskin, Purvi Gupta, and Matthew Masarik for useful discussions about various aspects of this work. I am also grateful to John Duchi for pointing out an error in an earlier version of this

dissertation.

I am very grateful for the financial support of the Natural Sciences and Engineering Research Council of Canada, the National Science Foundation, the Society for Industrial and Applied Mathematics, and the Rackham Predoctoral Fellowship. This support allowed me to travel to numerous workshops and conferences to present my work, and freed me of other responsibilities so that I could devote a majority of my time to research.

Finally, I would like to thank my family, especially my wife Hayley. She has given her unequivocal support throughout this journey, for which this mere expression of thanks does not suffice.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF ALGORITHMS

ABSTRACT

Hamilton-Jacobi equations for sorting and percolation problems

by

Jeff Calder

Co-Chair: Professor Selim Esedoḡlu
Co-Chair: Professor Alfred O. Hero III

In this dissertation we prove continuum limits for some sorting and percolation problems that are important in mathematical, scientific, and engineering contexts. The first problem we study is non-dominated sorting, which is a fundamental combinatorial problem in multi-objective optimization. The sorting can be viewed as arranging points in Euclidean space into fronts according to a partial order. We show that these fronts converge almost surely to the level sets of a function that satisfies a Hamilton-Jacobi equation in the viscosity sense. Of course, multi-objective optimization is ubiquitous in scientific and engineering contexts, and, as it turns out, non-dominated sorting is also equivalent to the longest chain problem, which has a long history in probability and combinatorics. We present a fast numerical scheme for solving this Hamilton-Jacobi equation and prove convergence and various properties of the scheme. We then show how to use the scheme to design a fast approximate non-dominated sorting algorithm and we demonstrate the algorithm on

synthetic data as well as a large-scale real-world dataset.

The second problem we study is directed last passage percolation (DLPP), which is a stochastic growth model with applications in directed polymer growth, queuing systems, and stochastic particle systems. DLPP is closely related to the longest chain problem, and by using similar techniques we prove that a DLPP model with macroscopic and discontinuous weights has a continuum limit that corresponds to solving a Hamilton-Jacobi equation. We further prove convergence of a numerical scheme for this Hamilton-Jacobi equation and present an algorithm based on dynamic programming for finding the asymptotic shapes of maximal directed paths.

CHAPTER I

Introduction

In this dissertation we prove continuum limits for some sorting and percolation problems that are important in mathematical, scientific, and engineering contexts. The problems we study are non-dominated sorting, or the longest chain problem, and directed last passage percolation. Non-dominated sorting is a type of sorting widely used in multi-objective optimization, which is ubiquitous in science and engineering. It is also equivalent to the longest chain problem, which has applications in materials science, pure mathematics, and computational biology, among other fields. Directed last passage percolation is a stochastic growth process closely to the longest chain problem, and has applications in directed polymer growth, queuing systems, and stochastic particle systems.

A common thread tying these combinatorial problems together is their formulation as discrete variational problems. The continuum limits are continuous variational problems obtained by a type of homogenization. These variational problems are then associated with Hamilton-Jacobi equations through a well-known technique from optimal control. One of the fundamental contributions of this work is the identification of Hamilton-Jacobi equations for these continuum limits. This allows efficient and accurate computation of the limits, as well as opening up another avenue

for theoretical study.

We give some applications of these continuum limits as well. As a theoretical application, we are able to show, using techniques from the theory of viscosity solutions of Hamilton-Jacobi equations, that non-dominated sorting is asymptotically stable under bounded random perturbations. We then show how to design a fast approximate non-dominated sorting algorithm based on a fast numerical scheme for the continuum limit Hamilton-Jacobi equation. We evaluate the algorithm on a real-world large-scale dataset and show that it is significantly faster than non-dominated sorting while maintaining a high degree of sorting accuracy. For directed last passage percolation, we show how to efficiently compute the asymptotic shapes of optimal paths by solving the continuum limit Hamilton-Jacobi equation and applying dynamic programming. Aside from these applications, there are undoubtedly many others that we leave for future work, and we discuss some possibilities in Chapter VI. Let us now describe each of these combinatorial problems in further detail.

1.1 Non-dominated sorting

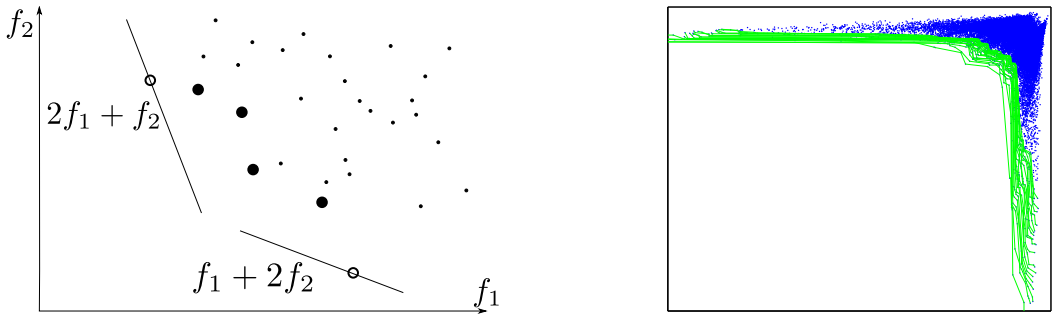
Consider the following discrete multi-objective optimization problem: Given several objective functions $f_i : S \to [0, \infty)$, where $i = 1, \dots, d$ and $S = \{x^1, \dots, x^n\}$ is a finite set, find $x \in S$ minimizing *all* of the objectives in some appropriate sense. Since it is generally impossible to find $x \in S$ that minimizing all of the objectives simultaneously when $d > 1$, a basic approach is to combine the objectives together into one objective. Such approaches are typically called *scalarization* [32], and a basic, though commonly used technique is to form a linear combination

$$f_\alpha = \alpha_1 f_1 + \cdots \alpha_d f_d.$$

Such an approach, which is called *linear scalarization*, suffers from two significant drawbacks. First, the minimizer or minimizers of f_α depend on the choice of weights $\alpha = (\alpha_1, \ldots, \alpha_d)$. There is in general no principled approach for selecting the weights without some *a priori* knowledge about the relative importance of each objective, and searching over all possible combinations of weights is prohibitively expensive. Second, linear scalarization misses many important and relevant solutions. For example, suppose our set has three elements $S = \{a, b, c\}$ and $(f_1(a), f_2(a)) = (1, 0)$, $(f_1(b), f_2(b)) = (0, 1)$ and $(f_1(c), f_2(c)) = (0.6, 0.6)$. It is easy to see that for any selection of weights $\alpha = (\alpha_1, \alpha_2)$, either a or b (or both) will be the minimizer of f_α, and the solution c will be ignored. However, c is more desirable than a in terms of f_1 and more desirable than b in terms of f_2. There is therefore no *a priori* reason we should prefer a and b over c.

A more natural notion of solution to a multi-objective optimization problem is the notion of Pareto-optimality. We say a feasible solution $x \in S$ is *Pareto-optimal* if for every $y \in S$, we have $f_i(y) > f_i(x)$ for some i, or $f_i(y) = f_i(x)$ for all i. In other words, there are no feasible solutions that are at least as desirable in all objectives and strictly better in one. The notion of Pareto-optimality is originally due to Vilfredo Pareto, an Italian economist who pioneered the idea for problems in economics, and it has since proven to be a powerful concept with applications in numerous other fields of science and engineering [32]. For example, multi-objective optimization problems alone have found applications in control theory and path planning [72, 63, 67], gene selection and ranking [88, 49, 48, 50, 36, 37, 38], data clustering [47], database systems [60, 75] and image processing and computer vision [73, 24].

We denote the collection of Pareto-optimal elements by $\mathcal{F}_1$ and it is called the *first Pareto front*. The second Pareto front, $\mathcal{F}_2$, consists of the Pareto-optimal elements



(a) (b)

Figure 1.1: (a) Depiction of nonconvexities in the first Pareto front. The large points are Pareto-optimal, but only the hollow points can be obtained by linear scalarization. (b) Depiction of nonconvexities in the Pareto fronts in a real-world example discussed in more detail in Chapter IV.

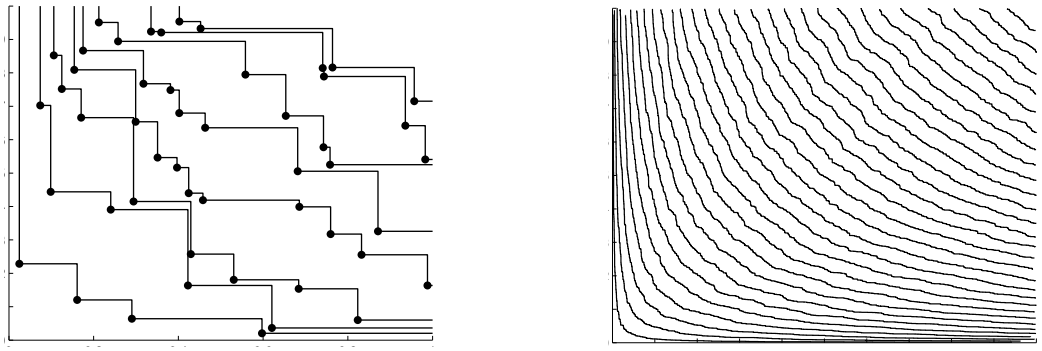
of $S \setminus \mathcal{F}_1$, and in general

$$\mathcal{F}_k = \text{Pareto optimal elements of } S \setminus \bigcup_{j<k} \mathcal{F}_j.$$

The Pareto front that a particular feasible solution lies on is useful for ranking feasible solutions to the multi-objective optimization problem. Any minimizer of a linear scalarization f_α is Pareto-optimal, however, not all Pareto-optimal solutions are minimizers of f_α for some choice of weights. In our simple example above, a, b and c are all Pareto-optimal, though c is not the minimizer of any linear scalarization. In general, linear scalarization finds the Pareto-optimal points on the boundary of the convex hull of the Pareto front. See Figure 1.1 for an illustration. We will explore this more in Chapter II.

We should mention at this point that there are other non-linear scalarization approaches that are capable of finding most (or even all) of the Pareto-optimal feasible solutions. For example, Athan and Papalambros [4] consider a linear combination of powers of the objectives as follows:

$$f_{\alpha,p} = \alpha_1 f_1^p + \cdots + \alpha_d f_d^p.$$



(a) $n = 50$ points (b) $n = 10^6$ points

Figure 1.2: Examples of Pareto fronts for $X_1, \dots X_n$ chosen from the uniform distribution on $[0, 1]^2$. In (b), 29 equally spaced fronts are depicted.

It turns out that every Pareto-optimal solution can be obtained by appropriate choices of α and p. As is the case for linear scalarization, the selection of appropriate weights and exponent is challenging, and the search space grows exponentially with dimension. Furthermore, depending on the geometry of the first front, the exponent p may have to be quite large in order to obtain all Pareto-optimal points, and this can become difficult computationally [4].

The process of arranging the set S into Pareto-fronts is called *non-dominated sorting* [28, 29]. If we set $X_i = (f_1(x^i), \dots, f_d(x^i)) \in \mathbb{R}^d$ for $i = 1, \dots, n$, then it is clear that non-dominated sorting depends only on the points $X_1, \dots, X_n \in \mathbb{R}^d$. Thus, we will from now on consider non-dominated sorting as a combinatorial sorting problem applied to points in $\mathbb{R}^d$. Figure 1.2(a) shows the Pareto-fronts obtained by non-dominated sorting of $n = 50$ points $X_1, \dots, X_n$ independent and identically distributed on $[0, 1]^2$, and Figure 1.2(b) shows the same for $n = 10^6$ points.

Although we have described non-dominated sorting in the context of a discrete optimization problem, it is a fundamental tool in continuous optimization as well. Many state of the art algorithms for continuous optimization involve a large number of discrete subproblems, each of which requires non-dominated sorting. The

most common examples are the so-called genetic and evolutionary algorithms for continuous multi-objective optimization [29, 39, 40, 28, 89]. The applications of non-dominated sorting are not restricted to optimization; indeed, there are further striking applications in combinatorics [34, 66], molecular biology [76, 1], graph theory [66], Young Tableaux [95, 34] and even in physical layout problems in the design of integrated circuits [1].

1.2 Longest chain in Euclidean space

Let $X_1, \ldots, X_n$ be independent and identically distributed (*i.i.d.*) random variables on $\mathbb{R}^d$ with density function $f \in L^1(\mathbb{R}^d)$. The points form a partially ordered set $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ under the partial order

$$x \leqq y \iff x_i \leq y_i \text{ for } i = 1, \ldots, d. \tag{1.1}$$

Let $\ell(n)$ denote the length of a longest chain—a totally ordered subset—in $\mathcal{X}_n$. For $x \in \mathbb{R}^d$ let $u_n(x)$ denote the length of a longest chain in $\mathcal{X}_n$ consisting of points less than or equal to x.

When f is a smooth density on $[0,1]^d$, hence $\ell(n) = u_n(1, \ldots, 1)$, the problem of studying the asymptotics of $\ell(n)$ has a long history. It begins with Ulam's famous problem [92] of finding the length of a longest increasing subsequence of a random permutation. Hammersley [46] made some of the first breakthroughs in understanding Ulam's problem. He observed that the distribution of the length of a longest increasing subsequence among n numbers chosen uniformly at random is the same as the distribution of $\ell(n)$ for uniformly distributed points on $[0,1]^2$. Using subadditive ergodic theory, Hammersley showed that $n^{-\frac{1}{2}}\ell(n)$ converges almost surely to a constant c as $n \to \infty$, and he conjectured that $c = 2$. In subsequent papers, Vershik and Kerov [93] and Logan and Shepp [65] showed that $c \leq 2$ and $c \geq 2$, respectively.

Hammersley's results were generalized by Bollobás and Winkler [17] to uniformly distributed points on $[0,1]^d$; they showed that there exist positive constants c_d such that $n^{-\frac{1}{d}}\ell(n) \to c_d$ almost surely as $n \to \infty$, and $c_d \nearrow e$ as $d \to \infty$. The only known values of c_d are $c_1 = 1$ and $c_2 = 2$. Deuschel and Zeitouni [31] generalized Hammersley's results in another direction. For $X_1, \ldots, X_n$ *i.i.d.* on $[0,1]^2$ with C^1 density function $f : [0,1]^2 \to \mathbb{R}$, bounded away from zero, they showed that $n^{-\frac{1}{2}}\ell(n) \to 2\overline{J}$ in probability, where $\overline{J}$ is the supremum of the energy

$$J(\varphi) = \int_0^1 \sqrt{\varphi'(x) f(x, \varphi(x))}\, dx,$$

over all $\varphi : [0,1] \to [0,1]$ nondecreasing and right continuous.

The longest chain problem is actually equivalent to non-dominated sorting. To see this, let us first give a definition of non-dominated sorting that is independent of the underlying multi-objective optimization problem. Given a set of points $\mathcal{X}_n = \{X_1, \ldots, X_n\}$, the first Pareto front is the collection of minimal points in $\mathcal{X}_n$ with respect to the partial order $\leqq$, i.e.,

$$\mathcal{F}_1 = \text{minimal elements of } \mathcal{X}_n.$$

As before, the deeper Pareto fronts are defined recursively as follows:

$$\mathcal{F}_k = \text{minimal elements of } \mathcal{X}_n \setminus \bigcup_{j<k} \mathcal{F}_j.$$

Let us suppose now that the points $X_1, \ldots, X_n$ are all distinct. Then it is clear that a point X_i is in the first Pareto front if and only if the longest chain consisting of points less than or equal to X_i has length 1, i.e.,

$$X_i \in \mathcal{F}_1 \iff u_n(X_i) = 1.$$

By stripping off the first Pareto front and repeating the same argument, it is clear

that in general we have

$$X_i \in \mathcal{F}_k \iff u_n(X_i) = k.$$

This observation is essential. It says that studying the asymptotic shapes of the Pareto fronts $\mathcal{F}_1, \mathcal{F}_2, \ldots$ is equivalent to studying the longest chain function u_n. Notice in Figure 1.2, the points X_i that are on the same Pareto front are connected by a continuous staircase curve that represents the jump set of u_n. It is interesting to note that in the combinatorics literature, the partition $\mathcal{X}_n = \mathcal{F}_1 \cup \mathcal{F}_2 \cup \cdots$ is called the *canonical antichain partition* [34]. The longest chain problem in two dimensions is also closely related to patience sorting [3].

1.3 Directed last passage percolation

The directed last passage percolation (DLPP) problem can be formulated as follows: Let $X(i,j)$ be nonnegative independent random variables defined on the lattice $\mathbb{N}^2$, and define the last passage time from $(1,1)$ to (M,N) by

$$L(M,N) = \max_{p \in \Pi_{M,N}} \sum_{(i,j)\in p} X(i,j), \tag{1.2}$$

where $\Pi_{M,N}$ denotes the set of up/right paths from $(1,1)$ to (M,N) in $\mathbb{N}^2$. Of interest are the asymptotics of L as $M, N \to \infty$, and their first order fluctuations.

DLPP is a stochastic growth model that has many applications in mathematical and scientific contexts. One important application of DLPP is zero-temperature directed polymer growth in a random environment, which is an important model in statistical mechanics [25, 52, 54, 18]. The model describes a hydrophilic polymer chain wafting in a water solution containing randomly placed hydrophobic molecules (impurities) that repel the individual monomers in the polymer chain. Due to thermal fluctuations and the random positions of impurities, the shape of the polymer

chain is best understood as a random object. The statistical mechanical model for a directed polymer assumes that the shape of the polymer can be described by a directed path $p \in \Pi_{M,N}$, thus suppressing entanglement and U-turns. The presence, or strength, of an impurity at site (i, j) is described by a random variable $X(i, j)$, and the energy of a path $p \in \Pi_{M,N}$ is given by

$$-\beta \sum_{(i,j)\in p} X(i,j), \tag{1.3}$$

where $\beta = 1/T > 0$ is the inverse temperature. The typical shape of a polymer is one that minimizes (1.3). Of interest is the quenched polymer distribution on paths defined by

$$Q(p; M, N) = \frac{1}{Z(M,N)} \exp\left(\beta \sum_{(i,j)\in p} X(i,j)\right), \tag{1.4}$$

where $p \in \Pi_{M,N}$ and the normalization factor $Z(M, N)$ is called the *partition function*, and is given by

$$Z(M,N) = \sum_{p\in\Pi_{M,N}} \exp\left(\beta \sum_{(i,j)\in p} X(i,j)\right). \tag{1.5}$$

In the zero-temperature limit, i.e., $\beta \to \infty$, the quenched polymer distribution concentrates around paths maximizing (1.3), and we formally have

$$\lim_{\beta\to\infty} \frac{1}{\beta} \log\left(Z(M,N)\right) = \max_{p\in\Pi_{M,N}} \sum_{(i,j)\in p} X(i,j) = L(M,N),$$

Directed polymers are related to several other stochastic models for growing surfaces, such as directed invasion percolation, ballistic deposition, polynuclear growth, and low temperature Ising models [62].

DLPP with *i.i.d.* exponential weights $X(i, j)$ is also equivalent to the totally asymmetric simple exclusion process (TASEP), which is an important stochastic interacting particle system [35, 82], and to randomly growing Young diagrams [58, 94, 83].

The dynamics of TASEP involve a particle configuration on the lattice $\mathbb{Z}$, evolving in time, with the dynamical rule that a particle jumps to the right after an exponential waiting time if the right neighboring site is empty. The correspondence between DLPP and TASEP proceeds via the following stochastic corner growth model: Partition $\mathbb{R}^2$ into squares defined by the edges of the lattice $\mathbb{Z}^2$. Imagine that at time $t = 0$, all the squares in $[0, \infty)^2$ are colored white, while the remaining squares are colored black. For each $(i, j) \in \mathbb{N}^2$, assign a passage time random variable $X(i, j)$ to square (i, j). The dynamic rule governing the growth process is the following: A white square at location (i, j) is colored black exactly $X(i, j)$ time units after both its south and west neighbors become black. The time until square (M, N) is colored black is exactly $L(M, N)$—the last passage time from $(1, 1)$ to (M, N)—and the set of all black squares is a randomly growing Young diagram.

There is a one-to-one correspondence between TASEP configurations, and configurations of black and white squares in the corner growth model. The idea is that when a white square is colored black, it corresponds to a particle jumping from a site j to its necessarily vacant neighbor $j + 1$. The explicit correspondence is as follows: For every edge separating a white and black square, assign a value of 1 to vertical edges, and a value of 0 to horizontal edges. The TASEP configuration corresponds exactly to reading these binary values sequentially from $(1, \infty)$ to $(\infty, 1)$. We give this correspondence more rigorously in Section 5.1.2 (see Figure 5.2). There are further applications of DLPP in queueing theory [5, 45], and the model is also related to greedy lattice animals [70].

One quantity of interest in DLPP is the time constant, U, given by

$$U(x) := \lim_{N \to \infty} \frac{1}{N} L\left(\lfloor Nx \rfloor\right). \tag{1.6}$$

The exact form of U is known for *i.i.d.* geometric weights [58], and *i.i.d.* exponential weights [82], and is given by

$$U(x) = \mu(x_1 + x_2) + 2\sigma\sqrt{x_1 x_2}, \tag{1.7}$$

where μ and σ^2 are the mean and variance, respectively, of the either geometric or exponential weights. For more general distributions, Martin [71] showed that U is continuous on $[0, \infty)^2$ and gave the following asymptotics at the boundary:

$$U(1, \alpha) = \mu + 2\sigma\sqrt{\alpha} + o(\sqrt{\alpha}).$$

The fluctuations of L for geometric and exponential weights are non-Gaussian, and instead follow the Tracy-Widom distribution asymptotically [58]. It is an open problem to determine $U(x)$ and the fluctuations of L for weights other than geometric and exponential.

Several authors have considered the DLPP problem with independent weights $X(i, j)$ that are either geometric or exponential, but not identically distributed and instead have a macroscopic inhomogeneity. For example, in the exponential DLPP setting, one would assume that $X(i, j)$ are independent and exponentially distributed with mean $\lambda(i, N^{-1}, jN^{-1})$ where $\lambda : \mathbb{R}^d \to \mathbb{R}$ is the macroscopic inhomogeneity and N is the asymptotic parameter. In the exponential case with continuous λ, Rolla and Teixeira [81] showed that U has a variational interpretation. Their result is in many ways analogous to the variational problem for the longest chain problem [31]. Macroscopic inhomogeneities have also been considered for TASEP [43], and for other similar growth models [80]. In particular, Georgiu et al. [43] proved a hydrodynamic limit for TASEP with a spatially (but not temporally) inhomogeneous jump rate c, which may admit discontinuities. Their result gives the limiting density profile in terms of a variational problem, and they connected this to a conservation law in the

special case that the rate $c(s)$ is piecewise constant with one jump, i.e.,

$$c(s) = \begin{cases} c_1, & s \leq 0 \\ \\ c_2, & s > 0. \end{cases}$$

In the context of exponential DLPP, this would be equivalent to assuming that the macroscopic mean $\lambda : [0,\infty)^2 \to [0,\infty)$ is given by $\lambda(x) = c_1^{-1}$ for $x_1 >= x_2$ and $\lambda(x) = c_2^{-1}$ otherwise.

1.4 Summary

For each of the aforementioned combinatorial problems, we have discovered continuum limits that correspond to solving certain Hamilton-Jacobi equations in the viscosity sense. For non-dominated sorting, the solution of the Hamilton-Jacobi equation gives the asymptotic shapes of the Pareto fronts, which are the level sets of the longest chain function; while for directed last passage percolation (DLPP), it gives the limiting time constant in the presence of macroscopic inhomogeneities. Much of this dissertation is devoted to rigorously proving these continuum limits, while the rest is devoted to applications of these new results.

This dissertation is organized as follows: In Chapter II, we give some statistical results concerning the convexity (or lack-thereof) of the first Pareto front. In Chapter III, we give the continuum limit for non-dominated sorting and rigorously prove the convergence. In Chapter IV, we explore applications of this continuum limit. In particular, we prove convergence of a fast numerical scheme and show how it can be used to perform fast approximate non-dominated sorting. Finally, in Chapter V we prove an analogous continuum limit for DLPP and show how to efficiently compute the asymptotic shapes of optimal paths via dynamic programming.

CHAPTER II

Statistical properties of Pareto fronts

The distribution of the number of points on the first Pareto front was first studied by Barndorff-Nielsen and Sobel [13]. The problem has garnered much attention since. Bai et al. [6] and Hwang and Tsai [53] provide good surveys of recent results. We study here the statistical properties of the Pareto-optimal points that are missed by linear scalarization. Special cases of the results in this section were published in [51].

2.1 Asymptotic convexity of Pareto fronts

Let $X_1, \ldots, X_n$ be *i.i.d.* on $[0,1]^d$ with density $f : [0,1]^d \to [0,\infty)$, and let $\mathcal{F}_1, \mathcal{F}_2, \ldots$ denote the Pareto fronts associated with $\mathcal{X}_n = \{X_1, \ldots, X_n\}$. Let $h_n : [0,1]^d \to \mathbb{R}$ denote the Pareto depth function defined by

$$h_n(x) = \max\{i \in \mathbb{N} : \mathcal{F}_i \leqq x\}, \tag{2.1}$$

where for simplicity we set $\mathcal{F}_0 = \{(-1, \ldots, -1)\}$, and we write $\mathcal{F}_i \leqq x$ if there exists $y \in \mathcal{F}_i$ such that $y \leqq x$. The function h_n is a (random) piecewise constant function that "counts" the Pareto fronts associated with $X_1, \ldots, X_n$.

It is immediate, based on the discussion in Chapter I, that we have the following alternative characterization of h_n:

Proposition II.1. *Suppose $f \in L^1([0,1]^d)$ is a density. Then $h_n(x) = u_n(x)$ with probability one for all $x \in [0,1]^d$.*

It is well-known [32] that Pareto methods outperform more traditional linear scalarization methods when the Pareto fronts are non-convex. Qualitatively, there are two types of non-convexities in Pareto fronts: 1) Non-convexities in the asymptotic shapes of the fronts induced by the density f and 2) local small-scale non-convexities due to the random positions of the samples $X_1, \ldots, X_n$. We study here non-convexities of the first kind, for which we make the following definition:

Definition II.2. Given a density $f : [0,1]^d \to [0,\infty)$, we say that f yields *macroscopically convex* Pareto fronts if for $X_1, \ldots, X_n$ drawn *i.i.d.* from f we have that the almost sure limit $U(x) := \lim_{n\to\infty} n^{-\frac{1}{d}} h_n(x)$ exists for all x and $U : [0,1]^d \to \mathbb{R}$ is quasiconcave.

Recall that U is said to be *quasiconcave* if the super level sets

$$\left\{ x \in [0,1]^d \,:\, U(x) \geq a \right\}$$

are convex for all $a \in \mathbb{R}$. Since the Pareto fronts are encoded into the level sets of h_n, the asymptotic shape of the Pareto fronts is dictated by the level sets of the function U. Hence the fronts are asymptotically convex on a macroscopic scale exactly when U is quasiconcave.

We now give a partial characterization of densities f that yield macroscopically convex Pareto fronts.

Theorem II.3. *Let $f : [0,1]^d \to (0,\infty)$ be a continuous, log-concave, and separable density, i.e., $f(x) = f_1(x_1) \cdots f_d(x_d)$. Then f yields macroscopically convex Pareto fronts.*

Proof. We denote by $F : [0,1]^d \to \mathbb{R}$ the cumulative distribution function (CDF) associated with the density f, which is defined by

$$F(x) = \int_0^{x_1} \cdots \int_0^{x_d} f(y_1, \ldots, y_d)\, dy_1 \cdots dy_d. \tag{2.2}$$

Let $X_1, \ldots, X_n$ be *i.i.d.* with density f, and let h_n denote the associated Pareto depth function, and u_n the associated longest chain function.

We claim that

$$n^{-\frac{1}{d}} u_n(x) \longrightarrow U(x) \text{ almost surely as } n \to \infty, \tag{2.3}$$

where $U(x) = c_d F(x)^{\frac{1}{d}}$, and c_d is a positive constant. In fact, the convergence is actually uniform on $[0,1]^d$ with probability one, but this is not necessary for the proof. To see this: Define $\Phi : [0,1]^d \to [0,1]^d$ by

$$\Phi(x) = \left(\int_0^{x_1} f_1(t)\, dt, \ldots, \int_0^{x_d} f_d(t)\, dt \right).$$

Since f is continuous and strictly positive, $\Phi : [0,1]^d \to [0,1]^d$ is a C^1-diffeomorphism. Setting $Y_i = \Phi(X_i)$, we easily see that $Y_1, \ldots, Y_d$ are independent and uniformly distributed on $[0,1]^d$. It is also easy to see that Φ preserves the partial order $\leqq$, i.e.,

$$x \leqq z \iff \Phi(x) \leqq \Phi(z).$$

Let $x \in [0,1]^d$, set $y = \Phi(x)$, and define $\mathcal{Y}_n = \Phi(\mathcal{X}_n)$. By our above observations we have

$$u_n(x) = \max\{\ell \in \mathbb{N} : \exists\, y_1 \leqq \cdots \leqq y_\ell \leqq y \text{ in } \mathcal{Y}_n\}.$$

Let $i_1 < \cdots < i_N$ denote the indices of the random variables among $Y_1, \ldots, Y_n$ that are less than or equal to y and set $Z_k = Y_{i_k}$ for $k = 1, \ldots, N$. Note that N is binomially distributed with parameter $p := F(x)$ and that $u_n(x)$ is the length of the longest chain among N uniformly distributed points in the hypercube $\{z \in [0,1]^d :$

$z \leqq y\}$. By [17, Remark 1] we have $N^{-\frac{1}{d}} u_n(x) \to c_d$ almost surely as $n \to \infty$ where $c_d < e$ are dimensional constants. Since $n^{-1}N \to p$ almost surely as $n \to \infty$, we have

$$n^{-\frac{1}{d}} u_n(x) = \left(n^{-\frac{1}{d}} N^{\frac{1}{d}}\right) N^{-\frac{1}{d}} u_n(x) \to c_d p^{\frac{1}{d}}$$

almost surely as $n \to \infty$. The proof of (2.3) is completed by recalling Proposition II.1.

In the context of Definition II.2, we have $U(x) = c_d F(x)^{\frac{1}{d}}$. Hence U is quasiconcave if and only if the cumulative distribution function F is quasiconcave. A sufficient condition for quasiconcavity of F is log-concavity of f [79], which completes the proof. □

Theorem II.3 indicates that Pareto methods are largely redundant when f is a log-concave separable density.

It would be very interesting to extend Theorem II.3 to arbitrary non-separable density functions f. When f is non-separable there is no simple integral expression like (2.2) for U. We show in Chapter III that U is instead characterized as the viscosity solution of a Hamilton-Jacobi partial differential equation. This makes the non-separable case substantially more difficult.

2.2 Microscopic non-convexities

We now present some results concerning microscopic non-convexities in the Pareto fronts. These are non-convexities that exist even when the fronts are macroscopically convex, and are due to the random positions of the samples $X_1, \ldots, X_n$.

By non-convexities, we mean Pareto-optimal points that cannot be obtained by linear scalarization, as defined in Chapter I. It is well known [32] that linear scalarization will identify exactly those Pareto-optimal points on the boundary of the convex

hull of

$$\mathcal{G}^n := \bigcup_{x \in \mathcal{F}_1^n} (x + \mathbb{R}_+^d),$$

where $\mathbb{R}_+^d = (0, \infty)^d$, and $\mathcal{F}_1^n$ denotes the first Pareto front of the set $\mathcal{X}_n = \{X_1, \ldots, X_n\}$. Although this is a common motivation for Pareto optimization methods, there are, to the best of our knowledge, no results in the literature regarding how many points on the Pareto front are missed by scalarization.

We define

$$\mathcal{L}^n = \bigcup_{\alpha \in \mathbb{R}_+^d} \operatorname*{argmin}_{x \in \mathcal{X}_n} \left\{ \sum_{i=1}^{d} \alpha_i x_i \right\}.$$

The subset $\mathcal{L}^n \subset \mathcal{F}_1^n$ contains all Pareto-optimal points that can be obtained by some selection of non-negative weights for linear scalarization. Let K_n denote the cardinality of $\mathcal{F}_1^n$, and let L_n denote the cardinality of $\mathcal{L}^n$.

When $X_1, \ldots, X_n$ are uniformly distributed on the unit hypercube, Barndorff-Nielsen and Sobel [13] showed that

$$E(K_n) = \frac{n}{(d-1)!} \int_0^1 (1-x)^{n-1} (-\log x)^{d-1} \, dx,$$

from which one can easily obtain the asymptotics

$$E(K_n) = \frac{(\log n)^{d-1}}{(d-1)!} + O((\log n)^{d-2}).$$

Many more recent works have studied the variance of K_n and have proven central limit theorems for K_n. All of these works assume that $X_1, \ldots, X_n$ are uniformly distributed on $[0,1]^d$. For a summary, see [6] and [53]. Other works have studied K_n for more general distributions on domains that have smooth "non-horizontal" boundaries near the Pareto front [14] and for multivariate normal distributions on $\mathbb{R}^d$ [55]. The "non-horizontal" condition excludes hypercubes. To the best of our knowledge there are no results on the asymptotics of K_n for non-uniformly distributed

points on the unit hypercube. This is of great importance as it is impractical in multi-criteria optimization (or anomaly detection) to assume that the coordinates of the points are independent. Typically the coordinates of $X_i \in \mathbb{R}^d$ are the images of the *same* feasible solution under several *different* criteria, which will not in general be independent.

Here we develop results on the size of the gap between the number of items L_n discoverable by scalarization compared to the number of items K_n discovered on the Pareto front. The larger the gap, the more suboptimal scalarization is relative to Pareto depth analysis. Since $x \in \mathcal{L}^n$ if and only if x is on the boundary of the convex hull of $\mathcal{G}^n$, the size of $\mathcal{L}^n$ is related to the convexity (or lack thereof) of the Pareto front.

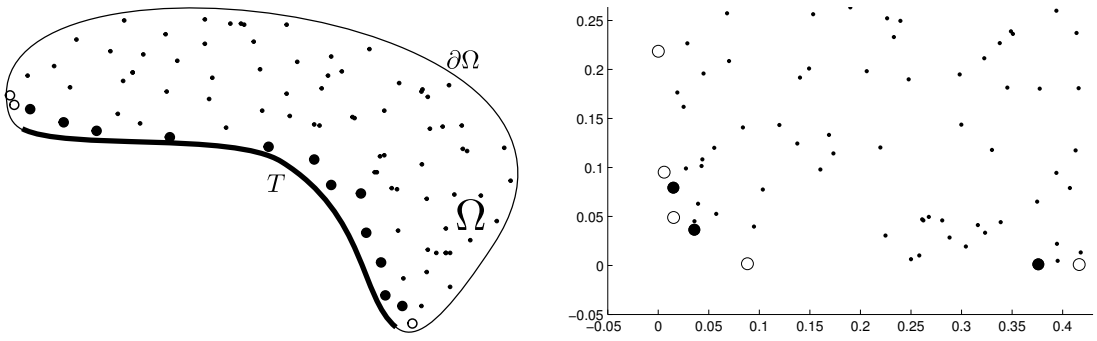
Suppose first that $X_1, \ldots, X_n$ are distributed on some domain $\Omega \subset \mathbb{R}^d$ with a continuous density function $f : \overline{\Omega} \to \mathbb{R}$ that is strictly positive on $\overline{\Omega}$. Let $T \subset \partial\Omega$ be a portion of the boundary of Ω such that

$$\inf_{z \in T} \min(\nu_1(z), \ldots, \nu_d(z)) > 0,$$

and

$$\{y \in \overline{\Omega} \,:\, \forall i \;\; y_i \leq x_i\} = \{x\}, \quad \text{for all} \;\; x \in T,$$

where $\nu : \partial\Omega \to \mathbb{R}^d$ is the unit inward normal to $\partial\Omega$. The conditions on T guarantee that a portion of the first Pareto front will concentrate near T as $n \to \infty$. If we suppose that T is contained in the interior of the convex hull of Ω, then points on the portion of the Pareto front near T cannot be obtained by linear scalarization, as they are on a non-convex portion of the front. Such non-convexities are a direct result of the geometry of the domain Ω and are depicted in Fig. 2.1(a). As before,



(a) (b)

Figure 2.1: (a) Non-convexities in the Pareto front induced by the geometry of the domain Ω. (b) Non-convexities due to randomness in the points. In each case, the larger points are Pareto-optimal, and the large black points *cannot* be obtained by scalarization.

we call such non-convexities *macroscopic non-convexities*. We showed in [51] that

$$E(K_n - L_n) \geq \gamma n^{\frac{d-1}{d}} + O(n^{\frac{d-2}{d}}),$$

as $n \to \infty$, where γ is a positive constant given by

$$\gamma = \frac{1}{d}(d!)^{\frac{1}{d}}\Gamma\left(\frac{1}{d}\right)\int_T f(z)^{\frac{d-1}{d}}(\nu_1(z)\cdots\nu_d(z))^{\frac{1}{d}}dz.$$

This gives a lower bound on the asymptotic expectation of the number of Pareto-optimal points missed by linear scalarization due to macroscopic non-convexities. It has recently come to our attention that a stronger result was proven previously by Baryshnikov and Yukich [14] in an unpublished manuscript.

Here, we study *microscopic* non-convexities in the Pareto front. These non-convexities are strictly due to randomness in the positions of the points and occur even when the domain Ω is convex (see Fig. 2.1(b) for a depiction of such non-convexities). In the following, we assume that $X_1, \ldots, X_n$ are *i.i.d.* on the unit hypercube $[0,1]^d$ with a bounded density function $f : [0,1]^d \to \mathbb{R}^d$ that is continuous at the origin and strictly positive on $[0,1]^d$. Under these assumptions on f, it turns out that the asymptotics of $E(K_n)$ and $E(L_n)$ are independent of f. Hence our

results are applicable to a wide range of problems without the need to know detailed information about the density f.

Our first result is

Theorem II.4. *Assume $f : [0,1]^d \to [\sigma, M]$ is continuous at the origin, and $0 < \sigma < M < \infty$. Then*

$$E(K_n) \sim c_{n,d} := \frac{(\log n)^{d-1}}{(d-1)!} \text{ as } n \to \infty.$$

We give the proof of Theorem II.4 after some preliminary results. Our second result concerns $E(L_n)$. We are not able to get the exact asymptotics of $E(L_n)$, so we provide upper and lower asymptotic bounds.

Theorem II.5. *Assume $f : [0,1]^d \to [\sigma, M]$ is continuous at the origin, and $0 < \sigma < M < \infty$. Then*

$$\tfrac{d!}{d^d} c_{n,d} + o((\log n)^{d-1}) \le E(L_n) \le \tfrac{3d-1}{4d-2} c_{n,d} + o((\log n)^{d-1})$$

as $n \to \infty$.

The proof of Theorem II.5 is also given after some preliminary results. Combining Theorems II.4 and II.5, we arrive at our main result:

Theorem II.6 (Scalarization Gap Theorem)**.** *Assume $f : [0,1]^d \to [\sigma, M]$ is continuous at the origin, and $0 < \sigma < M < \infty$. Then*

$$\frac{d-1}{4d-2} c_{n,d} + o((\log n)^{d-1}) \le E(K_n - L_n) \le \left(1 - \frac{d!}{d^d}\right) c_{n,d} + o((\log n)^{d-1}), \tag{2.4}$$

as $n \to \infty$.

The scalarization gap theorem shows that the fraction of Pareto-optimal points that *cannot* be obtained by linear scalarization is at least $\frac{d-1}{4d-2}$.

2.2.1 Proofs

We first present a general result on the expectation of K_n. Let $F : [0,1]^d \to \mathbb{R}$ denote the cumulative distribution function of f, defined by

$$F(x) = \int_0^{x_1} \cdots \int_0^{x_d} f(y_1, \ldots, y_d)\, dy_1 \cdots dy_d.$$

Proposition II.7. *For any $n \geq 1$ we have*

$$E(K_n) = n \int_{[0,1]^d} f(x)\,(1 - F(x))^{n-1}\, dx.$$

Proof. Let E_i be the event that $X_i \in \mathcal{F}^n$ and let χ_{E_i} be indicator random variables for E_i. Then

$$E(K_n) = E\left(\sum_{i=1}^{n} \chi_{E_i}\right) = \sum_{i=1}^{n} P(E_i) = nP(E_1).$$

Conditioning on X_1 we obtain

$$E(K_n) = n \int_{[0,1]^d} f(x) P(E_1 \,|\, X_1 = x) dx.$$

Noting that $P(E_1 \,|\, X_1 = x) = (1 - F(x))^{n-1}$ completes the proof. □

The following simple proposition is essential in the proofs of Theorem II.4 and II.5:

Proposition II.8. *Let $0 < \delta \leq 1$ and $a > 0$. For $a \leq \delta^{-d}$ we have*

$$n \int_{[0,\delta]^d} (1 - a x_1 \cdots x_d)^{n-1}\, dx = \frac{c_{n,d}}{a} + O((\log n)^{d-2}), \tag{2.5}$$

and for $a \leq 1$ we have

$$n \int_{[0,1]^d \setminus [0,\delta]^d} (1 - a x_1 \cdots x_d)^{n-1}\, dx = O((\log n)^{d-2}). \tag{2.6}$$

Proof. We will give a sketch of the proof as similar results are well-known [6]. Assume $\delta = 1$ and let Q_n denote the quantity on the left hand side of (2.5). Making the change of variables $y_i = x_i$ for $i = 1, \ldots, d-1$ and $t = x_1 \cdots x_d$, we see that

$$Q_n = n \int_0^1 \int_t^1 \int_{\frac{t}{y_{d-1}}}^1 \cdots \int_{\frac{t}{y_2 \cdots y_{d-1}}}^1 \frac{(1-at)^{n-1}}{y_1 \cdots y_{d-1}} \, dy_1 \cdots dy_{d-1} dt.$$

By computing the inner $d-1$ integrals we find that

$$Q_n = \frac{n}{(d-1)!} \int_0^1 (-\log t)^{d-1} (1-at)^{n-1} dt,$$

from which the asymptotics (2.5) can be easily obtained by another change of variables $u = nat$, provided $a \leq 1$. For $0 < \delta < 1$, we make the change of variables $y = x/\delta$ to find that

$$Q_n = \delta^d n \int_{[0,1]^d} (1 - a\delta^d y_1 \cdots y_d)^{n-1} \, dy.$$

We can now apply the above result provided $a\delta^d \leq 1$. The asymptotics in (2.5) show that

$$n \int_{[0,1]^d} (1 - ax_1 \cdots x_d)^{n-1} \, dx = n \int_{[0,\delta]^d} (1 - ax_1 \cdots x_d)^{n-1} \, dx + O((\log n)^{d-2}),$$

when $a \leq 1$, which gives the second result (2.6). □

We now give the proof of Theorem II.4.

Proof. Let $\varepsilon > 0$ and choose $\delta > 0$ such that

$$f(0) - \varepsilon \leq f(x) \leq f(0) + \varepsilon \quad \text{for any } x \in [0, \delta]^d,$$

and $f(0) < \delta^{-d}$. Since $\sigma \leq f \leq M$, we have that $F(x) \geq \sigma x_1 \cdots x_d$ for all $x \in [0,1]^d$. Since f is a probability density on $[0,1]^d$, we must have $\sigma \leq 1$. Since $\sigma > 0$, we can

apply Proposition II.8 to find that

$$n\int_{[0,1]^d\setminus[0,\delta]^d} f(x)(1-F(x))^{n-1}\,dx \le Mn\int_{[0,1]^d\setminus[0,\delta]^d}(1-\sigma x_1\cdots x_d)^{n-1}\,dx$$

$$= O((\log n)^{d-2}). \tag{2.7}$$

For $x\in[0,\delta]^d$, we have

$$(f(0)-\varepsilon)x_1\cdots x_d \le F(x) \le (f(0)+\varepsilon)x_1\cdots x_d.$$

Combining this with Proposition II.8, and the fact that $f(0)-\varepsilon<\delta^{-d}$ we have

$$n\int_{[0,\delta]^d} f(x)(1-F(x))^{n-1}\,dx$$

$$\le (f(0)+\varepsilon)n\int_{[0,\delta]^d}(1-(f(0)-\varepsilon)x_1\cdots x_d)^{n-1}\,dx$$

$$= \frac{f(0)+\varepsilon}{f(0)-\varepsilon}\cdot c_{n,d} + O((\log n)^{d-2}). \tag{2.8}$$

Combining (2.7) and (2.8) with Proposition (II.7) we have

$$E(K_n) \le \frac{f(0)+\varepsilon}{f(0)-\varepsilon}\cdot c_{n,d} + O((\log n)^{d-2}).$$

It follows that

$$\limsup_{n\to\infty} c_{n,d}^{-1}E(K_n) \le \frac{f(0)+\varepsilon}{f(0)-\varepsilon}.$$

By a similar argument we can obtain

$$\liminf_{n\to\infty} c_{n,d}^{-1}E(K_n) \ge \frac{f(0)-\varepsilon}{f(0)+\varepsilon}.$$

Since $\varepsilon>0$ was arbitrary, we see that

$$\lim_{n\to\infty} c_{n,d}^{-1}E(K_n) = 1. \qquad \square$$

The proof of Theorem II.5 is split into the following two lemmas. It is well-known, and easy to see, that $x\in\mathcal{L}^n$ if and only if $x\in\mathcal{F}^n$ and x is on the boundary of the convex hull of $\mathcal{G}^n$ [32]. This fact will be used in the proof of Lemma II.9.

Lemma II.9. *Assume $f : [0,1]^d \to \mathbb{R}$ is continuous at the origin and there exists $\sigma, M > 0$ such that $\sigma \le f \le M$. Then*

$$E(L_n) \le \frac{3d-1}{4d-2} \cdot c_{n,d} + o((\log n)^{d-1}) \quad \text{as} \quad n \to \infty.$$

Proof. Let $\varepsilon > 0$ and choose $0 < \delta < \frac{1}{2}$ so that

$$f(0) - \varepsilon \le f(x) \le f(0) + \varepsilon \quad \text{for any } x \in [0, 2\delta]^d, \tag{2.9}$$

and $3f(0) \le \delta^{-d}$. As in the proof of Proposition II.7 we have $E(L_n) = nP(X_1 \in \mathcal{L}^n)$, so conditioning on X_1 we have

$$E(L_n) = n \int_{[0,1]^d} f(x) P(X_1 \in \mathcal{L}^n \,|\, X_1 = x)\, dx.$$

As in the proof of Theorem II.4, we have

$$\begin{aligned} n \int_{[0,1]^d \setminus [0,\delta]^d} f(x) P(X_1 \in \mathcal{L}^n \,|\, X_1 = x)\, dx &\le n \int_{[0,1]^d \setminus [0,\delta]^d} f(x)(1 - F(x))^{n-1}\, dx \\ &= O((\log n)^{d-2}), \end{aligned}$$

and hence

$$E(L_n) = n \int_{[0,\delta]^d} f(x) P(X_1 \in \mathcal{L}^n \,|\, X_1 = x)\, dx + O((\log n)^{d-2}). \tag{2.10}$$

Fix $x \in [0,\delta]^d$ and define $A = \{y \in [0,1]^d : \forall i,\ y_i \le x_i\}$ and

$$B_i = \left\{ y \in [0,1]^d \ : \ \forall j \ne i,\ y_j < x_j \text{ and } x_i < y_i < 2x_i - \frac{x_i}{x_j} y_j \right\},$$

for $i = 1, \dots, d$, and note that $B_i \subset [0, 2\delta]^d$ for all i. See Fig. 2.2 for an illustration of these sets for $d = 3$.

We claim that if at least two of $B_1, \dots, B_d$ contain samples from $X_2, \dots, X_n$, and $X_1 = x$, then $X_1 \notin \mathcal{L}^n$. To see this, assume without loss of generality that B_1 and B_2 are nonempty and let $y \in B_1$ and $z \in B_2$. Set

$$\widetilde{y} = \left(y_1, 2x_2 - \frac{x_2}{x_1} y_1, x_3, \dots, x_d \right)$$

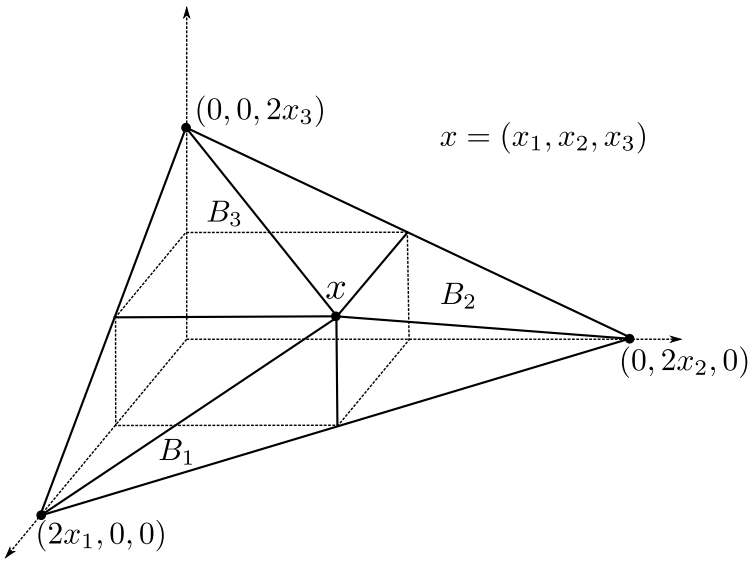


Figure 2.2: Depiction of the sets B_1, B_2 and B_3 from the proof of Lemma II.9 in the case that $d = 3$.

$$\widetilde{z} = \left(2x_1 - \frac{x_1}{x_2} z_2, z_2, x_3, \ldots, x_d\right).$$

By the definitions of B_1 and B_2 we see that $y_i \leq \widetilde{y}_i$ and $z_i \leq \widetilde{z}_i$ for all i, hence $\widetilde{y}, \widetilde{z} \in \mathcal{G}_n$. Let $\alpha \in (0, 1)$ such that

$$\alpha y_1 + (1 - \alpha)\left(2x_1 - \frac{x_1}{x_2} z_2\right) = x_1.$$

A short calculation shows that $x = \alpha\widetilde{y} + (1 - \alpha)\widetilde{z}$ which implies that x is in the interior of the convex hull of $\mathcal{G}_n$, proving the claim.

Let E denote the event that at most one of $B_1, \ldots, B_d$ contains a sample from $X_2, \ldots, X_n$, and let F denote the event that A contains no samples from $X_2, \ldots, X_n$. Then by the observation above we have

$$P(X_1 \in \mathcal{L}^n \,|\, X_1 = x) \leq P(E \cap F \,|\, X_1 = x) = P(E \cap F). \tag{2.11}$$

For $i = 1, \ldots, d$, let E_i denote the event that B_i contains no samples from $X_2, \ldots, X_n$. It is not hard to see that

$$E = \bigcup_{i=1}^{d} \left(\bigcap_{j \neq i} E_j \setminus \bigcap_{j} E_j\right) \bigcup \left(\bigcap_{j} E_j\right).$$

Furthermore, the events in the unions above are mutually exclusive (disjoint) and $\cap_j E_j \subset \cap_{j\neq i} E_j$ for $i = 1, \ldots, d$. It follows that

$$\begin{aligned}
&P(E \cap F) \\
&= \sum_{i=1}^{d} \left(P\left(\cap_{j\neq i} E_j \cap F\right) - P\left(\cap_j E_j \cap F\right)\right) + P\left(\cap_j E_j \cap F\right) \\
&= \sum_{i=1}^{d} P\left(\cap_{j\neq i} E_j \cap F\right) - (d-1) P\left(\cap_j E_j \cap F\right) \\
&= \sum_{i=1}^{d} \left(1 - F(x) - \int_{\cup_{j\neq i} B_j} f(y)\, dy\right)^{n-1} - (d-1)\left(1 - F(x) - \int_{\cup_j B_j} f(y)\, dy\right)^{n-1}. \qquad (2.12)
\end{aligned}$$

A simple computation shows that $|B_j| = \frac{1}{d} x_1 \cdots x_d$ for $j = 1, \ldots, d$. Since $A, B_i \subset [0, 2\delta]^d$, we have by (2.9) that

$$(f(0) - \varepsilon) x_1 \cdots x_d \leq F(x) \leq (f(0) + \varepsilon) x_1 \cdots x_d,$$

and

$$\frac{1}{d}(f(0) - \varepsilon) x_1 \cdots x_d \leq \int_{B_j} f(y)\, dy \leq \frac{1}{d}(f(0) + \varepsilon) x_1 \cdots x_d.$$

Inserting these into (2.12) and combining with (2.11) we have

$$\begin{aligned}
P(X_1 \in \mathcal{L}^n \,|\, X_1 = x) \leq d \Bigg(1 - \frac{2d-1}{d}(f(0) - \varepsilon) x_1 \cdots x_d\Bigg)^{n-1} \\
- (d-1)\left(1 - 2(f(0) + \varepsilon) x_1 \cdots x_d\right)^{n-1}.
\end{aligned}$$

We can now insert this into (2.10) and apply Proposition II.8 (since $3f(0) \leq \delta^{-d}$) to obtain

$$E(L_n) \leq \left(\frac{d^2}{2d-1} \frac{f(0) + \varepsilon}{f(0) - \varepsilon} - \frac{d-1}{2} \frac{f(0) - \varepsilon}{f(0) + \varepsilon}\right) c_{n,d} + O((\log n)^{d-2}).$$

Since $\varepsilon > 0$ was arbitrary, we find that

$$\limsup_{n\to\infty} \; c_{n,d}^{-1} E(L_n) \leq \left(\frac{d^2}{2d-1} - \frac{d-1}{2}\right) = \frac{3d-1}{4d-2}. \qquad \square$$

Lemma II.10. *Assume $f : [0,1]^d \to \mathbb{R}$ is continuous and there exists $\sigma, M > 0$ such that $\sigma \le f \le M$. Then*

$$E(L_n) \ge \frac{d!}{d^d} \cdot c_{n,d} + o((\log n)^{d-1}) \quad \text{as} \quad n \to \infty.$$

Proof. Let $\varepsilon > 0$ and choose $0 < \delta < 1/d$ so that

$$f(0) - \varepsilon \le f(x) \le f(0) + \varepsilon \quad \text{for} \quad x \in [0, d\delta]^d, \tag{2.13}$$

and

$$\frac{d^d}{d!}(f(0) + \varepsilon) \le \delta^{-d}. \tag{2.14}$$

As in the proof of Lemma II.9 we have

$$E(L_n) = n \int_{[0,\delta]^d} f(x) P(X_1 \in \mathcal{L}^n \,|\, X_1 = x)\, dx + O((\log n)^{d-2}). \tag{2.15}$$

Fix $x \in (0,\delta)^d$, set $\nu = \left(\frac{1}{x_1}, \dots, \frac{1}{x_d}\right)$ and

$$A = \big\{ y \in [0,1]^d \;\; | \;\; y \cdot \nu \le x \cdot \nu \big\}.$$

Note that A is a simplex with an orthogonal corner at the origin and side lengths $d \cdot x_1, \dots, d \cdot x_d$. A simple computation shows that $|A| = \frac{d^d}{d!} x_1 \cdots x_d$. By (2.13) we have

$$\int_A f(y)\, dy \le (f(0) + \varepsilon)|A| = \frac{d^d}{d!}(f(0) + \varepsilon) x_1 \cdots x_d.$$

It is easy to see that if A is empty and $X_1 = x$ then $X_1 \in \mathcal{L}^n$, hence

$$\begin{aligned} P(X_1 \in \mathcal{L}^n \,|\, X_1 = x) &\ge \left(1 - \int_A f(y)\, dy\right)^{n-1} \\ &\ge \left(1 - \frac{d^d}{d!}(f(0) + \varepsilon) x_1 \cdots x_d\right)^{n-1}. \end{aligned}$$

Inserting this into (2.15) and noting (2.14), we can apply Proposition II.8 to obtain

$$E(L_n) \ge \frac{d!}{d^d} \frac{f(0) - \varepsilon}{f(0) + \varepsilon} c_{n,d} + O((\log n)^{d-2}),$$

and hence

$$\limsup_{n\to\infty} \; c_{n,d}^{-1} E(L_n) \geq \frac{d!}{d^d}. \qquad \square$$

CHAPTER III

Continuum limit for non-dominated sorting

Let $X_1, \ldots, X_n$ be *i.i.d.* random variables on $\mathbb{R}^d$ with density function $f \in L^1(\mathbb{R}^d)$. The points form a partially ordered set $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ under the partial order

$$x \leqq y \iff x_i \leq y_i \quad \text{for} \quad i = 1, \ldots, d. \tag{3.1}$$

Let $u_n(x)$ denote the length of a longest chain[1] in $\mathcal{X}_n$ consisting of points less than or equal to x.

The goal of this chapter is to study the asymptotics of u_n, and hence the asymptotics of non-dominated sorting—the link between the two problems was discussed in Chapter I. Our main result, Theorem III.2, states that $n^{-\frac{1}{d}} u_n$ converges almost surely to a continuous function U that can be characterized as the viscosity solution of a Hamilton-Jacobi equation. Our proof is based on linking the asymptotics of u_n to a variational problem, which is a generalization of the variational problem discovered by Deuschel and Zeitouni [31] to higher dimensions. The Hamilton-Jacobi equation satisfied by U is the Hamilton-Jacobi-Bellman equation [10] for the corresponding variational problem. We describe our main result in Section 3.1, and postpone the proofs to Sections 3.3 and 3.5. Most of the results in this chapter have been published

[1] A *chain* is a totally ordered subset of $\mathcal{X}_n$.

in the SIAM Journal on Mathematical Analysis [20]. Theorems III.11 and III.16 are generalizations of the corresponding theorems from [20].

3.1 Main result

For $x, y \in \mathbb{R}^d$, we write $x \leq y$ if $x \leqq y$ and $x \neq y$. When $x_i < y_i$ for $i = 1, \ldots, d$, we write $x < y$, and we set $\mathbb{R}^d_+ = \{x \in \mathbb{R}^d \,:\, x > 0\}$. We will always assume $d \geq 2$. For $s, t \in \mathbb{R}$, $s \leq t$ and $s < t$ will retain their usual definitions. Let $\Omega \subset \mathbb{R}^d$ and let $f : \mathbb{R}^d \to [0, \infty)$. We place the following assumptions on f and Ω:

(H0) There exists an open and bounded set $\Omega \subset \mathbb{R}^d_+$ with Lipschitz boundary such that f is non-negative and uniformly continuous on Ω and $f = 0$ on $\mathbb{R}^d \setminus \Omega$.

It is worthwhile to take a moment to motivate the hypothesis (H0). Consider the following multi-objective optimization problem

$$\min\{F(x) \,:\, x \in \mathcal{K}\}, \tag{3.2}$$

where $F(x) = (f_1(x), \ldots, f_d(x))$ with $f_i : \mathcal{K} \to [0, \infty)$ for all i, and $\mathcal{K}$ is the set of feasible solutions. This formulation includes many types of constrained optimization problems, where the constraints are implicitly encoded into $\mathcal{K}$. If $x_1, \ldots, x_n$ are feasible solutions in $\mathcal{K}$, then these solutions are ranked, with respect to the optimization problem (3.2), by performing non-dominated sorting on $X_1 = F(x_1), \ldots, X_n = F(x_n)$. Thus the domain Ω of $X_1, \ldots, X_n$ is given by $\Omega = F(\mathcal{K})$. Supposing that $x_1, \ldots, x_n$ are, say, uniformly distributed on $\mathcal{K}$, then the induced density f of $X_1, \ldots, X_n$ on $\mathbb{R}^d$ will be nonzero on Ω and identically zero on $\mathbb{R}^d \setminus \Omega$. Thus, the constraint that feasible solutions must lie in $\mathcal{K}$ directly induces a discontinuity in f along $\partial\Omega$.

Set

$$\mathcal{A} = \left\{\gamma \in C^1([0,1];\mathbb{R}^d) \,:\, \gamma'(t) \geq 0 \ \text{ for all } \ t \in [0,1]\right\}. \tag{3.3}$$

Recall that $\gamma\,'(t) \geq 0$ means that $\gamma_i'(t) \geq 0$ for $i = 1, \ldots, d$ and $\gamma\,'(t) \neq 0$. Define $J : \mathcal{A} \to [0, \infty)$ by

$$J(\gamma) = \int_0^1 f(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}}\, dt, \tag{3.4}$$

and $U : \mathbb{R}^d \to \mathbb{R}$ by

$$U(x) = \sup_{\gamma \in \mathcal{A}\,:\, \gamma(1) \leqq x} J(\gamma). \tag{3.5}$$

We make the following definition.

Definition III.1. Given a domain $\mathcal{O} \subset \mathbb{R}^d$, we say that a function $u : \mathcal{O} \to \mathbb{R}$ is *Pareto-monotone* if $x \leqq y \implies u(x) \leq u(y)$ for all $x, y \in \mathcal{O}$.

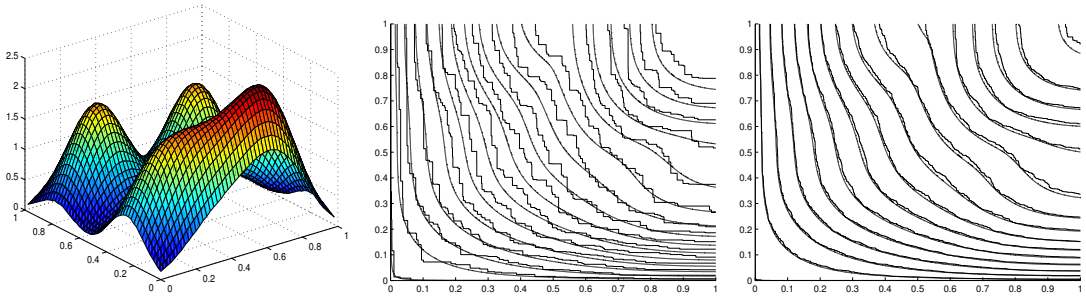
In Section 3.3, we show that U is a Pareto-monotone viscosity solution of the Hamilton-Jacobi partial differential equation (PDE)

$$\text{(P)} \quad \begin{cases} U_{x_1} \cdots U_{x_d} = \dfrac{1}{d^d} f & \text{on } \mathbb{R}^d_+ \\ \qquad\qquad U = 0 & \text{on } \partial\mathbb{R}^d_+. \end{cases}$$

The PDE (P) should be interpreted as the Hamilton-Jacobi-Bellman equation for the value function U. We note that f need only be Borel-measurable, bounded and have compact support in $\mathbb{R}^d_+$ for U to be a viscosity solution of (P). The stronger assumption (H0) is needed to prove that U is the unique Pareto-monotone viscosity solution of (P) (see Theorem III.19). Our main result is

Theorem III.2. *Assume f satisfies (H0) and let $X_1, \ldots, X_n$ be* i.i.d. *with density f. Then there exists a positive constant c_d such that*

$$n^{-\frac{1}{d}} u_n \longrightarrow c_d U \quad \text{in} \ \ L^\infty(\mathbb{R}^d) \ \ \text{almost surely.}$$



(a) Density f (b) $n = 10^4$ points (c) $n = 10^6$ points

Figure 3.1: Comparison of the Pareto fronts for $n = 10^4$ and $n = 10^6$ samples $X_1, \ldots, X_n$, and the level sets of U for the multi-modal density f depicted in a).

Figure 3.1 gives an illustration of the convergence given in Theorem III.2 for a multi-modal density f depicted in the figure. The constants c_d are the same as those given by Bollobás and Winkler [17]. In particular, $c_1 = 1$, $c_2 = 2$ and $c_d \nearrow e$ as $d \to \infty$. When f is a product density, i.e., $f(x) = f_1(x_1) \cdots f_d(x_d)$, the value function U is given by

$$U(x) = \left(\int_{0 \leqq y \leqq x} f(y)\, dy \right)^{\frac{1}{d}} = \left(\int_0^{x_i} f_1(t)\, dt \right)^{\frac{1}{d}} \cdots \left(\int_0^{x_d} f_d(t)\, dt \right)^{\frac{1}{d}}. \tag{3.6}$$

For the case $f = 1$ and $d = 2$, Aldous and Diaconis [2, p. 204] provided a non-rigorous derivation of (P) by viewing the problem as an interacting particle process. They used this to motivate their proof that $c = 2$ in Ulam's problem, but make no rigorous statements about the relationship between (P) and the longest chain problem. A similar, though tangentially related, PDE also appears in growth models in multiple dimensions that are defined through the height of a random partial order [85, p. 209].

Theorem III.2 provides a new tool with which to study the asymptotics of non-dominated sorting and the longest chain problem. As an example of the applicability of this result, we show in Theorem III.26 that non-dominated sorting is asymptotically stable under bounded random perturbations. Evidently, Theorem III.2 reduces

the problem of non-dominated sorting to solving a Hamilton-Jacobi equation. In Chapter IV we explore fast non-dominated sorting algorithms based on the continuum limit given by Theorem III.2.

3.2 Motivation

As motivation, let us give an informal derivation of the Hamilton-Jacobi PDE (P). Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is continuous and $n^{-\frac{1}{d}} u_n \to u \in C^1(\mathbb{R}^d)$ uniformly. Fix n large enough so that $n^{-\frac{1}{d}} u_n \approx u$. Then the k^{th} Pareto front should be well approximated by the level set $\{y \,:\, u(y) = n^{-\frac{1}{d}} k\}$. It is not hard to see that u should be Pareto-monotone (recall Definition III.1), and hence it is reasonable to assume that $u_{x_i} > 0$ for all i. Fix $x, v \in \mathbb{R}^d$ with $\langle Du(x), v\rangle > 0$, where $Du(x)$ denotes the gradient of u at x, and consider the quantity $n^{\frac{1}{d}}(u(x+v) - u(x))$. This is approximately the number of Pareto fronts passing between x and $x+v$. When counting these fronts, we may restrict ourselves to the region

$$A = \{y \,:\, u(y) \geq u(x) \;\text{ and }\; y \leqq x + v\}.$$

This is because any samples in $\{y \,:\, u(y) < u(x)\}$ will be on a previous Pareto front and only samples that are less than $x + v$ can influence the Pareto rank of $x + v$. See Figure 3.2 for a depiction of this region and some quantities from the derivation. Since $u_{x_i}(x) > 0$ for all i, and u is C^1, A is well approximated by a simplex for small $|v|$, and furthermore, the samples within A are approximately uniformly distributed. Let m denote the number of samples falling in A. By scaling the simplex into a standard simplex, without disrupting the Pareto ordering within A, it is reasonable to conjecture that the number of Pareto fronts within A (or the length of a longest chain in A) is approximately $cm^{\frac{1}{d}}$ for some constant c, independent of x. For simplicity we take $c = 1$.

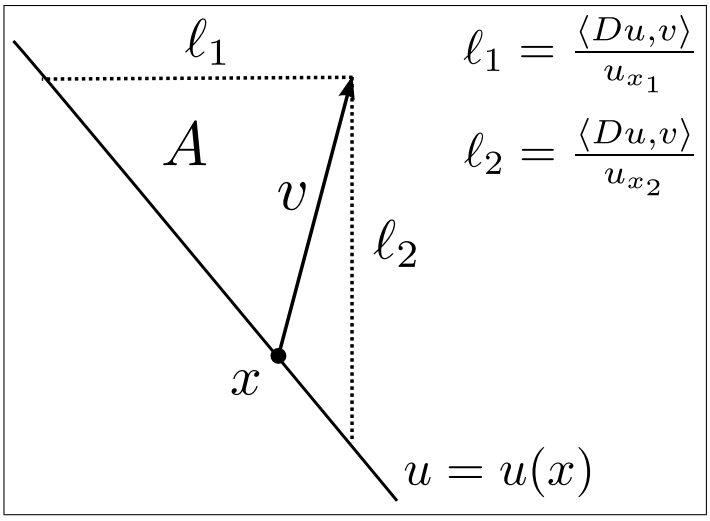


Figure 3.2: Some quantities from the informal derivation of the Hamilton-Jacobi PDE (P).

By the law of large numbers, we have $m \approx n \int_A f(y)\, dy$. Hence when $|v| > 0$ is small we have

$$n^{\frac{1}{d}}(u(x+v) - u(x)) \approx \left(n \int_A f(y)\, dy \right)^{\frac{1}{d}} \approx n^{\frac{1}{d}} |A|^{\frac{1}{d}} f(x)^{\frac{1}{d}}, \tag{3.7}$$

where $|A|$ denotes the Lebesgue measure of A. Let $\ell_1, \ldots, \ell_d$ denote the side lengths of the simplex A. Then $|A| \approx c\, \ell_1 \cdots \ell_d$ for a constant c which we again take to be 1. Since $x + v - \ell_i e_i$ lies approximately on the tangent plane to the level set $\{y : u(y) = u(x)\}$, we see that

$$\langle Du(x), v - \ell_i e_i \rangle \approx 0.$$

Rearranging the above we see that $\ell_i \approx u_{x_i}(x)^{-1} \langle Du(x), v \rangle$, and hence

$$|A| \approx u_{x_1}(x)^{-1} \cdots u_{x_d}(x)^{-1} \langle Du(x), v \rangle^d. \tag{3.8}$$

For small $|v|$, we can combine (3.8) and (3.7) to obtain

$$\langle Du(x), v \rangle \approx u(x+v) - u(x) \approx f(x)^{\frac{1}{d}} u_{x_1}(x)^{-\frac{1}{d}} \cdots u_{x_d}(x)^{-\frac{1}{d}} \langle Du(x), v \rangle.$$

Simplifying, we see that u should satisfy

$$u_{x_1} \cdots u_{x_d} = f \quad \text{on} \quad \mathbb{R}^d, \tag{3.9}$$

up to scaling by a constant.

Although this derivation is informal, it is straightforward and conveys the essence of the result. It is difficult, however, to construct a rigorous proof based on these heuristics. There are two main reasons for this. First, it supposes that $n^{-\frac{1}{d}} u_n$ converges to a limit u, which is not obvious. Second, it is essential that $u \in C^1$, as we require A to be an approximate simplex. Solutions of (3.9) are in general not smooth, and can have points of non-differentiability due to crossing characteristics. This is true even in the case that f is smooth, and is related to the geometry of Ω.

3.3 Analysis of variational problem

Before studying the variational problem (3.5), we recall some aspects of the theory of optimal control [10] that are relevant to our problem. We will describe the infinite horizon optimal control problem, but the discussion below applies with minor modifications to other variants of optimal control, such as finite horizon or undiscounted problems with exit times. The state of the control problem, $y(t)$, is assumed to obey the dynamics

$$\begin{cases} y'(t) = g(y(t), \alpha(t)), \quad t > 0 \\ y(0) = x, \end{cases} \tag{3.10}$$

where $\alpha : [0, \infty) \to A$ is the control, A is a topological space, and $g : \mathbb{R}^d \times A \to \mathbb{R}^d$. Given an initial condition $x \in \mathbb{R}^d$, the solution of (3.10) is denoted $y_x(\cdot)$. Let

$$\mathcal{A} := \{\text{measurable functions } [0, \infty) \to A\}.$$

The goal in optimal control is to select the control $\alpha \in \mathcal{A}$ to minimize the cost functional

$$J(x,\alpha) := \int_0^\infty c(y_x(t),\alpha(t))e^{-\lambda t}\,dt, \tag{3.11}$$

where $\lambda > 0$ and $c : \mathbb{R}^d \times A \to \mathbb{R}$. The value function for this problem is

$$v(x) := \inf_{\alpha\in\mathcal{A}} J(x,\alpha). \tag{3.12}$$

Under sufficient regularity assumptions on c and g (discussed below), the value function is a Hölder- (or Lipschitz) continuous viscosity solution of the Hamilton-Jacobi-Bellman equation

$$\lambda v + H(x,Dv) = 0 \quad \text{on} \quad \mathbb{R}^d, \tag{3.13}$$

where

$$H(x,p) = \sup_{a\in A}\{-\langle g(x,a),p\rangle - c(x,a)\}. \tag{3.14}$$

Although the variational problem (3.5) can be cast in this framework, the assumptions on the running cost $c(\cdot,\cdot)$ in the existing literature are too restrictive. For our variational problem, we have $\lambda = 0$, $g(x,a) = a$, $A = \mathbb{R}^d_+$,

$$c(x,a) = -f(x)^{\frac{1}{d}}(a_1\cdots a_d)^{\frac{1}{d}}, \tag{3.15}$$

and $U(x) = -v(x)$. In the proofs of Theorems III.24 and III.25, we require the standard optimal control theory to hold for f piecewise constant on arbitrarily small grids. In the standard reference on optimal control [10], it is assumed that $x \mapsto c(x,a)$ is uniformly continuous. This assumption is then used to prove regularity of the value function v. There is relatively little research devoted to relaxing the regularity condition on c. There are some results for the optimal control problem associated

with the Eikonal equation [74, 21, 30], which allow c to have discontinuities. These results assume that $A = \mathbb{R}^d$ and make essential use of either Lipschitzness of v, or uniform continuity and/or coercivity of $p \mapsto H(x,p)$, none of which hold for the variational problem (3.5). Soravia [86] and Garavello and Soravia [41] considered a running cost of the form $c(x,a) = c_1(x,a) + c_2(x)$, where c_1 is continuous and c_2 is Borel-measurable, and showed that the standard optimal control results hold with minor modifications. This is incompatible with (3.15) when f is not continuous. A similar program is carried out for differential games here [42]. Barles et al. [11] study optimal control on multi-domains, where the discontinuity in c is assumed to lie in a half-space.

Under the assumption that f is compactly supported, bounded and Borel-measurable, the standard results on optimal control hold for the variational problem (3.5) with minor modifications to the proofs. In particular, in Lemma III.3 we show that U is Hölder-continuous with exponent $\frac{1}{d}$, and in Theorem III.10 we show that U is a viscosity solution of (P). The uniqueness of viscosity solutions of (P) under the assumption that f satisfies (H0) is a more delicate problem. This is addressed in Section 3.4.

In our main result, Theorem III.2, we assume that f satisfies (H0), which is stronger than Borel-measurability. We assume Borel-measurability in much of this section so that our results apply to piecewise constant densities, which are used to approximate f in the proofs of Theorems III.24 and III.25. To be more precise, we set

(3.16)

$$\mathcal{B} = \{f : \mathbb{R}^d \to \mathbb{R} \,:\, f \text{ is bounded, Borel-measurable, and } \operatorname{supp}(f) \subset [0,1]^d\}.$$

We note that the assumption $\operatorname{supp}(f) \subset [0,1]^d$ is not restrictive, as we can make

a simple scaling argument to obtain the case where f has compact, but arbitrary, support in $\mathbb{R}^d$. The compact support assumption is not essential; it is mainly invoked to simplify the exposition. Following arguments similar to those in Chapter V, it is relatively straightforward to accommodate densities f with unbounded supports as well.

We now introduce some new notation. We will write $\gamma \leqq x$ whenever $\gamma(t) \leqq x$ for all $t \in [0,1]$. We write $\gamma_1 \leqq \gamma_2$ whenever $\gamma_1(1) \leqq \gamma_2(0)$. The same definitions apply to $\leq, <, \geqq, \geqq$ and $>$ with obvious modifications. For $y \in \mathbb{R}^d$ and $r > 0$ we set $B_r(y) = \{x \in \mathbb{R}^d \,:\, |x - y| < r\}$. For $x, y \in \mathbb{R}^d$ we set

$$w(x,y) = \begin{cases} \sup\{J(\gamma) \,:\, \gamma \in \mathcal{A} \text{ and } x \leqq \gamma \leqq y\} & \text{if } x \leqq y \\ 0 & \text{otherwise.} \end{cases} \tag{3.17}$$

3.3.1 Basic properties of U

We establish here some basic properties of U. Namely, in Lemma III.3 we establish Hölder-continuity of U, and in Lemma III.7, we establish a dynamic programming principle for U.

Lemma III.3. *Let $f \in \mathcal{B}$. Then U is Hölder-continuous with exponent $\frac{1}{d}$ and Hölder seminorm $[U]_{\frac{1}{d}} \leq \|f\|^{\frac{1}{d}}_{L^\infty(\mathbb{R}^d)}$.*

Proof. Let $x, z \in \mathbb{R}^d$ and let $\varepsilon > 0$. Choose $\gamma \in \mathcal{A}$ with $\gamma \leqq x$ and $J(\gamma) \geq U(x) - \varepsilon$. Since $f(x) = 0$ for $x \notin [0,1]^d$, we may assume that $\gamma(t) \in [0,1]^d$ for all $t \in [0,1]$. Set

$$s = \sup\{t \in [0,1] \,:\, \gamma(t) \leqq z\}.$$

If for all $t \in [0,1]$ we have $\gamma(t) \not\leqq z$, then set $s = 0$. We claim that

$$U(z) \geq U(x) - \int_s^1 f(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}} \, dt - \varepsilon. \tag{3.18}$$

To see this: In the case that $s > 0$, we have $\gamma(s) \leqq z$ and hence

$$\begin{aligned} U(z) &\geq \int_0^s f(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}}\, dt \\ &= J(\gamma) - \int_s^1 f(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}}\, dt \\ &\geq U(x) - \int_s^1 f(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}}\, dt - \varepsilon. \end{aligned}$$

In the case that $s = 0$, we have

$$\begin{aligned} U(z) \geq 0 = J(\gamma) &- \int_0^1 f(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}}\, dt \\ &\geq U(x) - \int_0^1 f(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}}\, dt - \varepsilon. \end{aligned}$$

Hence (3.18) is established. Suppose $s < 1$. Then there must exist i such that $\gamma_i(s) \geq z_i$. It follows that

$$\int_s^1 \gamma_i'(t)\, dt = \gamma_i(1) - \gamma_i(s) \leq x_i - z_i = |x_i - z_i|.$$

Applying the generalized Hölder inequality we see that

$$\begin{aligned} \int_s^1 f(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}}\, dt &\leq \|f\|_{L^\infty(\mathbb{R}^d)}^{\frac{1}{d}} \left(\int_s^1 \gamma_1'(t)\, dt \right)^{\frac{1}{d}} \cdots \left(\int_s^1 \gamma_d'(t)\, dt \right)^{\frac{1}{d}} \\ &\leq \|f\|_{L^\infty(\mathbb{R}^d)}^{\frac{1}{d}} |x_i - z_i|^{\frac{1}{d}} \prod_{j \neq i} (\gamma_j(1) - \gamma_j(s))^{\frac{1}{d}} \\ &\leq \|f\|_{L^\infty(\mathbb{R}^d)}^{\frac{1}{d}} |x_i - z_i|^{\frac{1}{d}}. \end{aligned}$$

Inserting this into 3.18 we obtain

$$U(x) - U(z) \leq \|f\|_{L^\infty(\mathbb{R}^d)}^{\frac{1}{d}} |x_i - z_i|^{\frac{1}{d}} + \varepsilon \leq \|f\|_{L^\infty(\mathbb{R}^d)}^{\frac{1}{d}} |x - z|^{\frac{1}{d}} + \varepsilon \tag{3.19}$$

If $s = 1$ then inspecting (3.18), we see that $U(x) - U(z) \leq \varepsilon$, which implies (3.19). Sending $\varepsilon \to 0$ we find that $U(x) - U(z) \leq \|f\|_{L^\infty(\mathbb{R}^d)}^{\frac{1}{d}} |x - z|^{\frac{1}{d}}$. We can reverse the roles of x and z in the preceding argument to obtain the opposite inequality. $\square$

Remark III.4. By a similar argument, we can show that $w : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is Hölder-continuous with exponent $1/d$ and $[w]_{\frac{1}{d}} \leq \|f\|^{\frac{1}{d}}_{L^\infty(\mathbb{R}^d)}$.

In general, the statement in Lemma III.3 cannot be strengthened. For example, if $f \equiv 1$ then we have $U(x) = (x_1 \dots x_d)^{\frac{1}{d}} \notin C^{0,\alpha}([0,\infty)^d)$ for any $\alpha > 1/d$. However, it is easy to see that $U(x) = (x_1 \cdots x_d)^{\frac{1}{d}} \in C^{0,1}([h,R]^d)$ for any $0 < h < R < \infty$. Our next result shows that this is true a more general class of densities f.

Lemma III.5. *Let $0 < h < R$, and suppose that f is Lipschitz and $f \geq m > 0$ on $[0,R]^d$. Then we have*

$$\|U\|_{C^{0,1}([h,R]^d)} \leq C\left(h, R, m\right)\left(1 + \|f\|_{C^{0,1}([0,R]^d)}\right). \tag{3.20}$$

Proof. Let $x, y \in [h,R]^d$ and let us first suppose that $x \leqq y$. Let $\varepsilon > 0$ and $\gamma \in \mathcal{A}$ such that $\gamma(1) = y$ and $J(\gamma) \geq U(y) - \varepsilon$. Define the mapping $\Psi : [0,y] \to [0,x]$ by

$$\Psi(z) = \left(\frac{x_1}{y_1} z_1, \dots, \frac{x_d}{y_d} z_d\right), \tag{3.21}$$

and set $\overline{\gamma}(t) = \Psi(\gamma(t))$ for $t \in [0,1]$. Notice that we have $\gamma(1) = y$ and hence $\overline{\gamma}(1) = x$. Therefore we have

$$|U(x) - U(y)| = U(y) - U(x) \leq J(\gamma) - J(\overline{\gamma}) + \varepsilon. \tag{3.22}$$

Since $x, y \in [h,R]^d$ we have

$$|z - \Psi(z)| = \left|\left(z_1 - \frac{x_1}{y_1} z_1, \dots, z_d - \frac{x_d}{y_d} z_d\right)\right| \leq Rh^{-1}|x - y| \tag{3.23}$$

for any $z \in [0,R]^d$. It follows that

$$|\gamma(t) - \overline{\gamma}(t)| \leq Rh^{-1}|x - y| \quad \text{for all } t \in [0,1]. \tag{3.24}$$

Since $f \geq m > 0$ and f is Lipschitz, it follows that

$$|f(\gamma(t))^{\frac{1}{d}} - f(\overline{\gamma}(t))^{\frac{1}{d}}| \leq d^{-1} m^{\frac{1}{d}-1} |f(\gamma(t)) - f(\overline{\gamma}(t))| \overset{(3.24)}{\leq} C[f]_1 |x - y|, \tag{3.25}$$

for all $t \in [0,1]$, where $C = C(h,R,m)$.

By the definition of Ψ and $\overline{\gamma}$ we have

$$\overline{\gamma}_1'(t)\cdots\overline{\gamma}_d'(t) = \frac{x_1\cdots x_d}{y_1\cdots y_d}\gamma_1'(t)\cdots\gamma_d'(t) \quad \text{for all } t\in[0,1]. \tag{3.26}$$

Now set $\Phi(z) = (z_1\cdots z_d)^{\frac{1}{d}}$. Then we have $|D\Phi(z)| \le CR/h$ for $z \in [h,R]^d$, and it follows that

$$\frac{(x_1\cdots x_d)^{\frac{1}{d}}}{(y_1\cdots y_d)^{\frac{1}{d}}} = \frac{\Phi(y)+\Phi(x)-\Phi(y)}{\Phi(y)} \ge 1 - C|x-y|, \tag{3.27}$$

where $C = C(h,R)$. Combining (3.25), (3.26) and (3.27) we have

$$\begin{aligned}
J(\overline{\gamma}) &\overset{(3.26)}{=} \int_0^1 f(\overline{\gamma}(t))^{\frac{1}{d}} \frac{(x_1\cdots x_d)^{\frac{1}{d}}}{(y_1\cdots y_d)^{\frac{1}{d}}} \left(\gamma_1'(t)\cdots\gamma_d'(t)\right)^{\frac{1}{d}}\, dt \\
&\overset{(3.25),(3.27)}{\ge} \int_0^1 \left(f(\gamma(t))^{\frac{1}{d}} - C[f]_1|x-y|\right)(1-C|x-y|)\left(\gamma_1'(t)\cdots\gamma_d'(t)\right)^{\frac{1}{d}}\, dt \\
&\ge J(\gamma) - C\left(1+\|f\|_{C^{0,1}([0,R]^d)}\right)|x-y| \int_0^1 \left(\gamma_1'(t)\cdots\gamma_d'(t)\right)^{\frac{1}{d}}\, dt \\
&\ge J(\gamma) - CR\left(1+\|f\|_{C^{0,1}([0,R]^d)}\right)|x-y|, \qquad (3.28)
\end{aligned}$$

where $C = C(h,R,m)$ and we applied the generalized Hölder inequality in the last line. Combining (3.28) with (3.22) and sending $\varepsilon \to 0$ we have

$$|U(x)-U(y)| \le C\left(1+\|f\|_{C^{0,1}([0,R]^d)}\right)|x-y| \tag{3.29}$$

for all $x, y \in [h,R]^d$ with $x \leqq y$. If we do not have $x \leqq y$, then we set

$$\widehat{x} = (\min(x_1,y_1),\ldots,\min(x_d,y_d))$$

and note that $\widehat{x} \leqq x$, $\widehat{x} \leqq y$ and $\widehat{x} \in [h,R]^d$. Then since we have

$$|U(x)-U(y)| \le |U(\widehat{x})-U(x)| + |U(\widehat{x})-U(y)|,$$

we see that (3.29) holds for all $x, y \in [h,R]^d$.

To complete the proof, we note that a straightforward application of Hölder's inequality yields

$$\|U\|_{L^\infty([0,R]^d)} \leq R\|f\|^{\frac{1}{d}}_{L^\infty([0,R]^d)} \leq R\left(1 + \|f\|_{L^\infty([0,R]^d)}\right). \qquad \square$$

Remark III.6. A natural question is whether $U \in W^{1,p}([0,1]^d)$ for some $p \geq 1$. In the special case of $f \equiv 1$ we have $U = (x_1 \cdots x_d)^{\frac{1}{d}}$, and it is easily verified that $U \in W^{1,p}([0,1]^d)$ for any $p < d/(d-1)$. It is reasonable to suspect that such an estimate holds for more general densities f. By Lemma III.5, we see that for f Lipschitz and strictly positive, U is Lipschitz away from the boundary $\partial\mathbb{R}^d_+$, hence the issue boils down to obtaining estimates on the size of DU near $\partial\mathbb{R}^d_+$. Inspecting Lemma III.5, one can see that if $y - x = te_i$ for some $t > 0, i$, then the constant in the statement of the Lemma is given by $C(h, R, m) = C(R, m)h^{-1}$. This implies that $x_i U_{x_i} \in L^\infty([0,1]^d)$, but unfortunately does not give us any information about membership in $W^{1,p}([0,1]^d)$. We suspect that the constant in Lemma III.5 can be improved to $C(h, R, m) = C(R, m)h^{\frac{1}{d}-1}$, but at present we are not able to prove this.

We now have the following dynamic programming principle for U.

Lemma III.7 (Dynamic Programming Principle)**.** *Let $f \in \mathcal{B}$. Then for any $r > 0$ and $y \in \mathbb{R}^d$ we have*

$$U(y) = \max_{x \in \partial B_r(y) \,:\, x \leq y} \Big\{ U(x) + w(x, y) \Big\}. \tag{3.30}$$

Proof. Let us denote the right hand side of (3.30) by $v(y)$. We first show that $U(y) \leq v(y)$. Let $\varepsilon > 0$ and let $\gamma \in \mathcal{A}$ such that $\gamma \leqq y$ and $J(\gamma) \geq U(y) - \varepsilon$. Suppose that $|\gamma(1) - y| \geq r$. Then there exists $x \in \partial B_r(y)$ such that $\gamma(1) \leqq x \leqq y$ and hence

$$U(y) \leq J(\gamma) + \varepsilon \leq U(x) + \varepsilon \leq v(y) + \varepsilon.$$

If $|\gamma(0) - y| \leq r$ then there exists $x \in \partial B_r(y)$ such that $x \leq \gamma \leq y$ and hence

$$v(y) \geq w(x, y) \geq J(\gamma) \geq U(y) - \varepsilon.$$

Finally, suppose that $|\gamma(1) - y| < r$ and $|\gamma(0) - y| > r$. Then there exists $0 < s < 1$ such that $|\gamma(s) - y| = r$. Set $x = \gamma(s)$ and define $\gamma^1, \gamma^2 \in \mathcal{A}$ by

$$\gamma^1(t) = \gamma(st) \quad \text{and} \quad \gamma^2(t) = \gamma(s + t(1 - s)) \quad \text{for} \quad t \in [0, 1].$$

Note that $\gamma^1 \leqq x$ and $x \leqq \gamma^2 \leqq y$. Since J is invariant under a change of parametrization of γ, we see that

$$U(y) \leq J(\gamma) + \varepsilon = J(\gamma^1) + J(\gamma^2) + \varepsilon \leq U(x) + w(x, y) + \varepsilon \leq v(y) + \varepsilon.$$

Sending $\varepsilon \to 0$ we obtain $U(y) \leq v(y)$.

We now show that $U(y) \geq v(y)$. By Lemma III.3 and Remark III.4, there exists $x \in \partial B_r(y)$ with $x \leq y$ such that

$$v(y) = U(x) + w(x, y).$$

Let $\varepsilon > 0$ and let $\gamma^1, \gamma^2 \in \mathcal{A}$ with $\gamma^1 \leqq x$ and $x \leqq \gamma^2 \leqq y$ such that

$$J(\gamma^1) \geq U(x) - \frac{\varepsilon}{2} \quad \text{and} \quad J(\gamma^2) \geq w(x, y) - \frac{\varepsilon}{2}.$$

Since $\gamma^1 \leqq \gamma^2 \leqq y$, we can concatenate γ^1 and γ^2 to find that $U(y) \geq J(\gamma^1) + J(\gamma^2)$. Thus we have

$$U(y) \geq U(x) + w(x, y) - \varepsilon = v(y) - \varepsilon.$$

Sending $\varepsilon \to 0$ yields $U(y) \geq v(y)$. □

3.3.2 Hamilton-Jacobi-Bellman equation for U

We digress momentarily to recall the definition of viscosity solution of

$$H(x, Du) = 0 \quad \text{on} \quad \mathcal{O}, \tag{3.31}$$

where $\mathcal{O} \subset \mathbb{R}^d$ is open, $H : \mathcal{O} \times \mathbb{R}^d \to \mathbb{R}$ is locally bounded with $p \mapsto H(x,p)$ continuous for every $x \in \mathcal{O}$, and $u : \mathcal{O} \to \mathbb{R}$ is the unknown function. For more information on viscosity solutions of Hamilton-Jacobi equations, we refer the reader to [10, 27].

We denote by $\mathrm{USC}(\mathcal{O})$ (resp. $\mathrm{LSC}(\mathcal{O})$) the set of upper semicontinuous (resp. lower semicontinuous) functions on $\mathcal{O}$. For $u : \mathcal{O} \to \mathbb{R}$, the *superdifferential* of u at $x \in \mathcal{O}$, denoted $D^+u(x)$, is the set of all $p \in \mathbb{R}^d$ satisfying

$$u(y) \leq u(x) + \langle p, y - x\rangle + o(|x-y|) \quad \text{as} \quad \mathcal{O} \ni y \to x. \tag{3.32}$$

Similarly, the *subdifferential* of u at $x \in \mathcal{O}$, denoted $D^-u(x)$, is the set of all $p \in \mathbb{R}^d$ satisfying

$$u(y) \geq u(x) + \langle p, y - x\rangle + o(|x-y|) \quad \text{as} \quad \mathcal{O} \ni y \to x. \tag{3.33}$$

Equivalently, we may set

$$D^+u(x) = \{D\varphi(x) \,:\, \varphi \in C^1(\mathcal{O}) \ \text{ and } \ u - \varphi \text{ has a local max at } x\},$$

and

$$D^-u(x) = \{D\varphi(x) \,:\, \varphi \in C^1(\mathcal{O}) \ \text{ and } \ u - \varphi \text{ has a local min at } x\}.$$

Definition III.8. A *viscosity subsolution* of (3.31) is a function $u \in \mathrm{USC}(\mathcal{O})$ satisfying

$$\liminf_{y \to x} H(y,p) =: H_*(x,p) \leq 0 \quad \text{for all} \quad x \in \mathcal{O} \quad \text{and} \quad p \in D^+u(x). \tag{3.34}$$

Similarly, a *viscosity supersolution* of (3.31) is a function $u \in \mathrm{LSC}(\mathcal{O})$ satisfying

$$\limsup_{y \to x} H(y,p) =: H^*(x,p) \geq 0 \quad \text{for all} \quad x \in \mathcal{O} \quad \text{and} \quad p \in D^-u(x). \tag{3.35}$$

The functions H_* and H^* are the lower and upper semicontinuous envelopes of H with respect to the spatial variable, respectively. We will often say u is a viscosity solution of

$$H(x, Du) \leq 0 \quad (\text{resp. } H(x, Du) \geq 0) \text{ on } \quad \mathcal{O},$$

to indicate that u is a viscosity subsolution (resp. supersolution) of (3.31). If u is a viscosity subsolution and supersolution of (3.31), then we say that u is a *viscosity solution* of (3.31). Notice that viscosity solutions defined in this way are necessarily continuous.

After a basic proposition, we establish in Theorem III.10 that U is a *Pareto-monotone* viscosity solution of (P).

Proposition III.9. *Let $\mathcal{O} \subset \mathbb{R}^d$ be open and let $v : \mathcal{O} \to \mathbb{R}$ be Pareto-monotone. Then*

$$D^+v(x) \cup D^-v(x) \subset \overline{\mathbb{R}^d_+} \text{ for all } x \in \mathcal{O}.$$

Proof. Let $x \in \mathcal{O}$ and $p \in D^+v(x)$. For any index i and small enough $t > 0$, we have $x \leqq x + te_i \in \mathcal{O}$. Since v is Pareto-monotone, we have

$$v(x) \leq v(x + te_i) \leq v(x) + p_i t + o(t) \text{ as } t \searrow 0.$$

Hence $p_i \geq o(t)/t$ as $t \searrow 0$ which implies that $p_i \geq 0$. The proof for $D^-v(x)$ is similar. □

Theorem III.10. *Let $f \in \mathcal{B}$. Then the value function U defined by* (3.5) *is a Pareto-monotone viscosity solution of the Hamilton-Jacobi equation*

$$U_{x_1} \cdots U_{x_d} = \frac{1}{d^d} f \quad \text{on} \quad \mathbb{R}^d. \tag{3.36}$$

Furthermore, U satisfies

(i) Whenever $\operatorname{supp}(f) \subset \{x \in \mathbb{R}^d : 0 \leqq x \leqq z\}$, *we have*

$$U(x_1, \dots, x_d) = U(\min(x_1, z_1), \cdots \min(x_d, z_d)) \text{ for all } x \in \mathbb{R}^d_+,$$

(ii) $U(x) = 0$ *for every* $x \in \mathbb{R}^d \setminus \mathbb{R}^d_+$.

Proof. It follows from the definition of U (3.5) that U is Pareto-monotone, and (ii) follows from the fact that $\operatorname{supp}(f) \subset [0,1]^d$.

For (i), let $z \in \mathbb{R}^d$ such that $\operatorname{supp}(f) \subset \{x \in \mathbb{R}^d : 0 \leqq x \leqq z\}$ and let $x \in \mathbb{R}^d_+$ such that $x_i > z_i$ for some i. Set $\widehat{x} = (\min(x_1, z_1), \dots, \min(x_d, z_d))$. Since U is Pareto-monotone we have $U(\widehat{x}) \leq U(x)$. Let $\varepsilon > 0$ and $\gamma \in \mathcal{A}$ such that $\gamma \leqq x$ and $U(x) \leq J(\gamma) + \varepsilon$. Let

$$s = \sup\{t : \gamma(t) \leqq \widehat{x}\}.$$

If for all $t \in [0,1]$ we have $\gamma(t) \not\leqq \widehat{x}$, then set $s = 0$. If $s = 1$, then $\gamma \leqq \widehat{x}$ and hence $U(x) \leq J(\gamma) + \varepsilon \leq U(\widehat{x}) + \varepsilon$. If $s = 0$ then for every $t \in [0,1]$, $\gamma(t) \notin \operatorname{supp}(f)$, and hence $J(\gamma) = 0$. It follows that

$$U(x) \leq J(\gamma) + \varepsilon = \varepsilon \leq U(\widehat{x}) + \varepsilon.$$

If $0 < s < 1$, then for any $t > s$, $\gamma_i(t) > z_i$ for some i, and hence $f(\gamma(t)) = 0$. Set $\gamma^1(t) = \gamma(st)$ for $t \in [0,1]$. Then $\gamma^1 \leqq \widehat{x}$ and $J(\gamma) = J(\gamma^1)$, hence $U(x) \leq J(\gamma) + \varepsilon = J(\gamma^1) + \varepsilon \leq U(\widehat{x}) + \varepsilon$. Sending $\varepsilon \to 0$ we see that $U(x) \leq U(\widehat{x})$ and hence $U(x) = U(\widehat{x})$.

We now show that U is a viscosity supersolution of (3.36). Let $y \in \mathbb{R}^d$, let $a \in \mathbb{R}^d_+$, and set $\gamma(t) = y - a(1-t)$. By Lemma III.7 we have

$$\begin{aligned} U(y) &\geq U(y - a(1-t)) + \int_t^1 f(y - a(1-s))^{\frac{1}{d}} (a_1 \cdots a_d)^{\frac{1}{d}} \, ds \\ &\geq U(y - a(1-t)) + (1-t) f_*(y) (a_1 \cdots a_d)^{\frac{1}{d}} + o(1-t) \text{ as } t \nearrow 1. \end{aligned} \tag{3.37}$$

Let $p \in D^- U(y)$. Since $y - a(1-t) \to y$ as $t \nearrow 1$, we have

$$\begin{aligned}\langle p, (1-t)a\rangle &\overset{(3.33)}{\geq} U(y) - U(y - a(1-t)) + o(1-t) \\ &\overset{(3.37)}{\geq} (1-t) f_*(y)(a_1 \cdots a_d)^{\frac{1}{d}} + o(1-t) \quad \text{as} \quad t \nearrow 1.\end{aligned}$$

Sending $t \nearrow 1$ we obtain

$$\langle p, a\rangle \geq f_*(y)^{\frac{1}{d}} (a_1 \cdots a_d)^{\frac{1}{d}}.$$

Since $a > 0$ was arbitrary, we obtain

$$\sup_{a>0} \left\{ -\langle p, a\rangle + f_*(y)^{\frac{1}{d}} (a_1 \cdots a_d)^{\frac{1}{d}} \right\} \leq 0. \tag{3.38}$$

Since U is Pareto-monotone, Proposition III.9 yields $p \geq 0$. Hence if $f_*(y) = 0$ then U is trivially a viscosity supersolution of (3.36) at y. We may therefore suppose that $f_*(y) > 0$. Fix i and set $a_j = 1$ for $j \neq i$. By (3.38) we have

$$\sup_{a_i>0} \left\{ -\sum_{j \neq i} p_j + a_i^{\frac{1}{d}} f_*(y)^{\frac{1}{d}} - a_i p_i \right\} \leq 0.$$

Since a_i can be arbitrarily large, we must have $p_i > 0$ for the above to hold. Substituting $a_i = p_i^{-1}$ into (3.38) and simplifying we obtain

$$p_1 \cdots p_d \geq \frac{1}{d^d} f_*(y).$$

Thus U is a viscosity supersolution of (3.36).

We now show that U is a viscosity subsolution of (3.36). Let $y \in \mathbb{R}^d$, let $\varepsilon > 0$, and let $p \in D^+ U(y)$. By Lemmas III.3 and III.7 and Remark III.4, for every $r > 0$ there exists $x \in \partial B_r(y)$ with $x \leq y$ such that $U(y) = U(x) + w(x, y)$. Hence there exists $\gamma \in \mathcal{A}$ with $x \leqq \gamma \leqq y$ such that

$$U(y) \leq \int_0^1 f(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}} \, dt + U(x) + \varepsilon r.$$

By Hölder's inequality

$$\begin{aligned} U(y) - U(x) &\le \int_0^1 f(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}} \, dt + \varepsilon r \\ &\le (f^*(y)^{\frac{1}{d}} + o(1)) \left(\int_0^1 \gamma_1'(t) \, dt \right)^{\frac{1}{d}} \cdots \left(\int_0^1 \gamma_d'(t) \, dt \right)^{\frac{1}{d}} + \varepsilon r \\ &\le f^*(y)^{\frac{1}{d}} |x_1 - y_1|^{\frac{1}{d}} \cdots |x_d - y_d|^{\frac{1}{d}} + o(r) + \varepsilon r, \end{aligned}$$

as $r \searrow 0$. Since $x \to y$ as $r \searrow 0$, we have

$$\langle p, y - x \rangle \overset{(3.32)}{\le} U(y) - U(x) + o(r) \le f^*(y)^{\frac{1}{d}} |x_1 - y_1|^{\frac{1}{d}} \cdots |x_d - y_d|^{\frac{1}{d}} + o(r) + \varepsilon r,$$

as $r \searrow 0$. Choose $r > 0$ small enough so that $o(r)/r \le \varepsilon$, and set $a = (y - x)/r$. Then we have

$$-\langle p, a \rangle + f^*(y)^{\frac{1}{d}} (a_1 \cdots a_d)^{\frac{1}{d}} \ge -2\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we see that

$$\sup_{a \ge 0 \,:\, |a| = 1} \left\{ -\langle p, a \rangle + f^*(y)^{\frac{1}{d}} (a_1 \cdots a_d)^{\frac{1}{d}} \right\} \ge 0.$$

Since U is Pareto-monotone, we have $p \ge 0$. If $p_i = 0$ for some i, then $p_1 \cdots p_d \le f^*(y)/d^d$. Thus we may assume that $p_i > 0$ for all i. Then the supremum above is attained at some $a > 0$ with $|a| = 1$. By scaling a so that $a_1 \cdots a_d = 1$, we see that

$$\sup_{a > 0 \,:\, a_1 \cdots a_d = 1} \left\{ -\langle p, a \rangle + f^*(y)^{\frac{1}{d}} \right\} \ge 0. \tag{3.39}$$

Since $p_i > 0$ for all i, we have that

$$\limsup_{|a| \to \infty, \ a > 0} -\langle p, a \rangle + f^*(y)^{\frac{1}{d}} = -\infty.$$

It follows that the supremum in (3.39) is attained at some $a^* > 0$. Introducing a Lagrange multiplier $\lambda > 0$, the necessary conditions for a^* to be a maximizer of the above constrained optimization problem are

$$p_i = \frac{\lambda}{a_i^*} \ \text{ for all } i \in \{1, \ldots, d\} \ \text{ and } \ a_1^* \cdots a_d^* = 1.$$

It follows that $\lambda = (p_1 \cdots p_d)^{\frac{1}{d}}$ and $a_i^* = p_i^{-1}(p_1 \cdots p_d)^{\frac{1}{d}}$. Substituting this into (3.39) we have

$$p_1 \cdots p_d \leq \frac{1}{d^d} f^*(y),$$

which completes the proof. □

3.4 Comparison principle

We aim here to establish that U is the unique viscosity solution of (P) under hypothesis (H0) on f, which in general allows f to be discontinuous. The standard results on uniqueness of viscosity solutions [10, 27] assume uniformly continuous dependence on spatial variables. There has been some recent work relaxing this condition, as it is important in many applications. Tourin [90] considered Hamilton-Jacobi equations of the form $H(x, Du) = 0$, where $x \mapsto H(x, p)$ is allowed to have a discontinuity along a smooth surface, and proved a comparison principle under the assumption that $p \mapsto H(x, p)$ is convex and uniformly continuous. Neither assumption holds for (P), although the non-convexity can be easily remedied. Deckelnick and Elliot [30] prove a comparison principle for Lipschitz viscosity solutions of Eikonal-type equations of the form $H(Du) = f$, where f satisfies a regularity condition similar to (H0), but slightly more general. As exhibited by the solution $U(x) = (x_1 \cdots x_d)^{\frac{1}{d}}$ of (P) for $f = 1$, solutions of (P) are not in general Lipschitz continuous. Camilli and Siconolfi [22] proposed a new notion of viscosity solution for Hamilton-Jacobi equations in which H has measurable dependence on the spatial variable x. They obtain general uniqueness results under the assumption that $p \mapsto H(x, p)$ is quasiconvex and coercive. Their results do not apply to (P) due to the coercivity assumption.

The main result in this section, Theorem III.19, establishes uniqueness for (P) under hypothesis (H0). Let us give a sketch of the proof now. Let u be a Pareto-

monotone viscosity solution of (P). We first prove a standard comparison principle, in Theorem III.11, for uniformly continuous f. We can then define the regularized value functions U_ε and U^ε by replacing f by its inf and sup convolutions f_ε and f^ε, respectively, in (3.5). Since f^ε and f_ε are Lipschitz continuous, the comparison principle from Theorem III.11 yields $U_\varepsilon \leq u \leq U^\varepsilon$. The proof is completed by showing that $U_\varepsilon, U^\varepsilon \to U$ as $\varepsilon \to 0$, where U is the value function defined by (3.5). We establish a more general result in Lemma III.18, the proof of which relies on the second comparison principle, Theorem III.16. This comparison principle holds for f satisfying (H0) under the additional assumption that the subsolution is truncatable, as per Definition III.14. As pointed out in Remark III.15, the value function U is truncatable, so Theorem III.16 is applicable in the proof of Lemma III.18.

So that our results apply to the Hamilton-Jacobi equations that appear later in Chapter V, we will consider the following general Hamilton-Jacobi equation

$$\begin{cases} H(x, Du) = 0 \quad \text{on } \mathbb{R}^d_+, \\ \qquad\qquad u = \varphi \quad \text{on } \partial\mathbb{R}^d_+. \end{cases} \tag{3.40}$$

Here, $\varphi : \partial\mathbb{R}^d_+ \to \mathbb{R}$ is continuous and Pareto-monotone, $H : \mathbb{R}^d_+ \times \mathbb{R}^d \to \mathbb{R}$ is the Hamiltonian, and $u : [0,\infty)^d \to \mathbb{R}$ is the unknown function.

For now, we place the following assumptions on H:

(H1) For every $x \in \mathbb{R}^d_+$, the mapping $H(x,\cdot) : \mathbb{R}^d \to \mathbb{R}$ is monotone non-decreasing.

(H2) There exists a modulus of continuity m such that

$$H(x,p) - H(y,p) \leq m(|p||x-y| + |x-y|) \tag{3.41}$$

for all $p \in [0,\infty)^d$ and $x, y \in \mathbb{R}^d_+$.

Assumption (H1) generalizes our Hamlitonian $H(x,p) = p_1 \cdots p_d - f(x)$, and assumption (H2) is the standard spatial regularity assumption [10, 27]. We will relax

(H2) later to allow for discontinuous spatial dependence.

We now give a comparison principle for Hamiltonians H satisfying (H1) and (H2).

Theorem III.11. *Suppose that H satisfies (H1) and (H2). Let $u \in USC([0,\infty)^d)$ be a viscosity solution of*

$$H(x, Du) \leq 0 \quad \text{in } \mathbb{R}^d_+, \tag{3.42}$$

let $v \in LSC([0,\infty)^d)$ be a Pareto-monotone viscosity solution of

$$H(x, Dv) \geq a \quad \text{in } \mathbb{R}^d_+, \tag{3.43}$$

where $a > 0$, and suppose that $u \leq v$ on $\partial\mathbb{R}^d_+$. Then $u \leq v$ on $\mathbb{R}^d_+$.

The proof of Theorem III.11 is based on the auxiliary function technique, which is standard in the theory of viscosity solutions [27, 10], with modifications to incorporate the lack of compactness resulting from the unbounded domain $\mathbb{R}^d_+$. A standard technique for dealing with unbounded domains is to assume the Hamiltonian H is uniformly continuous in the gradient p and modify the auxiliary function (see, for example [10, Theorem 3.5]). Since (P) is not uniformly continuous in the gradient, we cannot use this technique. In our previous work [20], we included an additional boundary condition at infinity to induce compactness. It turns out that this is not necessary, and in the proof of Theorem III.11, we instead heavily exploit the structure of the Hamiltonian, namely (H1), to produce the required compactness.

Proof. Since v is Pareto-monotone, it is bounded below by $v(0)$. Without loss of generality we may assume that $v(0) = 0$. Let $h > 0$ and set $v_h(x) = v(x) + h(x_1 + x_2)$. It follows from (H1) that v_h is a viscosity solution of (3.43). Assume that

$\sup_{\mathbb{R}^d_+}(u - v_h) > 0$. Let $\Psi : \mathbb{R} \to \mathbb{R}$ be a C^1 function satisfying

$$\begin{cases} \Psi(t) = t & \text{for all } t \leq 1, \\ \Psi(t) \leq 2 & \text{for all } t \in \mathbb{R}, \\ 0 < \Psi'(t) \leq 1 & \text{for all } t \in \mathbb{R}. \end{cases} \tag{3.44}$$

For $c > 0$ set $\overline{u}(x) = c\Psi(c^{-1}u(x))$, and choose c large enough so that

$$\delta := \sup_{\mathbb{R}^d_+}(\overline{u} - v_h) > 0.$$

Since Ψ is C^1 and $\Psi' > 0$, it is a standard application of the chain rule [10] to show that $\overline{u}$ is a viscosity solution of

$$H\left(x, \Psi'(c^{-1}u(x))^{-1}D\overline{u}\right) \leq 0 \ \text{ on } \mathbb{R}^d_+. \tag{3.45}$$

Since $\Psi'(t) \in (0, 1]$ for all $t \geq 0$, we can apply (H1) to (3.45) to find that $\overline{u}$ is a viscosity solution of (3.42).

For $\alpha > 0$ we define

$$\Phi_\alpha(x, y) = \overline{u}(x) - v_h(y) - \frac{\alpha}{2}|x - y|^2, \tag{3.46}$$

and $M_\alpha = \sup_{\mathbb{R}^d_+ \times \mathbb{R}^d_+} \Phi_\alpha$. Since $\overline{u} \leq 2c$ and $v_h \geq 0$, we have by (3.46) that

$$|x - y| \leq \frac{2c}{\sqrt{\alpha}} \ \text{ whenever } \Phi_\alpha(x, y) \geq 0. \tag{3.47}$$

Since $v_h(y) \geq h(y_1 + y_2)$ we have

$$\Phi_\alpha(x, y) \leq 2c - h(y_1 + y_2). \tag{3.48}$$

Since Φ_α is upper semicontinuous and $M_\alpha \geq \delta > 0$, it follows from (3.47) and (3.48) that for every $\alpha > 0$ there exist $x_\alpha, y_\alpha \in [0, \infty)^d$ such that

$$\Phi_\alpha(x_\alpha, y_\alpha) = M_\alpha \geq \delta > 0, \tag{3.49}$$

and

$$y_{\alpha,1} + y_{\alpha,2} \leq \frac{2c}{h}. \tag{3.50}$$

Furthermore, by (3.47) and (3.50) we see that, upon passing to a subsequence if necessary, we have $x_\alpha, y_\alpha \to x_0$ for some $x_0 \in [0,\infty)^d$. Since $(x,y) \mapsto \overline{u}(x) - v_h(y)$ is upper semicontinuous we have

$$\limsup_{\alpha\to\infty} M_\alpha \leq \limsup_{\alpha\to\infty} \overline{u}(x_\alpha) - v_h(y_\alpha) \leq \overline{u}(x_0) - v_h(x_0).$$

Since $M_\alpha \geq \overline{u}(x_0) - v_h(x_0)$ for all α we have that $M_\alpha \to u(x_0) - v(x_0) = \delta > 0$ as $\alpha \to \infty$ and hence

$$\alpha|x_\alpha - y_\alpha|^2 \longrightarrow 0. \tag{3.51}$$

Since $\overline{u} \leq v_h$ on $\partial\mathbb{R}^d_+$ we must have $x_0 \in \mathbb{R}^d_+$, and therefore $x_\alpha, y_\alpha \in \mathbb{R}^d_+$ for α large enough.

Set $p = \alpha(x_\alpha - y_\alpha)$. By (3.49) we have that

$$p \in D^+\overline{u}(x_\alpha) \cap D^- v_h(y_\alpha).$$

Therefore we have

$$H(x_\alpha, p) \leq 0 \ \text{ and } \ H(y_\alpha, p) \geq a.$$

Subtracting the above inequalities and invoking (H2) we have

$$0 < a \leq H(y_\alpha,p) - H(x_\alpha,p) \leq m(|p||x_\alpha - y_\alpha| + |x_\alpha - y_\alpha|) \leq m(\alpha|x_\alpha - y_\alpha|^2 + |x_\alpha - y_\alpha|).$$

Sending $\alpha \to \infty$ we arrive at a contradiction. Therefore $u \leq v_h$, and sending $h \to 0^+$ completes the proof. □

We now specialize Theorem III.11 to our PDE (P).

Corollary III.12. *Suppose $f : \mathbb{R}^d_+ \to \mathbb{R}$ is uniformly continuous with* $\mathrm{supp}(f) \subset [0,1]^d$. *Let $u \in USC([0,\infty)^d)$ and $v \in LSC([0,\infty)^d)$ be viscosity sub- and supersolutions, respectively, of*

$$u_{x_1} \cdots u_{x_d} = f \quad \text{on} \quad \mathbb{R}^d_+, \tag{3.52}$$

and suppose that v is Pareto-monotone. If $u \le v$ on $\partial\mathbb{R}^d_+$ then $u \le v$ on $\mathbb{R}^d_+$.

Proof. Consider the Hamiltonian

$$H(x,p) = \max(p_1, 0) \cdots \max(p_d, 0) - f(x).$$

It is easy to verify that H satisfies (H1) and (H2), and we have $H(x, Du) \le 0$ in the viscosity sense.

For $\theta > 0$, set $v_\theta(x) = v(x) + \theta^{\frac{1}{d}}(x_1 + \cdots x_d)$. Fix $x \in \mathbb{R}^d_+$ and $p \in D^- v_\theta(x)$. It is easy to see that $p - \theta^{\frac{1}{d}}(1, \ldots, 1) \in D^- v(x)$. Hence we have

$$(p_1 - \theta^{\frac{1}{d}}) \cdots (p_d - \theta^{\frac{1}{d}}) - f(x) \ge 0.$$

Since v is Pareto-monotone, it follows from Proposition III.9 that $p_i - \theta^{\frac{1}{d}} \ge 0$ for all i, and therefore

$$\begin{aligned} p_1 \cdots p_d = (p_1 - \theta^{\frac{1}{d}} + \theta^{\frac{1}{d}}) \cdots (p_d - \theta^{\frac{1}{d}} + \theta^{\frac{1}{d}}) &\ge (p_1 - \theta^{\frac{1}{d}}) \cdots (p_d - \theta^{\frac{1}{d}}) + \theta \\ &\ge f(x) + \theta. \end{aligned}$$

Since $p_i \ge 0$ for all i, it follows that $H(x, Dv_\theta) \ge \theta > 0$ in the viscosity sense. Since $u \le v \le v_\theta$ on $\partial\mathbb{R}^d_+$, we can apply Theorem III.11 to find that $u \le v_\theta$ on $\mathbb{R}^d_+$. The proof is completed by sending $\theta \to 0$. □

We now aim to extend this comparison principle to Hamiltonians with discontinuous spatial dependence. We first introduce some notation. Let us denote an

interval associated to $\leqq$ by

$$[x, y] = \{z \in \mathbb{R}^d \,:\, x \leqq z \leqq y\},$$

where $x, y \in \mathbb{R}^d$ and $x \leqq y$. For $\xi \in [0, \infty)^d$, let us denote by $\pi_\xi : \mathbb{R}^d \to [0, \xi]$ the projection mapping $[0, \infty)^d$ onto the convex set $[0, \xi]$. For $x \in [0, \infty)^d$, we have

$$\pi_\xi(x) = \Big(\min(x_1, \xi_1), \ldots, \min(x_d, \xi_d) \Big).$$

We now make the following definitions.

Definition III.13. Given a function $u : [0, \infty)^d \to \mathbb{R}$ and $\xi \in [0, \infty)^d$, we define the *ξ-truncation* of u by $u^\xi := u \circ \pi_\xi$.

Definition III.14. Let u be a viscosity solution of

$$H(x, Du) \leq 0 \quad \text{on } \mathbb{R}^d_+. \tag{3.53}$$

We say that u is *truncatable* if for every $\xi \in \mathbb{R}^d_+$, the ξ-truncation u^ξ is a viscosity solution of (3.53).

Remark III.15. We remark that the value function U given by (3.5) is a truncatable viscosity solution of (3.36). To see this, fix $\xi \in \mathbb{R}^d_+$ and set $U^\xi = U \circ \pi_\xi$. Then it follows from Theorem III.10 and (3.5) that

$$U^\xi(x) = \sup_{\gamma \in \mathcal{A} \,:\, \gamma \leqq x} \int_0^1 \chi(\gamma(t)) f(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}} \, dt,$$

where $\chi : \mathbb{R}^d \to \{0, 1\}$ is the characteristic function of $[0, \xi]$. Therefore we have

$$U^\xi_{x_1} \cdots U^\xi_{x_d} = \chi f \leq f \quad \text{on} \quad \mathbb{R}^d$$

in the viscosity sense, hence U is truncatable.

We now relax (H2) and allow the Hamiltonian H to have discontinuous spatial dependence. Given a set $\mathcal{O} \subset \mathbb{R}^d_+$ we assume H satisfies

(H3)$_\mathcal{O}$ There exists a modulus of continuity m such that for all $\xi \in \mathcal{O}$ there exists $\varepsilon_\xi > 0$ and $\mathbf{v}_\xi \in \mathbb{S}^{d-1}$ such that

$$H(y,p) - H(y+\varepsilon\mathbf{v},p) \leq m(|p|\varepsilon+\varepsilon) \tag{3.54}$$

for all $p \in \mathbb{R}^d$, $y \in B_{\varepsilon_\xi}(\xi)$, $\varepsilon \in (0,\varepsilon_\xi)$, and $\mathbf{v} \in \mathbb{S}^{d-1}$ with $|\mathbf{v}-\mathbf{v}_\xi| < \varepsilon_\xi$.

This hypothesis is similar to one used by Deckelnick and Elliott [30] to prove uniqueness of viscosity solutions to Eikonal-type Hamilton-Jacobi equations with discontinuous spatial dependence.

If we assume the subsolution is truncatable, then we can prove the following comparison principle, which holds for Hamiltonians H with discontinuous spatial dependence.

Theorem III.16. *Suppose that H satisfies (H3)$_\mathcal{O}$ for some $\mathcal{O} \subset \mathbb{R}^d_+$. Let $u \in C([0,\infty)^d)$ be a truncatable viscosity solution of (3.42) and let $v \in C([0,\infty)^d)$ be a Pareto-monotone viscosity solution of (3.43). Suppose that $u \leq v$ on $[0,\infty)^d \setminus \mathcal{O}$. Then $u \leq v$ on $\mathbb{R}^d_+$.*

As in the proof of Theorem III.11, the proof below is based on the standard auxiliary function technique [27]. The proof is similar to [30, Theorem 2.3] in the way that (H3)$_\mathcal{O}$ is used, however, we cannot assume Lipschitzness of v. The truncatability condition on u in a sense replaces the Lipschitz condition on v in [30, Theorem 2.3].

Proof. Suppose that

$$\lambda := \sup_{\mathbb{R}^d_+}(u-v) > 0.$$

Let

$$R = \sup\left\{ r > 0 \,:\, u \le v + \frac{\lambda}{2} \ \text{ on } \ D_r \right\}, \tag{3.55}$$

where

$$D_r = \{ x \in \mathbb{R}^d_+ \,:\, x_1 + \cdots + x_d < r \}. \tag{3.56}$$

Since $\mathcal{O} \subset \mathbb{R}^d_+$, we have by hypothesis that $u \le v$ on $\partial\mathbb{R}^d_+$. Therefore, since u and v are continuous we have $R \in (0, \infty)$. By (3.55) there exists $\xi_0 \in \mathbb{R}^d_+ \cap \partial D_R$ such that

$$\begin{aligned} &u(\xi_0) = v(\xi_0) + \frac{\lambda}{2} \ \text{ and} \\ &\text{every neighborhood of } \xi_0 \text{ contains some } y \in \mathbb{R}^d_+ \text{ with } u(y) > v(y) + \frac{\lambda}{2}. \end{aligned} \tag{3.57}$$

For $t > 0$ set $\xi = \xi_0 + (t, \ldots, t)$ and

$$\mathcal{G} = \{ x \in [0, \infty)^d \,:\, x \leqq \xi \}. \tag{3.58}$$

Let u^ξ denote the ξ-truncation of u. By (3.57) and (3.55) we see that

$$\delta := \sup_{\mathbb{R}^d_+} (u^\xi - v) > \frac{\lambda}{2} > 0 \tag{3.59}$$

By (3.57) we have $u(\xi_0) > v(\xi_0)$, and hence $\xi_0 \in \mathcal{O}$. Let ε_{ξ_0} and $\mathbf{v}_{\xi_0} \in \mathbb{S}^1$ be as given in $(\mathrm{H3})_{\mathcal{O}}$. Choose $t > 0$ small enough, and $\varepsilon_{\xi_0} > 0$ smaller if necessary, so that $\mathcal{G} \setminus D_R \in B_{\varepsilon_{\xi_0}}(\xi_0) \subset \mathbb{R}^d_+$. For $\alpha > 0$ define

$$\Phi_\alpha(x, y) = u^\xi(x) - v(y) - \frac{\alpha}{2} \left| x - y - \frac{1}{\sqrt{\alpha}} \mathbf{v}_{\xi_0} \right|^2. \tag{3.60}$$

We claim that for α large enough, there exists $x_\alpha, y_\alpha \in B_{\varepsilon_{\xi_0}}(\xi_0)$ such that

$$M_\alpha := \sup_{\mathbb{R}^d_+ \times \mathbb{R}^d_+} \Phi_\alpha = \Phi_\alpha(x_\alpha, y_\alpha). \tag{3.61}$$

To see this, first substitute $y = x - \frac{1}{\sqrt{\alpha}} \mathbf{v}_{\xi_0}$ into (3.61) to find

$$M_\alpha \ge u^\xi(x) - v\left(x - \frac{1}{\sqrt{\alpha}} \mathbf{v}_{\xi_0} \right),$$

for any $x \in \mathbb{R}^d_+$ such that $x - \frac{1}{\sqrt{\alpha}} \in \mathbb{R}^d_+$. Since u^ξ and v are continuous, it follows from (3.59) that

$$\liminf_{\alpha\to\infty} M_\alpha \geq \sup_{\mathbb{R}^d_+}(u^\xi - v) = \delta > \frac{\lambda}{2} > 0. \tag{3.62}$$

Since u^ξ is bounded, and v is monotone, we have by (3.60) that

$$|x - y| \leq \frac{C}{\sqrt{\alpha}} \text{ whenever } \Phi_\alpha(x, y) \geq 0. \tag{3.63}$$

Let $x, y \in \mathbb{R}^d_+$ such that $\Phi_\alpha(x, y) \geq 0$. Set $w = \pi_x(y) = \pi_y(x)$ and $\widehat{w} = \pi_\xi(w)$, and define

$$\widehat{x} = x + \widehat{w} - w \text{ and } \widehat{y} = y + \widehat{w} - w. \tag{3.64}$$

A short calculation shows that $\pi_\xi(x) = \pi_\xi(\widehat{x})$. Since $u^\xi = u \circ \pi_\xi$ we have

$$u^\xi(\widehat{x}) = u^\xi(\pi_\xi(\widehat{x})) = u^\xi(\pi_\xi(x)) = u^\xi(x). \tag{3.65}$$

Since v is Pareto-monotone and $\widehat{y} \leqq y$ we have by (3.65) that

$$u^\xi(\widehat{x}) - v(\widehat{y}) \geq u^\xi(x) - v(y) \tag{3.66}$$

Since $\widehat{x} - \widehat{y} = x - y$, we see from (3.66) and (3.60) that

$$\Phi_\alpha(\widehat{x}, \widehat{y}) \geq \Phi_\alpha(x, y). \tag{3.67}$$

Furthermore, by (3.63) we have

$$|\widehat{x} - \widehat{w}| = |x - w| \leq |x - y| \leq \frac{C}{\sqrt{\alpha}}.$$

Similarly we have $|\widehat{y} - \widehat{w}| \leq \frac{C}{\sqrt{\alpha}}$. Since $\widehat{w} \leqq \xi$ we have

$$\widehat{x}, \widehat{y} \in \mathcal{G}_\alpha := \left\{ x' \in [0, \infty)^d \,:\, x' \leqq \xi + \frac{C}{\sqrt{\alpha}}(1, \ldots, 1) \right\}.$$

It follows from this and (3.67) that for every $\alpha > 0$ there exists $x_\alpha, y_\alpha \in \mathcal{G}_\alpha$ such that $M_\alpha = \Phi_\alpha(x_\alpha, y_\alpha)$. By (3.63) we may pass to a subsequence if necessary to find $x_0 \in \mathcal{G}$ such that $x_\alpha, y_\alpha \to x_0$ as $\alpha \to \infty$. Then we have

$$\limsup_{\alpha\to\infty} M_\alpha \leq \lim_{\alpha\to\infty} u^\xi(x_\alpha) - v(y_\alpha) \leq \delta.$$

Combining this with (3.62) we have $M_\alpha \to \delta = u^\xi(x_0) - v(x_0)$ and

$$\frac{\alpha}{2}\left|x_\alpha - y_\alpha - \frac{1}{\sqrt{\alpha}}\mathbf{v}_{\xi_0}\right|^2 \longrightarrow 0. \tag{3.68}$$

Since $\delta > \lambda/2$, it follows from the definition of R (3.55) that $x_0 \in \mathcal{G} \setminus D_R \subset B_{\varepsilon_{\xi_0}}(\xi_0)$. Therefore, for $\alpha > 0$ large enough we have $x_\alpha, y_\alpha \in B_{\varepsilon_{\xi_0}}(\xi_0)$, which establishes the claim.

Letting $p = \alpha\left(x_\alpha - y_\alpha - \frac{1}{\sqrt{\alpha}}\right)$ we have by (3.61) that $p \in D^+u^\xi(x_\alpha) \cap D^-v(y_\alpha)$. By hypothesis we have

$$H^*(y_\alpha, p) \geq a. \tag{3.69}$$

Since u is truncatable, u^ξ is a viscosity solution of (3.42) and therefore

$$H_*(x_\alpha, p) \leq 0. \tag{3.70}$$

Subtracting (3.70) from (3.69) we have

$$a \leq H^*(y_\alpha, p) - H_*(x_\alpha, p). \tag{3.71}$$

Let $w_\alpha = x_\alpha - y_\alpha - \frac{1}{\sqrt{\alpha}}\mathbf{v}_{\xi_0}$ and note that

$$x_\alpha = y_\alpha + \varepsilon\mathbf{v},$$

where

$$\varepsilon = \frac{1}{\sqrt{\alpha}}|\mathbf{v}_{\xi_0} + \sqrt{\alpha}w_\alpha| = |x_\alpha - y_\alpha| \quad \text{and} \quad \mathbf{v} = \frac{\mathbf{v}_{\xi_0} + \sqrt{\alpha}w_\alpha}{|\mathbf{v}_{\xi_0} + \sqrt{\alpha}w_\alpha|}.$$

By (3.68) we have $\sqrt{\alpha}w_\alpha \to 0$. Therefore, for α large enough we have $|\mathbf{v}_{\xi_0} - \mathbf{v}| < \varepsilon_{\xi_0}$ and $\varepsilon < \varepsilon_{\xi_0}$. Since $y_\alpha \in B_{\varepsilon_{\xi_0}}(\xi_0)$ we can invoke $(\mathrm{H3})_{\mathcal{O}}$ to find that

$$H^*(y_\alpha, p) - H_*(x_\alpha, p) = H^*(y_\alpha, p) - H_*(y_\alpha + \varepsilon\mathbf{v}, p) \leq m(|p||x_\alpha - y_\alpha| + |x_\alpha - y_\alpha|). \tag{3.72}$$

Note that

$$\begin{aligned} |p||x_\alpha - y_\alpha| &= \alpha \left| x_\alpha - y_\alpha - \frac{1}{\sqrt{\alpha}}\mathbf{v}_{\xi_0} \right| |x_\alpha - y_\alpha| \\ &= \alpha \left| x_\alpha - y_\alpha - \frac{1}{\sqrt{\alpha}}\mathbf{v}_{\xi_0} \right| \left| x_\alpha - y_\alpha - \frac{1}{\sqrt{\alpha}}\mathbf{v}_{\xi_0} + \frac{1}{\sqrt{\alpha}}\mathbf{v}_{\xi_0} \right| \\ &\leq \alpha w_\alpha^2 + \sqrt{\alpha} w_\alpha. \end{aligned}$$

Combining this with (3.72) and (3.71) we have

$$0 < a \leq m(\alpha w_\alpha^2 + \sqrt{\alpha} w_\alpha + |x_\alpha - y_\alpha|).$$

Sending $\alpha \to \infty$ yields a contradiction. □

We now specialize Theorem III.16 to our Hamilton-Jacobi equation (P). Let us first recall the assumption on $f : \mathbb{R}^d \to [0, \infty)$:

(H0) There exists an open and bounded set $\Omega \subset \mathbb{R}^d_+$ with Lipschitz boundary such that f is non-negative and uniformly continuous on Ω and $f = 0$ on $\mathbb{R}^d \setminus \Omega$.

The assumption (H0) implies that $(\mathrm{H3})_{\mathbb{R}^d_+}$ holds. To see this: We only need to verify $(\mathrm{H3})_{\mathbb{R}^d_+}$ for $\xi \in \partial\Omega$, since f is uniformly continuous away from $\partial\Omega$. For any $\xi \in \partial\Omega$ there exists, by Lipchitzness of $\partial\Omega$, a real number $r > 0$ and a Lipschitz continuous function $\Psi : \mathbb{R}^{d-1} \to \mathbb{R}$ such that, upon relabelling and reorienting the coordinate axes if necessary, we have

$$\Omega \cap B_r(\xi) = \Big\{ y \in B_r(\xi) \,:\, y_d < \Psi(y_1, \ldots, y_{d-1}) \Big\}.$$

One can then use any $\mathbf{v}_\xi \in \mathbb{S}^{d-1}$ in the interior of the cone

$$\mathcal{K}_x = \left\{ y \in \mathbb{R}^d \,:\, y_d \geq 2\mathrm{Lip}(\Psi)\sqrt{y_1^2 + \cdots y_{d-1}^2} \right\}$$

provided ε_ξ is chosen small enough.

Corollary III.17. *Assume f satisfies (H0). Let u and v be viscosity sub- and supersolutions, respectively, of*

$$u_{x_1} \cdots u_{x_d} = f \quad \text{on} \quad \mathbb{R}^d_+, \tag{3.73}$$

and assume that u is truncatable and v is Pareto-monotone. Then $u \leq v$ on $\partial\mathbb{R}^d_+$ implies that $u \leq v$ on $\mathbb{R}^d_+$.

Proof. For $\theta > 0$, set $v_\theta(x) = v(x) + \theta^{\frac{1}{d}}(x_1 + \cdots + x_d)$. Then $u < v_\theta$ on $\partial\mathbb{R}^d_+$. As in the proof of Corollary III.12, v_θ satisfies

$$v_{\theta,x_1} \cdots v_{\theta,x_d} \geq f + \theta \quad \text{on} \quad \mathbb{R}^d_+ \tag{3.74}$$

in the viscosity sense. As remarked above, (H0) implies that $(\mathrm{H3})_{\mathbb{R}^d_+}$ holds, hence we can apply Theorem III.16 to find that $u \leq v_\theta$. Sending $\theta \to 0$ completes the proof. □

In order to prove a general uniqueness result without the truncatability assumption we require a perturbation result for the value function U with respect to sup and inf convolutions of the density f. Since a similar result, for a different type of perturbation, is required in the proof of our main result, Theorem III.2, we state a more general result in Lemma III.18. We first recall some notation standard in the theory of viscosity solutions. For a sequence of bounded functions $f_n : \mathbb{R}^d_+ \to \mathbb{R}$, the upper and lower limits are defined by

$$\limsup_{n\to\infty}{}^* f_n(x) := \lim_{j\to\infty} \sup \left\{ f_n(y) \,:\, n \geq j,\ y \in \mathbb{R}^d, \ \text{ and } \ |x - y| \leq \frac{1}{j} \right\},$$

and

$$\liminf_{n\to\infty *} f_n(x) := \lim_{j\to\infty} \inf \left\{ f_n(y) \,:\, n \geq j,\ y \in \mathbb{R}^d, \ \text{ and } \ |x-y| \leq \frac{1}{j} \right\}.$$

Lemma III.18. *Assume f satisfies (H0). Let $\{f_n\}_{n=1}^{\infty} \subset \mathcal{B}$ and suppose that*

$$f_* \leq \liminf_{n\to\infty *} f_n \quad \textit{and} \quad \limsup_{n\to\infty}{}^{*} f_n \leq f^*. \tag{3.75}$$

For each n, set

$$v_n(x) = \sup_{\gamma\in\mathcal{A}\,:\,\gamma\leqq x} \int_0^1 f_n(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t)\cdots\gamma_d'(t))^{\frac{1}{d}}\, dt.$$

Then $v_n \to U$ uniformly where U is the value function given by (3.5).

Proof. We claim that $\{f_n\}_{n=1}^{\infty}$ is a uniformly bounded sequence. To see this, suppose to the contrary that there exists a sequence x_n in $[0,1]^d$ such that $f_n(x_n) \to \infty$ as $n \to \infty$. By passing to a subsequence, if necessary, we may assume that $x_n \to x_0 \in [0,1]^2$ as $n \to \infty$. By the definition of the upper limit and (3.75), we have

$$f^*(x_0) \geq \limsup_{n\to\infty}{}^{*} f_n(x_0) = \infty,$$

which contradicts the assumption that f is uniformly continuous on Ω and hence bounded. This establishes the claim.

Since $\{f_n\}_{n=1}^{\infty}$ is uniformly bounded, there exists (by Lemma III.3) a constant C such that $[v_n]_{\frac{1}{d}} \leq C$ for all n. The sequence v_n is therefore bounded and equicontinuous, and by the Arzela-Ascoli theorem there exists a subsequence v_{n_k} and a Hölder-continuous function $v : \mathbb{R}^d \to \mathbb{R}$ such that $v_{n_k} \to v$ uniformly on compact sets in $\mathbb{R}^d$ as $k \to \infty$. By Theorem III.10 (i), (ii), we conclude that the convergence is actually uniform on $\mathbb{R}^d$. By Theorem III.10, each v_n is a Pareto-monotone truncatable viscosity solution of

$$v_{n,x_1}\cdots v_{n,x_d} = \frac{1}{d^d} f_n \quad \text{on} \quad \mathbb{R}^d_+.$$

By standard results on viscosity solutions (see [27, Remark 6.3]) and (3.75), we have that v is a Pareto-monotone viscosity solution of

$$v_{x_1} \cdots v_{x_d} = \frac{1}{d^d} f \quad \text{on} \quad \mathbb{R}^d_+.$$

By the assumption that $\operatorname{supp}(f_n) \subset [0,1]^d$, we have that $v_n(x) = 0$ for all $x \notin \mathbb{R}^d_+$, hence $v(x) = 0$ for all $x \notin \mathbb{R}^d_+$.

We claim that v is truncatable. To see this, fix $\xi \in \mathbb{R}^d_+$ and define $v^\xi = v \circ \pi_\xi$, $v^\xi_n = v_n \circ \pi_\xi$, and let χ denote the characteristic function of $[0, \xi]$. Since v_n is truncatable, v^ξ_n is a viscosity solution of

$$v^\xi_{n,x_1} \cdots v^\xi_{n,x_d} \leq \frac{1}{d^d} f_n \quad \text{on} \quad \mathbb{R}^d_+.$$

Since (3.75) holds and $v^\xi_n \to v^\xi$ uniformly, we can again apply standard results on viscosity solutions [27] to find that v^ξ is a viscosity subsolution of

$$v^\xi_{x_1} \cdots v^\xi_{x_d} = \frac{1}{d^d} f \quad \text{on} \quad \mathbb{R}^d_+,$$

which proves the claim.

By Corollary III.17 we have $v = U$ on $\mathbb{R}^d_+$. Since $U(x) = v(x) = 0$ for $x \notin \mathbb{R}^d_+$ we have $v = U$ on $\mathbb{R}^d$. The above argument can be used to show that every subsequence of v_n contains a uniformly convergent subsequence converging to U. It follows that $v_n \to U$ uniformly in $\mathbb{R}^d$ as $n \to \infty$. □

We now establish uniqueness of viscosity solutions of (P).

Theorem III.19. *Assume f satisfies (H0). Then there exists a unique Pareto-monotone viscosity solution u of*

$$\begin{cases} u_{x_1} \cdots u_{x_d} = f & \text{on } \mathbb{R}^d_+ \\ u = 0 & \text{on } \partial\mathbb{R}^d_+. \end{cases} \tag{3.76}$$

Proof. Existence is established by Theorem III.10. To prove uniqueness, let u be a Pareto-monotone viscosity solution of (3.76). We will show that $u = d \cdot U$, where U is the value function defined by (3.5). Let $\varepsilon > 0$ and consider the inf and sup convolutions of f, defined for $x \in \mathbb{R}^d_+$ by

$$f_\varepsilon(x) = \inf_{y \in \mathbb{R}^d_+} \left\{ f(y) + \frac{1}{\varepsilon}|x-y| \right\} \quad \text{and} \quad f^\varepsilon(x) = \sup_{y \in \mathbb{R}^d_+} \left\{ f(y) - \frac{1}{\varepsilon}|x-y| \right\}.$$

Recall that f_ε and f^ε are Lipschitz continuous with constant $1/\varepsilon$ and $f_\varepsilon \leq f \leq f^\varepsilon$. Without loss of generality, we may assume that $\overline{\Omega} \subset (0,1)^d$, and hence for $\varepsilon > 0$ small enough, we have $\operatorname{supp}(f_\varepsilon), \operatorname{supp}(f^\varepsilon) \subset [0,1]^d$. For $x \in \overline{\mathbb{R}^d_+}$, set

$$U^\varepsilon(x) = \sup_{\gamma \in \mathcal{A} \colon \gamma \leqq x} \int_0^1 f^\varepsilon(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}} \, dt,$$

and

$$U_\varepsilon(x) = \sup_{\gamma \in \mathcal{A} \colon \gamma \leqq x} \int_0^1 f_\varepsilon(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}} \, dt.$$

By Theorem III.10, $d \cdot U^\varepsilon$ is a viscosity solution of

$$v_{x_1} \cdots v_{x_d} = f^\varepsilon \quad \text{on} \quad \mathbb{R}^d_+. \tag{3.77}$$

Since $f \leq f^\varepsilon$ and u is a viscosity solution of (3.76), we see that u is a viscosity subsolution of (3.77). Since $u = U = 0$ on $\partial \mathbb{R}^d_+$ we can apply Corollary III.12 to find that $u \leq d \cdot U^\varepsilon$. By a similar argument, we have that $u \geq d \cdot U_\varepsilon$. Since $f_\varepsilon, f^\varepsilon \in \mathcal{B}$ and (3.75) is satisfied for the sequences $\{f^\varepsilon\}_{\varepsilon > 0}$ and $\{f_\varepsilon\}_{\varepsilon > 0}$, we have by Lemma III.18 that $U_\varepsilon, U^\varepsilon \to U$ uniformly in $\mathbb{R}^d$ as $\varepsilon \to 0$, and hence $u = d \cdot U$. □

3.5 Large sample asymptotics of u_n

The proof of Theorem III.2 is split into several steps. In Section 3.5.1, we prove a basic convergence result for piecewise constant density functions, which is a generalization of the results of Deuschel and Zeitouni [31]. In Section 3.5.2, we extend

the convergence result to densities that are continuous on Ω and vanish on $\mathbb{R}^d \setminus \Omega$ by considering a sequence of piecewise constant approximations to f, applying the results from Section 3.5.1, and passing to the limit. This requires a perturbation result for the energy J, which we obtained from the comparison principle for the associated Hamilton-Jacobi PDE (P) in Lemma III.18.

3.5.1 Piecewise constant densities

We aim to prove a basic convergence result for piecewise constant densities here. The proof is split into a lower bound, Theorem III.20, and an upper bound, Theorem III.23. We should note that the techniques used here are similar to those used by Deuschel and Zeitouni [31], who showed the same convergence result for C^1 densities on the unit hypercube in dimension $d = 2$.

Let us introduce some notation. For a finite set $S \subset \mathbb{R}^d$, let $\ell(S)$ denote the length of a longest increasing chain in S. The set function ℓ has an important invariance. If $\Psi : \mathbb{R}^d \to \mathbb{R}^d$ is a mapping that preserves the partial order $\leqq$, i.e., $x \leqq y \iff \Psi(x) \leq \Psi(y)$, then

$$\ell(S) = \ell(\Psi(S)) \text{ for any } S \subset \mathbb{R}^d. \tag{3.78}$$

For $A \subset \mathbb{R}^d$ we denote by $\chi_A : \mathbb{R}^d \to \mathbb{R}$ the characteristic function of the set A, which takes the value 1 on A and 0 on $\mathbb{R}^d \setminus A$. When A is Lebesgue measurable, we denote by $|A|$ the Lebesgue measure of A. We set $\mathbf{0}_d = (0, \ldots, 0) \in \mathbb{R}^d$ and $\mathbf{1}_d = (1, \ldots, 1) \in \mathbb{R}^d$. Given an integer L, we partition $[0,1)^d$ into L^d hypercubes of side length $1/L$. More precisely, for a multiindex $\alpha \in \mathbb{N}^L$ with $\|\alpha\|_\infty \leq L$, where $\|\alpha\|_\infty = \max(\alpha_1, \ldots, \alpha_L)$, we set

$$Q_{L,\alpha} = \{x \in [0,1)^d \,:\, \alpha - \mathbf{1}_d \leqq Lx < \alpha\}. \tag{3.79}$$

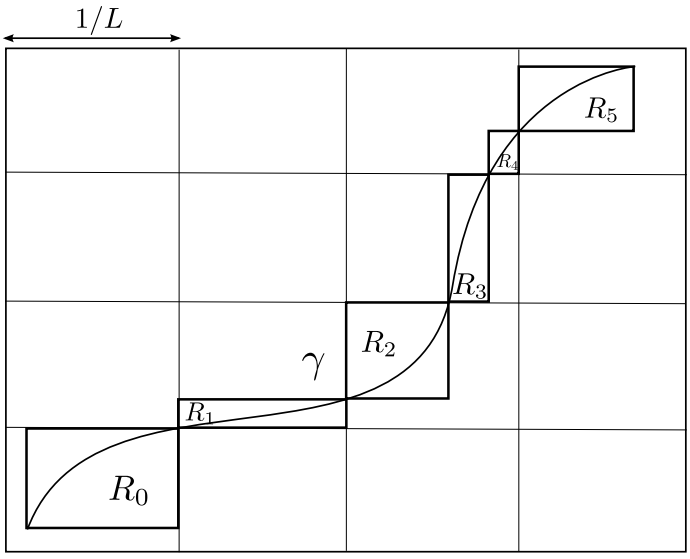


Figure 3.3: An illustration of some quantities from the proof of Theorem III.20.

We say that $f : [0,1)^d \to [0,\infty)$ is *L-piecewise constant* if f is constant on $Q_{L,\alpha}$ for all α. If f is L-piecewise constant then f is kL-piecewise constant for all $k \in \mathbb{N}$. For convenience, we also set

$$\overline{J} = \sup_{\gamma \in \mathcal{A}} J(\gamma) = \sup_{\gamma \in \mathcal{A}} \int_0^1 f(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}} \, dt.$$

We now establish an asymptotic lower bound on $\ell(\{X_1, \ldots, X_n\})$.

Theorem III.20. *Let $f : [0,1)^d \to [0,\infty)$ be L-piecewise constant, and let $X_1, \ldots, X_n$ be* i.i.d. *with density f. Then*

$$\liminf_{n \to \infty} \; n^{-\frac{1}{d}} \ell\left(\{X_1, \ldots, X_n\}\right) \geq c_d \overline{J} \quad \text{a.s.}$$

Proof. Let $\varepsilon > 0$ and select $\gamma \in \mathcal{A}$ with $J(\gamma) \geq \overline{J} - \frac{\varepsilon}{c_d}$. Without loss of generality, we may assume that $\gamma\,'(t) > 0$ for all $t \in [0,1]$. Let $s_1, \ldots, s_k$ denote the $k \leq dL$ times at which γ intersects the set

$$\bigcup_{\alpha} \partial Q_{L,\alpha} \cap (0,1)^d.$$

Set $s_0 = 0$ and $s_{k+1} = 1$. For $j = 0, \ldots, k$ set $I_j = [s_j, s_{j+1})$ and

$$R_j = \{x \in [0,1)^d \,:\, \gamma(s_j) \leqq x < \gamma(s_{j+1})\}. \tag{3.80}$$

See Figure 3.3 for an illustration of these quantities. For every j we have $R_j \subset Q_{L,\alpha}$ for some α. Recalling the definition of J (3.4) we have

(3.81)
$$J(\gamma) = \sum_{j=0}^{k} \int_{I_j} f(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}}\, dt = \sum_{j=0}^{k} f(\gamma(s_j))^{\frac{1}{d}} \int_{I_j} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}}\, dt,$$

where the second equality follows from the fact that f is constant on $R_j \subset Q_{L,\alpha}$. Applying the generalized Hölder inequality to (3.81) we have

$$J(\gamma) \leq \sum_{j=0}^{k} f(\gamma(s_j))^{\frac{1}{d}} \prod_{i=1}^{d} \left(\int_{I_j} \gamma_i'(t)\, dt \right)^{\frac{1}{d}} = \sum_{j=0}^{k} f(\gamma(s_j))^{\frac{1}{d}} |R_j|^{\frac{1}{d}}.$$

Setting $p_j = \int_{R_j} f(x)\, dx = f(\gamma(s_j))|R_j|$ we have

$$\overline{J} \leq J(\gamma) + \frac{\varepsilon}{c_d} \leq \sum_{j=0}^{k} p_j^{\frac{1}{d}} + \frac{\varepsilon}{c_d}. \tag{3.82}$$

Fix $j \in \{0, \ldots, k\}$. Let n_j denote the number of points from $X_1, \ldots, X_n$ falling in R_j and set

$$\ell_j(n) = \ell\left(\{X_1, \ldots, X_n\} \cap R_j\right). \tag{3.83}$$

Then n_j is Binomially distributed with parameters n and p_j. If f is identically zero on R_j then $\ell_j(n) = 0$ with probability one for all n, and $p_j = 0$, hence

$$n^{-\frac{1}{d}} \ell_j(n) = c_d p_j^{\frac{1}{d}} \quad \text{a.s.} \tag{3.84}$$

If f is not identically zero on R_j, then since $\gamma\,'(t) > 0$ for all t, we have $|R_j| > 0$ and hence $p_j > 0$. The conditional law $\rho_j := p_j^{-1} \cdot f \cdot \chi_{R_j}$ is then uniform on R_j. Let $i_1, \ldots, i_{n_j}$ be the indices of the n_j random variables out of $X_1, \ldots, X_n$ that belong

to R_j. Let $\Psi : R_j \to [0,1)^d$ be the injective affine transformation mapping R_j onto $[0,1)^d$. Then $\Psi(X_{i_1}), \ldots, \Psi(X_{i_{n_j}})$ are independent and uniformly distributed on $[0,1)^d$. By [17, Remark 1], we have

$$n_j^{-\frac{1}{d}} \ell\left(\{\Psi(X_{i_1}), \ldots, \Psi(X_{i_{n_j}})\}\right) \to c_d \quad \text{a.s.}$$

Since Ψ preserves the partial order $\leqq$, we have by (3.78) that

$$\ell_j(n) = \ell\left(\{X_{i_1}, \ldots, X_{i_{n_j}}\}\right) = \ell\left(\{\Psi(X_{i_1}), \ldots, \Psi(X_{i_{n_j}})\}\right).$$

Since $n^{-1} n_j \to p_j$ almost surely we have

$$n^{-\frac{1}{d}} \ell_j(n) = n_j^{-\frac{1}{d}} (n^{-1} n_j)^{\frac{1}{d}} \ell_j(n) \to c_d p_j^{\frac{1}{d}} \quad \text{a.s.} \tag{3.85}$$

Combining this (3.82), (3.84) and (3.85), we see that

$$n^{-\frac{1}{d}} \sum_{j=0}^{k} \ell_j(n) \to c_d \sum_{j=0}^{k} p_j^{\frac{1}{d}} \geq c_d \overline{J} - \varepsilon \quad \text{a.s.} \tag{3.86}$$

Since γ is a monotone curve (i.e., $\gamma'(t) \geq 0$), we can connect longest chains from each rectangle R_j together to form a chain in $[0,1)^d$. It follows that

$$\ell\left(\{X_1, \ldots, X_n\}\right) \geq \sum_{j=0}^{k} \ell_j(n). \tag{3.87}$$

Combining this with (3.86) we have

$$\liminf_{n\to\infty} \; n^{-\frac{1}{d}} \ell\left(\{X_1, \ldots, X_n\}\right) \geq c_d \overline{J} - \varepsilon \quad \text{a.s.},$$

which completes the proof. □

For the proof of the upper bound, we need to introduce some new notation. Let k_1 be an integer and set $\Delta x = 1/k_1$. Let k_2 be another integer and set $\Delta y = \Delta x / k_2$. For given k_1, k_2, we say that a sequence of multiindices $\mathbf{b} = (b_j)_{j=1}^{k_1} \subset \mathbb{N}^{d-1}$ is *admissible* if $b_1 \leqq \cdots \leqq b_{k_1}$ and $\|b_j\|_\infty \leq k_1 k_2$ for all j. We denote the set of admissible

multiindices by $\Phi(k_1, k_2)$. For $\mathbf{b} \in \Phi(k_1, k_2)$, define $z_{\mathbf{b},0}, z_{\mathbf{b},1}, \ldots, z_{\mathbf{b},k_1}$ in $[0,1]^d$ by $z_{\mathbf{b},0} = \mathbf{0}_d$ and $z_{\mathbf{b},j} = (b_j \Delta y, j\Delta x)$ for $j \geq 1$. Since $\mathbf{b}$ is admissible, $z_{\mathbf{b},0}, \ldots, z_{\mathbf{b},k_1}$ defines a chain in $[0,1]^d$. Define $\gamma_{\mathbf{b}} : [0,1] \to [0,1]^d$ to be the polygonal curve connecting the points $z_{\mathbf{b},0}, \ldots, z_{\mathbf{b},k_1}$, i.e.,

$$\gamma_{\mathbf{b}}(t) = z_{\mathbf{b},j-1} + \frac{1}{\Delta x}(z_{\mathbf{b},j} - z_{\mathbf{b},j-1})(t - (j-1)\Delta x)$$

for $t \in [(j-1)\Delta x, j\Delta x]$. For $\mathbf{b} \in \Phi(k_1, k_2)$ and $1 \leq j \leq k_1$, set

$$R_{\mathbf{b},j} = \left\{ x \in [0,1)^d \,:\, z_{\mathbf{b},j-1} - (\mathbf{1}_{d-1}, 0)\Delta y \leqq x < z_{\mathbf{b},j} \right\}.$$

For each rectangle $R_{\mathbf{b},j}$, we set $p_{\mathbf{b},j} = \int_{R_{\mathbf{b},j}} f(x)\, dx$. We say that a chain $x_1 \leqq x_2 \leqq \cdots \leqq x_m$ in $[0,1)^d$ is $\mathbf{b}$-*increasing* if

$$\{x_1, \ldots, x_m\} \subset \bigcup_{j=1}^{k_1} R_{\mathbf{b},j}.$$

It is not hard to see that for any k_1, k_2, every chain in $[0,1)^d$ is $\mathbf{b}$-increasing for some $\mathbf{b} \in \Phi(k_1, k_2)$. See Figure 3.4 for an illustration of the above definitions.

We first need a preliminary lemma which bounds the length of a longest chain within the narrow strip

$$T_j := [0,1]^{d-1} \times [(j-1)\Delta x, j\Delta x), \tag{3.88}$$

for any $j \in \{1, \ldots, k_1\}$. We note that the following lemma is a generalization of [31, Lemma 7]. The proof is based on the same idea of using a mixing process to embed $X_1, \ldots, X_n$ into another set of *i.i.d.* random variables that are uniform when restricted to the strip T_j.

Lemma III.21. *Let $f : [0,1)^d \to [0,\infty)$ be L-piecewise constant, and let $X_1, \ldots, X_n$ be* i.i.d. *with density f. Fix an integer $j \in \{1, \ldots, k_1\}$ and let $0 < \Delta x \leq \|f\|^{-1}_{L^\infty((0,1)^d)}$. Then*

$$\limsup_{n\to\infty}\; n^{-\frac{1}{d}}\, \ell\left(\{X_1, \ldots, X_n\} \cap T_j\right) \leq c_d \left(2\Delta x \|f\|_{L^\infty((0,1)^d)}\right)^{\frac{1}{d}} \quad \text{a.s.} \tag{3.89}$$

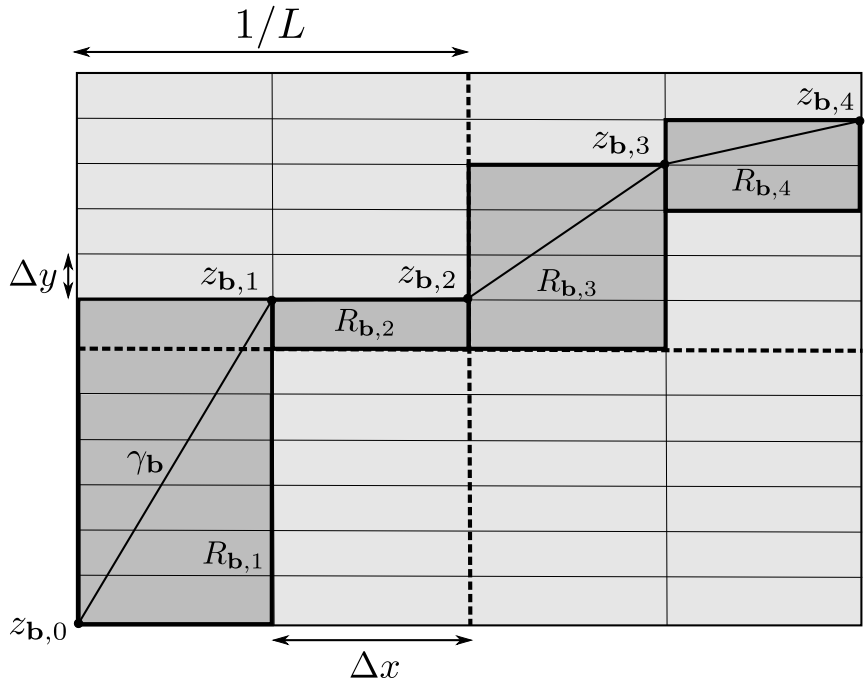


Figure 3.4: An illustration of the quantities $R_{\mathbf{b},j}$, $z_{\mathbf{b},j}$, $Q_{L,\alpha}$ and $\gamma_{\mathbf{b}}$ in two dimensions with $\mathbf{b} = (b_1, b_2, b_3, b_4) = (7, 7, 10, 11)$. In this case, the unit square is partitioned into four squares, $Q_{L,(1,1)}$, $Q_{L,(1,2)}$, $Q_{L,(2,1)}$ and $Q_{L,(2,2)}$, which are separated by dotted lines in the figure.

Proof. Set $M = \|f\|_{L^\infty((0,1)^d)}$ and let $g = f + (M - f) \cdot \chi_{T_j}$. Let $Y_1, \ldots, Y_n$ be *i.i.d.* according to the conditional density $\beta^{-1}(M-f)\cdot\chi_{T_j}$ where $\beta = \int_{T_j} M - f(x)\,dx$. Let $m_1, \ldots, m_n$ be Bernoulli zero-one random variables with parameter $(1+\beta)^{-1}$ and set

$$i_k = m_1 + \cdots + m_k. \tag{3.90}$$

Define $Z_1, \ldots, Z_n$ through the mixture process

$$Z_k = m_k X_{i_k} + (1 - m_k)Y_k.$$

Then $Z_1, \ldots, Z_n$ are *i.i.d.* with density $(1 + \beta)^{-1}g$. Let W denote the cardinality of the set $\{Z_1, \ldots, Z_n\} \cap T_j$. Then W is binomially distributed with parameters n and $p := (1 + \beta)^{-1}\Delta x M$. Since g is constant on T_j, we can use a similar argument to

that in Theorem III.20 to show that

$$n^{-\frac{1}{d}}\ell\left(\{Z_1,\ldots,Z_n\}\cap T_j\right)\to c_d p^{\frac{1}{d}} \quad \text{a.s.} \tag{3.91}$$

Let $m=i_n$ and note that

$$\ell\left(\{X_1,\ldots,X_m\}\right)=\ell\left(\{Z_k\,:\,m_k=1\}\cap T_j\right)\le \ell\left(\{Z_1,\ldots,Z_n\}\cap T_j\right),$$

and that $p\le \Delta x M$. Combining this with (3.91) we have

$$\limsup_{n\to\infty}\; n^{-\frac{1}{d}}\ell\left(\{X_1,\ldots,X_m\}\right)\le c_d(\Delta x M)^{\frac{1}{d}} \quad \text{a.s.} \tag{3.92}$$

Since m is Binomially distributed with parameters n and $(1+\beta)^{-1}$, we have $nm^{-1}\to 1+\beta$ almost surely and hence

$$\begin{aligned}\limsup_{n\to\infty}\; m^{-\frac{1}{d}}\ell\left(\{X_1,\ldots,X_m\}\right) &= \limsup_{n\to\infty}\,(nm^{-1})^{\frac{1}{d}}n^{-\frac{1}{d}}\ell\left(\{X_1,\ldots,X_m\}\right)\\ &\le (1+\beta)^{\frac{1}{d}}c_d(\Delta x M)^{\frac{1}{d}} \quad \text{a.s.}\end{aligned}$$

Since $\beta\le \Delta x M\le 1$ we have

$$\limsup_{n\to\infty}\; m^{-\frac{1}{d}}\ell\left(\{X_1,\ldots,X_m\}\right)\le c_d(2\Delta x M)^{\frac{1}{d}} \quad \text{a.s.}$$

The desired result (3.89) follows from noting that $n\mapsto m(n)$ is monotone nondecreasing along every sample path and $m\to\infty$ as $n\to\infty$ with probability one. □

The following short technical lemma is essential in the proof of Theorem III.23

Lemma III.22. *Let $f:[0,1)^d\to[0,\infty)$ be L-piecewise constant. For every $\varepsilon>0$ and $k_1\ge L$ we have*

$$\sum_{j\in\mathcal{H}_{\mathbf{b}}} p_{\mathbf{b},j}^{\frac{1}{d}}\le \overline{J}+\varepsilon. \tag{3.93}$$

for all $\mathbf{b}\in\Phi(k_1,k_2)$, the admissible multiindices, and $k_2\ge C\|f\|_{L^\infty((0,1)^d)}k_1^{d-1}/\varepsilon^d$, where

$$\mathcal{H}_{\mathbf{b}}=\{j\,:\,R_{\mathbf{b},j}\subset Q_{L,\alpha}\text{ for some }\alpha\}. \tag{3.94}$$

Proof. Let $k_1, k_2, \varepsilon > 0$, and $\mathbf{b} \in \Phi(k_1, k_2)$. Set $I_j = [(j-1)\Delta x, j\Delta x)$ and fix $j \in \{1, \dots, k_1\}$ and $t \in I_j$. Note that

$$|R_{\mathbf{b},j}| = \begin{cases} \Delta y^{d-1}\Delta x \prod_{i=1}^{d-1}(b_{j,i} - b_{j-1,i} + 1) & \text{if } j \geq 2 \\ \Delta y^{d-1}\Delta x \prod_{i=1}^{d-1}(b_{j,i} - b_{j-1,i}) & \text{if } j = 1, \end{cases}$$

and

$$\Delta x^d \gamma_{\mathbf{b},1}(t) \cdots \gamma_{\mathbf{b},d}(t) = \Delta y^{d-1}\Delta x \prod_{i=1}^{d-1}(b_{j,i} - b_{j-1,i}),$$

where we set $\mathbf{b}_0 = 0$ for convenience. A short computation shows that

$$|\Delta x(\gamma_{\mathbf{b},1}(t) \cdots \gamma_{\mathbf{b},d}(t))^{\frac{1}{d}} - |R_{\mathbf{b},j}|^{\frac{1}{d}}| \leq C\Delta x^{\frac{1}{d}}\Delta y^{\frac{1}{d}}, \tag{3.95}$$

where $C = (d-1)^{\frac{1}{d}}$. Since f is L-piecewise constant we have

$$f(\gamma_{\mathbf{b}}(t)) = f(z_{\mathbf{b},j-1}) = \frac{p_{\mathbf{b},j}}{|R_{\mathbf{b},j}|}, \tag{3.96}$$

for all $j \in \mathcal{H}_{\mathbf{b}}$ and $t \in I_j$. Noting that $\Delta x = |I_j|$ and recalling the definition of J (3.4) we have

$$\begin{aligned} \overline{J} \geq J(\gamma_{\mathbf{b}}) &\overset{(3.95)}{\geq} \sum_{j \in \mathcal{H}_{\mathbf{b}}} \frac{1}{|I_j|} \int_{I_j} f(\gamma_{\mathbf{b}}(t))^{\frac{1}{d}} (|R_{\mathbf{b},j}|^{\frac{1}{d}} - C\Delta x^{\frac{1}{d}}\Delta y^{\frac{1}{d}})\, dt \\ &= \sum_{j \in \mathcal{H}_{\mathbf{b}}} \frac{|R_{\mathbf{b},j}|^{\frac{1}{d}}}{|I_j|} \int_{I_j} f(\gamma_{\mathbf{b}}(t))^{\frac{1}{d}}\, dt - C \sum_{j \in H_{\mathbf{b}}} \frac{1}{|I_j|} \int_{I_j} f(\gamma_{\mathbf{b}}(t))^{\frac{1}{d}} \Delta x^{\frac{1}{d}}\Delta y^{\frac{1}{d}}\, dt \\ &\overset{(3.96)}{\geq} \sum_{j \in \mathcal{H}_{\mathbf{b}}} p_{\mathbf{b},j}^{\frac{1}{d}} - \|f\|_{L^\infty((0,1)^d)}^{\frac{1}{d}} k_1^{\frac{d-1}{d}} k_2^{-\frac{1}{d}}. \end{aligned} \tag{3.97}$$

Taking $k_2 \geq (C/\varepsilon)^d \|f\|_{L^\infty((0,1)^d)} k_1^{d-1}$ completes the proof. □

We now establish an asymptotic upper bound on $\ell(\{X_1, \dots, X_n\})$.

Theorem III.23. *Let $f : [0,1)^d \to [0, \infty)$ be L-piecewise constant, and let $X_1, \dots, X_n$ be* i.i.d. *with density f. Then*

$$\limsup_{n \to \infty} \; n^{-\frac{1}{d}} \ell(\{X_1, \dots, X_n\}) \leq c_d \overline{J} \quad \text{a.s.}$$

Proof. Let $k_2 > 0, k_1 \geq L, \varepsilon > 0$, and $\mathbf{b} \in \Phi(k_1, k_2)$. We suppose that $k_1 \geq L$ is a multiple of L so that f is k_1-piecewise constant. Let $\ell_{\mathbf{b}}(n)$ denote the length of a longest $\mathbf{b}$-increasing chain. Let n_j denote the number of $X_1, \ldots, X_n$ that belong $R_{\mathbf{b},j}$ and set

$$\ell_{\mathbf{b},j}(n) = \ell\left(\{X_1, \ldots, X_n\} \cap R_{\mathbf{b},j}\right). \tag{3.98}$$

Due to the monotonicity of $z_{\mathbf{b},0}, \ldots, z_{\mathbf{b},k_1}$, at most $(d-1)L$ of $R_{\mathbf{b},1}, \ldots, R_{\mathbf{b},k_1}$ can have a non-empty intersection with more than one hypercube $Q_{L,\alpha}$. It follows that $|\mathcal{H}_{\mathbf{b}}^c| \leq (d-1)L$, where $\mathcal{H}_{\mathbf{b}}^c = \{1, \ldots, k_1\} \setminus \mathcal{H}_{\mathbf{b}}$.

Since each $\mathbf{b}$-increasing chain is the union of chains in $R_{\mathbf{b},1}, \ldots, R_{\mathbf{b},k_1}$, we have

$$\ell_{\mathbf{b}}(n) \leq \sum_{j=1}^{k_1} \ell_{\mathbf{b},j}(n) = \sum_{j \in \mathcal{H}_{\mathbf{b}}^c} \ell_{\mathbf{b},j}(n) + \sum_{j \in \mathcal{H}_{\mathbf{b}}} \ell_{\mathbf{b},j}(n). \tag{3.99}$$

We will deal with each of the above sums separately. For the first term, set $M = \|f\|_{L^\infty((0,1)^d)}$ and let k_1 be large enough so that $\Delta x \leq 1/M$. Since $R_{\mathbf{b},j} \subset T_j$ for each j, we have by Lemma III.21 that

$$\limsup_{n\to\infty} \; n^{-\frac{1}{d}} \sum_{j \in \mathcal{H}_{\mathbf{b}}^c} \ell_{\mathbf{b},j}(n) \leq c_d |\mathcal{H}_{\mathbf{b}}^c| (2\Delta x M)^{\frac{1}{d}} \leq c_d (d-1) L (2M)^{\frac{1}{d}} k_1^{-\frac{1}{d}} \quad \text{a.s.}$$

Choose k_1 large enough so that

$$\limsup_{n\to\infty} \; n^{-\frac{1}{d}} \sum_{j \in \mathcal{H}_{\mathbf{b}}^c} \ell_{\mathbf{b},j}(n) \leq \frac{\varepsilon}{2} \quad \text{a.s.} \tag{3.100}$$

We now bound the second sum in (3.99). By Lemma III.22, choose $k_2 = k(M, k_1, \varepsilon)$ so that

$$\sum_{j \in \mathcal{H}_{\mathbf{b}}} p_{\mathbf{b},j}^{\frac{1}{d}} \leq \overline{J} + \frac{\varepsilon}{2c_d}, \tag{3.101}$$

for all $\mathbf{b} \in \Phi(k_1, k_2)$. For any $j \in \mathcal{H}_{\mathbf{b}}$, the conditional density ρ_j on $R_{\mathbf{b},j}$ is uniform. By a similar argument as in the proof of Theorem III.20, we have that

$$n^{-\frac{1}{d}} \ell_{\mathbf{b},j}(n) \to c_d p_{\mathbf{b},j}^{\frac{1}{d}} \quad \text{a.s.}$$

Combining this with (3.99), (3.100), and (3.101) we have

$$\limsup_{n\to\infty}\ n^{-\frac{1}{d}}\ell_{\mathbf{b}}(n) \le c_d\overline{J} + \varepsilon \quad \text{a.s.} \tag{3.102}$$

Since every chain in $X_1, \ldots, X_n$ is $\mathbf{b}$-increasing for some $\mathbf{b} \in \Phi(k_1, k_2)$, we have

$$\ell\left(\{X_1, \ldots, X_n\}\right) \le \max_{\mathbf{b}\in\Phi(k_1,k_2)} \ell_{\mathbf{b}}(n),$$

for every n. It follows that

$$\limsup_{n\to\infty}\ \ell\left(\{X_1, \ldots, X_n\}\right) \le c_d\overline{J} + \varepsilon \quad \text{a.s.},$$

which completes the proof. □

3.5.2 Continuous densities on Ω

We now generalize the convergence results on piecewise constant densities, Theorems III.20 and III.23, to continuous densities on Ω. Our main result, Theorem III.2, is proved at the end of the section. The idea of our approach is to divide $[0,1)^d$ into a large number of hypercubes, and to flatten f on each sub-cube. We can then apply the results from Section 3.5.1 and take the limit as the size of the sub-cubes tends to zero. In order to pass to the limit, we apply the perturbation result given in Lemma III.18.

Let $X_1, \ldots, X_n$ be *i.i.d.* with density f. We recall that $u_n(x)$ denotes the length of a longest chain among $X_1, \ldots, X_n$ consisting of points less than or equal to x under the partial order $\leqq$. In other words

$$u_n(x) = \ell\left(\{X_i \,:\, X_i \leqq x\}\right).$$

We also recall the definition of the value function U, defined in (3.5) by

$$U(x) = \sup_{\gamma\in\mathcal{A}\,:\,\gamma\leqq x} \int_0^1 f(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t)\cdots\gamma_d'(t))^{\frac{1}{d}}\, dt.$$

We now establish pointwise asymptotic upper and lower bounds on u_n.

Theorem III.24. *Let f satisfy (H0). Then for every $z \in \mathbb{R}^d$ we have*

$$\limsup_{n\to\infty} \; n^{-\frac{1}{d}} u_n(z) \le c_d U(z) \quad \text{a.s.} \tag{3.103}$$

Proof. Set $D = \{x \in \mathbb{R}^d \,:\, 0 \leqq x < z\}$ and $p = \int_D f(x)\,dx$. Suppose that $p = 0$. It follows from (H0) that f is lower semicontinuous, and hence $f(x) = 0$ for $x \leqq z$. Thus $u_n(z) = 0 = U(z)$ almost surely.

Suppose that $p > 0$ and let $\varepsilon > 0$. Let $k \in \mathbb{N}$ and partition D into k^d hypercubes $Q_{k,\alpha}$ for multiindices α with $\|\alpha\|_\infty \le k$. Define $f_k : \mathbb{R}^d \to [0,\infty)$ by

$$f_k(x) = \sum_\alpha \Big(\sup_{Q_{k,\alpha}} f\Big)\chi_{Q_{k,\alpha}}(x) + f(x)\chi_{\mathbb{R}^d\setminus D}(x), \tag{3.104}$$

and set $p_k = \int_D f_k(x)\,dx$. For every integer k, f_k is k-piecewise constant on D and $f \le f_k$. Define $v_k : \mathbb{R}^d \to \mathbb{R}$ by

$$v_k(x) = \sup_{\gamma\in\mathcal{A}\,:\,\gamma\leqq x} \int_0^1 f_k(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t)\cdots\gamma_d'(t))^{\frac{1}{d}}\,dt. \tag{3.105}$$

Note that the sequence f_k is uniformly bounded, Borel-measurable, and has compact support in $[0,1]^d$. Furthermore, it follows from (3.104) that (3.75) holds for the sequence f_k. Hence by Lemma III.18 we have that $v_k \to U$ uniformly as $k \to \infty$. Now fix k large enough so that

$$|v_k(z) - U(z)| \le \frac{\varepsilon}{c_d}. \tag{3.106}$$

Set

$$\lambda = \left(\int_{\mathbb{R}^d} f_k(x)\,dx\right)^{-1}, \tag{3.107}$$

and define $g = \lambda f_k$. Then $\lambda f \le g$ and we can write g as a convex combination of two distributions as follows:

$$g = \lambda f + (g - \lambda f).$$

Let $Y_1, \ldots, Y_n$ be *i.i.d.* with density $(1-\lambda)^{-1}(g - \lambda f)$, let $m_1, \ldots, m_n$ be Bernoulli random variables with parameter λ, and set

$$i_j = m_1 + \cdots + m_j.$$

Define

$$Z_j = m_j X_{i_j} + (1 - m_j) Y_j.$$

Then a simple computation shows that $Z_1, \ldots, Z_n$ are *i.i.d.* with density g. Let W denote the cardinality of $\{Z_1, \ldots, Z_n\} \cap D$. Since g is k-piecewise constant on D, we can apply Theorems III.20 and III.23 to obtain

$$\lim_{n\to\infty} W^{-\frac{1}{d}} \ell\left(\{Z_1, \ldots, Z_n\} \cap D\}\right) = c_d p_k^{-\frac{1}{d}} v_k(z) \quad \text{a.s.} \tag{3.108}$$

Note that W is Binomially distributed with parameters n and λp_k, hence $n^{-1}W \to \lambda p_k$ almost surely. Applying this to (3.108) we have

$$\begin{aligned} \lim_{n\to\infty} n^{-\frac{1}{d}} \ell\left(\{Z_1, \ldots, Z_n\} \cap D\right) &= \lim_{n\to\infty} (n^{-1}W)^{\frac{1}{d}} W^{-\frac{1}{d}} \ell\left(\{Z_1, \ldots, Z_n\} \cap D\right) \\ &= c_d \lambda^{\frac{1}{d}} v_k(z) \quad \text{a.s.} \end{aligned} \tag{3.109}$$

Set $m = i_n$. Note that m is Binomially distributed with parameters n and λ, and

$$u_m(z) = \ell\left(\{X_1, \ldots, X_m\} \cap D\right) \leq \ell\left(\{Z_1, \ldots, Z_n\} \cap D\right). \tag{3.110}$$

Combining (3.110) with (3.109) and the fact that $n^{-1}m \to \lambda$ as $n \to \infty$ we have

$$\begin{aligned} \limsup_{n\to\infty}\ m^{-\frac{1}{d}} u_m(z) &\overset{(3.110)}{\leq} \lim_{n\to\infty} (m^{-1}n)^{\frac{1}{d}} n^{-\frac{1}{d}} \ell\left(\{Z_1, \ldots, Z_n\} \cap D\right) \\ &\overset{(3.109)}{=} c_d v_k(z) \quad \text{a.s.} \end{aligned} \tag{3.111}$$

Recalling (3.106) we have

$$\limsup_{n\to\infty}\ m^{-\frac{1}{d}} u_m(z) \leq c_d U(z) + \varepsilon \quad \text{a.s.}$$

As in the proof of Lemma III.21, the proof is completed by noting that $n \mapsto m(n)$ is monotone nondecreasing along every sample path and $m \to \infty$ as $n \to \infty$ with probability one. □

Theorem III.25. *Let $f : \mathbb{R}^d \to \mathbb{R}$ satisfy (H0). Then for every $z \in \mathbb{R}^d$ we have*

$$\liminf_{n\to\infty} \; n^{-\frac{1}{d}} u_n(z) \geq c_d U(z) \quad \text{a.s.} \tag{3.112}$$

Proof. Let $\varepsilon > 0$. As in the proof of Theorem III.24, we set $D := \{x \in \mathbb{R}^d \,:\, 0 \leqq x < z\}$ and we may suppose that $p := \int_D f(x)\,dx > 0$. As before, let $k \in \mathbb{N}$ and partition D into k^d hypercubes $Q_{k,\alpha}$ for multiindices α with $\|\alpha\|_\infty \leq k$. Define $f_k : \mathbb{R}^d \to [0, \infty)$ by

$$f_k(x) = \sum_\alpha \Big(\inf_{Q_{k,\alpha}} f \Big) \chi_{Q_{k,\alpha}}(x) + f(x) \chi_{\mathbb{R}^d \setminus D}(x), \tag{3.113}$$

and set $p_k = \int_D f_k(x)\,dx$. Define

$$q(x) = \begin{cases} \frac{f_k(x)}{f(x)}, & \text{if } f(x) > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{3.114}$$

For any α such that $Q_{k,\alpha} \subset \Omega$, we have by (H0) and (3.113) that

$$f_k(x) \geq f(x) - m\left(\frac{\sqrt{d}}{k}\right) \text{ for } x \in Q_{k,\alpha}.$$

It follows that $\|q\|_{L^\infty(\mathbb{R}^d)} \nearrow 1$ as $k \to \infty$. As in the proof of Theorem III.24, we have that $v_k \to U$ uniformly as $k \to \infty$, where v_k is defined by (3.105). We can therefore fix k large enough so that

$$c_d \|q\|^{-\frac{1}{d}}_{L^\infty(\mathbb{R}^d)} v_k(z) \geq c_d U(z) - \varepsilon. \tag{3.115}$$

For $i = 1, \ldots, n$, let m_i be a Bernoulli zero-one random variable with parameter $\|q\|^{-1}_{L^\infty(\mathbb{R}^d)} q(X_i)$. Let $m = m_1 + \cdots + m_n$ and let $i_1, \ldots, i_m$ denote the indices for

which $m_i = 1$. We claim that $X_{i_1}, \ldots, X_{i_m}$ are *i.i.d.* with density $g := \lambda f_k$ where λ is defined in (3.107). To see this, first note since $f(x) = 0$ implies $f_k(x) = 0$, we have $q(x)f(x) = f_k(x)$ for all $x \in \mathbb{R}^d$. Thus

$$P(m_i = 1) = \int_{\mathbb{R}^d} P(m_i = 1 \,|\, X_i = x) f(x)\, dx = \int_{\mathbb{R}^d} \frac{q(x)}{\|q\|} f(x)\, dx = \frac{1}{\lambda \|q\|}, \tag{3.116}$$

where $\|q\| = \|q\|_{L^\infty(\mathbb{R}^d)}$. Let $j \geq 1$ and let $A \subset \mathbb{R}^d$ be measurable. We have

$$\begin{aligned} P(X_{i_j} \in A) &= P(X_i \in A \,|\, m_i = 1) \\ &= \frac{P(X_i \in A \text{ and } m_i = 1)}{P(m_i = 1)} \\ &\overset{(3.116)}{=} \lambda \|q\| \int_A P(m_i = 1 \,|\, X_i = x) f(x)\, dx \\ &= \lambda \|q\| \int_A \frac{q(x)}{\|q\|} f(x)\, dx \\ &= \int_A \lambda f_k(x)\, dx. \end{aligned}$$

By the construction of $X_{i_1}, \ldots, X_{i_m}$, they are independent random variables, hence the claim is established.

Let W denote the cardinality of $\{X_{i_1}, \ldots, X_{i_m}\} \cap D$. By Theorems III.20 and III.23, we have

$$\lim_{n \to \infty} W^{-\frac{1}{d}} \ell\left(\{X_{i_1}, \ldots, X_{i_m}\} \cap D\right) = c_d p_k^{-\frac{1}{d}} v_k(z) \quad \text{a.s.} \tag{3.117}$$

Define

$$w_i = \begin{cases} 1, & \text{if } m_i = 1 \text{ and } X_i \in D \\ 0, & \text{otherwise.} \end{cases}$$

Then $W = w_1 + \cdots w_n$. Each w_i is a Bernoulli zero-one random variable with parameter

$$P(w_i = 1) = P(m_i = 1 \text{ and } X_i \in D) = \int_D P(m_i = 1 \,|\, X_i = x) f(x)\, dx = \frac{p_k}{\|q\|}.$$

It follows that W is Binomially distributed with parameters n and $\|q\|^{-1}p_k$, and hence $n^{-1}W \to \|q\|^{-1}p_k$ almost surely. Combining this with (3.117) yields

$$\begin{aligned}\lim_{n\to\infty} n^{-\frac{1}{d}}\ell\left(\{X_{i_1},\dots,X_{i_m}\}\cap D\right) &= \lim_{n\to\infty}\left(n^{-1}W\right)^{\frac{1}{d}}W^{-\frac{1}{d}}\ell\left(\{X_{i_1},\dots,X_{i_m}\}\cap D\right)\\ &= c_d\|q\|^{-\frac{1}{d}}v_k(z) \quad \text{a.s.}\end{aligned} \tag{3.118}$$

Noting that

$$u_n(z) = \ell\left(\{X_1,\dots,X_n\}\cap D\right) \geq \ell\left(\{X_{i_1},\dots,X_{i_m}\}\cap D\right),$$

we have

$$\begin{aligned}\liminf_{n\to\infty}\ n^{-\frac{1}{d}}u_n(z) &\geq \lim_{n\to\infty} n^{-\frac{1}{d}}\ell\left(\{X_{i_1},\dots,X_{i_m}\}\cap D\right)\\ &\overset{(3.118)}{=} c_d\|q\|^{-\frac{1}{d}}v_k(z) \quad \text{a.s.}\end{aligned} \tag{3.119}$$

Recalling (3.115) we have

$$\liminf_{n\to\infty}\ n^{-\frac{1}{d}}u_n(z) \geq c_dU(z) - \varepsilon \quad \text{a.s.},$$

which completes the proof. □

We now have the proof of Theorem III.2.

Proof. Let $\varepsilon > 0$. Let $k \in \mathbb{N}$ and for a multiindex $\alpha \in \mathbb{Z}^d$, set $x_\alpha = \alpha/k$. Since U is uniformly continuous (by Lemma III.3) we can choose k large enough so that

$$|U(x_{\alpha+\mathbf{1}_d}) - U(x_\alpha)| \leq \frac{\varepsilon}{c_d} \tag{3.120}$$

for all $\alpha \in \mathbb{Z}^d$. Let I be the set of multiindices α for which $x_\alpha \in [0,1]^d$. Note that the cardinality of I is $(k+1)^d$. Since I is finite with cardinality independent of n, Theorems III.24 and III.25 yield

$$\lim_{n\to\infty}\sup_{\alpha\in I}|n^{-\frac{1}{d}}u_n(x_\alpha) - c_dU(x_\alpha)| = 0 \quad \text{a.s.} \tag{3.121}$$

Let $z \in (0,1]^d$. Then there exists $\alpha \in I$ such that $x_\alpha < z \leqq x_{\alpha+\mathbf{1}_d}$. By the Pareto-monotonicity of u_n and (3.120) we have

$$n^{-\frac{1}{d}} u_n(z) - c_d U(z) \leq n^{-\frac{1}{d}} u_n(x_{\alpha+\mathbf{1}_d}) - c_d U(z) \overset{(3.120)}{\leq} n^{-\frac{1}{d}} u_n(x_{\alpha+\mathbf{1}_d}) - c_d U(x_{\alpha+\mathbf{1}_d}) + \varepsilon.$$

By a similar argument, we have

$$n^{-\frac{1}{d}} u_n(z) - c_d U(z) \geq n^{-\frac{1}{d}} u_n(x_\alpha) - c_d U(x_\alpha) - \varepsilon,$$

and hence

$$\|n^{-\frac{1}{d}} u_n - c_d U\|_{L^\infty((0,1)^d)} \leq \sup_{\alpha \in I} |n^{-\frac{1}{d}} u_n(x_\alpha) - c_d U(x_\alpha)| + \varepsilon. \tag{3.122}$$

Combining (3.121) and (3.122) we have

$$\limsup_{n\to\infty} \|n^{-\frac{1}{d}} u_n - c_d U\|_{L^\infty((0,1)^d)} \leq \varepsilon \quad \text{a.s.},$$

and hence $\lim_{n\to\infty} \|n^{-\frac{1}{d}} u_n - c_d U\|_{L^\infty((0,1)^d)} = 0$ almost surely. The desired result now follows immediately from the boundary conditions on U proved in Theorem III.10 (i), (ii) and the fact that there are almost surely no samples in $\mathbb{R}^d \setminus (0,1)^d$. □

As a straightforward application of Theorem III.2, we can show that non-dominated sorting is stable under bounded random perturbations in the samples $X_1, \ldots, X_n$. For $\delta > 0$, we set

$$Z_i = X_i + Y_i \delta \quad \text{for} \quad i = 1, \ldots, n,$$

where $Y_1, \ldots, Y_n$ are *i.i.d.* with a continuous compactly supported density function $g : \mathbb{R}^d \to \mathbb{R}$. For $x \in \mathbb{R}^d$, set

$$u_n^\delta(x) = \ell\left(\{Z_i \,:\, Z_i \leqq x\}\right).$$

Theorem III.26 (Stability of non-dominated sorting)**.** *Let f satisfy (H0). There exist constants C_δ, depending only on δ, f, and g, such that*

$$\limsup_{n\to\infty} \; n^{-\frac{1}{d}} \|u_n^\delta - u_n\|_{L^\infty(\mathbb{R}^d)} \le C_\delta \quad \text{a.s.}$$

and $C_\delta \to 0$ as $\delta \to 0$.

Proof. Set $g_\delta(x) = \frac{1}{\delta^d} g\left(\frac{x}{\delta}\right)$. Then $Z_1, \dots, Z_n$ are *i.i.d.* with density $f_\delta := g_\delta * f$. Set

$$U^\delta(x) = \sup_{\gamma \in \mathcal{A} \,:\, \gamma \leqq x} \int_0^1 f_\delta(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}} \, dt.$$

Without loss of generality, we may suppose that $\overline{\Omega} \subset (0,1)^d$. Since $\operatorname{supp}(f) \subset \overline{\Omega}$ and g has compact support, we can take $\delta > 0$ small enough so that $\operatorname{supp}(f_\delta) \subset [0,1]^d$. It is not hard to see that (3.75) holds for the sequence f_δ. Since each f_δ is continuous and bounded with compact support in $[0,1]^d$, it follows from Lemma III.18 that $U^\delta \to U$ uniformly on $\mathbb{R}^d$. Note that

$$\begin{aligned} n^{-\frac{1}{d}} \|u_n^\delta - u_n\|_{L^\infty(\mathbb{R}^d)} \le \|n^{-\frac{1}{d}} u_n^\delta - c_d U^\delta\|_{L^\infty(\mathbb{R}^d)} + \|n^{-\frac{1}{d}} u_n - c_d U\|_{L^\infty(\mathbb{R}^d)} \\ + c_d \|U^\delta - U\|_{L^\infty(\mathbb{R}^d)}, \end{aligned}$$

for every n. Since f_δ is continuous on $(0,1)^d$ and $f_\delta(x) = 0$ for $x \notin (0,1)^d$, (H0) is satisfied for f_δ by taking $\Omega' = (0,1)^d$. We can therefore apply Theorem III.2 to obtain

$$\|n^{-\frac{1}{d}} u_n^\delta - c_d U^\delta\|_{L^\infty(\mathbb{R}^d)} + \|n^{-\frac{1}{d}} u_n - c_d U\|_{L^\infty(\mathbb{R}^d)} \to 0 \quad \text{a.s.}$$

The proof is completed by setting $C_\delta = c_d \|U^\delta - U\|_{L^\infty(\mathbb{R}^d)}$. □

CHAPTER IV

A PDE-based approach to non-dominated sorting

In this chapter we propose and study a fast numerical scheme for the Hamilton-Jacobi equation (P), and prove convergence of this scheme. We then show how the scheme can be used to design a fast approximate non-dominated sorting algorithm, and we evaluate the sorting accuracy of the new algorithm on both synthetic and real data. A fast approximate algorithm for non-dominated sorting has the potential to be a valuable tool for multiobjective optimization, especially in evolutionary algorithms which require frequent non-dominated sorting [29]. There are also potential applications in polynuclear growth of crystals in materials science [77], where the scheme could be used to simulate crystal growth in the presence of inhomogeneous growth rates.

We introduce the numerical scheme in Section 4.1 and prove convergence and regularity results for the scheme in Section 4.2. In Section 4.3 we demonstrate the numerical scheme on several density functions, and in Section 4.4 we propose a fast algorithm for approximate non-dominated sorting that is based on numerical solving (P).

4.1 Numerical scheme

Let us first fix some additional notation. Given $x, y \in \mathbb{R}^d$ with $x \leqq y$ we define the open and half-open intervals by

$$(x, y) = \{z \in \mathbb{R}^d : x < z < y\}, \quad (x, y] = \{z \in \mathbb{R}^d : x < z \leqq y\},$$

and make a similar definition for $[x, y)$. For any $x \in \mathbb{R}^d$ and $h > 0$, there exists unique $y \in h\mathbb{Z}^d$ and $z \in [0, h)^d$ such that $x = y + z$. We will denote y by $\lfloor x \rfloor_h$ so that $z = x - \lfloor x \rfloor_h$. We also denote $\mathbf{0} = (0, \ldots, 0) \in \mathbb{R}^d$ and $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^d$. For $z \in [\mathbf{0}, \infty)$, we recall that $\pi_z : \mathbb{R}^d \to [0, z]$ denotes the projection mapping $\mathbb{R}^d$ onto $[\mathbf{0}, z]$. For $x \in [\mathbf{0}, \infty)$ this mapping is given explicitly by

$$\pi_z(x) = (\min(x_1, z_1), \ldots, \min(x_d, z_d)).$$

In order to simplify notation in the chapter, let us consider the following Hamilton-Jacobi equation

$$\text{(P)} \quad \begin{cases} U_{x_1} \cdots U_{x_d} = f & \text{on } \mathbb{R}^d_+ \\ U = 0 & \text{on } \partial\mathbb{R}^d_+. \end{cases}$$

Notice we have simply removed the d^{-d} factor from the right hand side, compared to (P) in Chapter III. With this new definition of (P), Theorem III.2 gives that $n^{-\frac{1}{d}} u_n \to d^{-1} c_d U$ uniformly with probability one.

We now present a numerical scheme for (P). For a given $x \in [\mathbf{0}, \infty)$, the domain of dependence for (P) is $\{y : y \leqq x\}$. This can be seen from the connection to non-dominated sorting and the longest chain problem. It is thus natural to consider a scheme for (P) based on backward difference quotients, yielding

$$\prod_{i=1}^{d} (U_h(x) - U_h(x - he_i)) = h^d f(x), \tag{4.1}$$

where $U_h : h\mathbb{N}_0^d \to \mathbb{R}$ is the numerical solution of (P) and $e_1, \ldots, e_d$ are the standard basis vectors in $\mathbb{R}^d$. Under reasonable hypotheses on f, described in Section 4.2.2, there exists a unique Pareto-monotone viscosity solution of (P). As we wish to numerically approximate this Pareto-monotone solution we may assume that $U_h(x) \geq U_h(x - he_i)$ for all i. Given that f is non-negative, for any $f(x), U_h(x - e_1), \ldots, U_h(x - e_d)$, there is a unique $U_h(x)$ with

$$U_h(x) \geq \max(U_h(x - he_1), \ldots, U_h(x - he_d)),$$

satisfying (4.1). Hence the numerical solution U_h can be computed by visiting each grid point exactly once via any sweeping pattern that respects the partial order $\leqq$. The scheme therefore has linear complexity in the number of gridpoints. At each grid point, the scheme (4.1) can be solved numerically by either a binary search and/or Newton's method restricted to the interval

$$[\max(U_h(x - he_1), \ldots, U_h(x - he_d)), \max(U_h(x - he_1), \ldots, U_h(x - he_d)) + hf(x)^{1/d}].$$

In the case of $d = 2$, we can solve the scheme (4.1) explicitly via the quadratic formula

$$U_h(x) = \frac{1}{2}(U_h(x - he_1) + U_h(x - he_2)) + \frac{1}{2}\sqrt{(U_h(x - he_1) - U_h(x - he_2))^2 + 4h^2 f(x)}.$$

Now extend U_h to a function $U_h : [\mathbf{0}, \infty) \to \mathbb{R}$ by setting $U_h(x) = U_h(\lfloor x \rfloor_h)$. Defining

$$\Gamma_h = [\mathbf{0}, \infty) \setminus (h\mathbf{1}, \infty),$$

we see that U_h is a Pareto-monotone solution of the discrete scheme

$$\text{(S)} \quad \begin{cases} S(h, x, U_h) = f(\lfloor x \rfloor_h), & \text{if } x \in (h\mathbf{1}, \infty) \\ \\ U_h(x) = 0, & \text{if } x \in \Gamma_h, \end{cases}$$

where $S : \mathbb{R}_+ \times (h\mathbf{1}, \infty) \times X \to \mathbb{R}$ is defined by

$$S(h, x, u) = \prod_{i=1}^{d} \frac{u(x) - u(x - he_i)}{h}. \tag{4.2}$$

Here, X is the space of functions $u : [\mathbf{0}, \infty) \to \mathbb{R}$. In the next section we will study properties of solutions U_h of (S).

4.2 Convergence of numerical scheme

In this section we prove that the numerical solutions U_h defined by (S) converge uniformly to the viscosity solution of (P). A general framework for proving convergence of a finite-difference scheme to the viscosity solution of a non-linear second order PDE was developed by Barles and Souganidis [12]. Their framework requires that the scheme be stable, monotone, consistent, and that the PDE satisfy a *strong uniqueness property* [12]. The monotonicity condition is equivalent to ellipticity for second order equations, and plays a similar role for first order equations, enabling one to prove maximum and/or comparison principles for the discrete scheme. The strong uniqueness property refers to a comparison principle that holds for semicontinuous viscosity sub- and supersolutions.

The numerical scheme (S) is easily seen to be consistent; this simply means that

$$\lim_{\substack{y \to x \\ h \to 0}} S(h, y, \varphi) = \varphi_{x_1}(x) \cdots \varphi_{x_d}(x),$$

for all $\varphi \in C^1(\mathbb{R}^d_+)$. The scheme is stable [12] if the numerical solutions U_h are uniformly bounded in L^∞, independent of h. It is not immediately obvious that (S) is stable; stability follows from the discrete comparison principle for (S) (Lemma IV.1) and is proved in Lemma IV.3. The monotonicity property requires the following:

$$S(h, x, u) \leq S(h, x, v) \text{ whenever } u \geq v \text{ and } u(x) = v(x).$$

It is straightforward to verify that (S) is monotone when restricted to Pareto-monotone u, v. This is sufficient since we are only interested in the Pareto-monotone viscosity solution of (P). All that is left is to establish a strong uniqueness result for (P). Unfortunately such a result is not available under the hypothesis (H0) from Chapter III. Since f may be discontinuous along $\partial\Omega$, we can only establish a comparison principle for continuous viscosity sub- and supersolutions.

One way to rectify this situation is to break the proof into two steps. First prove convergence of the numerical scheme for f Lipschitz on $\mathbb{R}^d_+$. It is straightforward in this case to establish a strong uniqueness result for (P). Second, extend the result to f satisfying (H0) by an approximation argument using inf and sup convolutions. Although this approach is fruitful, we take an alternative approach as it yields an interesting regularity property for the numerical solutions. In particular, in Lemma IV.3 we establish approximate Hölder regularity of U_h of the form

$$|U_h(x) - U_h(y)| \leq C(|x-y|^{\frac{1}{d}} + h^{\frac{1}{d}}). \tag{4.3}$$

As we verify in Appendix A, the approximate Hölder estimate (4.3) along with the stability of (S) allows us to apply the Arzelà-Ascoli Theorem, with a slightly modified proof, to the sequence U_h. This allows us to substitute the ordinary uniqueness result given by Theorem III.19 in place of strong uniqueness.

4.2.1 Analysis of the numerical scheme

We first prove a discrete comparison principle for the scheme (S). This comparison principle is essential in proving stability of (S) and the approximate Hölder regularity result in Lemma IV.3. For the remainder of this section, we fix $h > 0$.

Lemma IV.1 (Comparison principle)**.** *Let $z \in (h\mathbf{1}, \infty)$ and suppose $u, v \in L^\infty_{loc}([\mathbf{0}, \infty))$*

are Pareto-monotone and satisfy

$$S(h,x,u) \le S(h,x,v) \quad \textit{for all } x \in (h\mathbf{1}, z]. \tag{4.4}$$

Then $u \le v$ *on* $\Gamma_h \cap [\mathbf{0}, z]$ *implies that* $u \le v$ *on* $[\mathbf{0}, z]$.

Proof. Suppose that $\sup_{[\mathbf{0},z]}(u-v) > 0$ and set

$$T_r = \left\{ x \in [\mathbf{0}, \infty) \,:\, \frac{1}{d}(x_1 + \cdots + x_d) \le r \right\},$$

and

$$R = \sup\{r > 0 \,:\, u \le v \text{ on } T_r \cap [\mathbf{0}, z]\}.$$

Since $u \le v$ on $\Gamma_h \cap [\mathbf{0}, z)$ and $\sup_{[\mathbf{0},z]}(u-v) > 0$, we must have $R \in [h, s]$, where $s = d^{-1}(z_1 + \cdots + z_d)$. By the definition of R, there exists $x \in (h\mathbf{1}, z]$ and $s < R$ such that

$$u(x) > v(x) \text{ and } x - he_i \in T_s \text{ for } i = 1, \ldots, d.$$

Since $s < R$, we have $u \le v$ on $T_s \cap [\mathbf{0}, z]$ and hence

$$u(x - he_i) \le v(x - he_i) \le v(x) \quad \text{for } i = 1, \ldots, d. \tag{4.5}$$

The second inequality above follows from Pareto-monotonicity of v. Since u and v are Pareto-monotone and $u(x) > v(x)$ we have

$$\prod_{i=1}^{d}(u(x) - u(x - he_i)) > \prod_{i=1}^{d}(v(x) - u(x - he_i)) \overset{(4.5)}{\ge} \prod_{i=1}^{d}(v(x) - v(x - he_i)).$$

Hence $S(h,x,u) > S(h,x,v)$, contradicting the hypothesis. □

Using the comparison principle, we can establish that numerical solutions of (S) satisfy an additional boundary condition at infinity.

Proposition IV.2. *Let $u \in L^\infty_{loc}([\mathbf{0},\infty))$ be Pareto-monotone with $u = 0$ on Γ_h. Suppose that for some $z \in (h\mathbf{1},\infty)$ we have*

$$\operatorname{supp}\{x \mapsto S(h,x,u)\} \subset [\mathbf{0},z]. \tag{4.6}$$

Then we have $u = u \circ \pi_z$.

Proof. Define $v = u \circ \pi_z$ and fix $x \in [\mathbf{0},\infty)$. Since u is Pareto-monotone and $\pi_z(x) \leqq x$, we have $v(x) = u(\pi_z(x)) \leq u(x)$. Hence $v \leq u$. Since $u = v$ on $[\mathbf{0},z]$ we have

$$S(h,x,u) = S(h,x,v) \quad \text{for all } x \in [\mathbf{0},z] \setminus \Gamma_h.$$

For $x \notin [\mathbf{0},z] \cup \Gamma_h$ we have $S(h,x,u) = 0$ by assumption. Since v is Pareto-monotone we have $S(h,x,v) \geq 0 = S(h,x,u)$ for such x, and hence $S(h,x,v) \geq S(h,x,u)$ for all $x \in [\mathbf{0},\infty) \setminus \Gamma_h$. Since $v = u = 0$ on Γ_h we can apply Lemma IV.1 to find that $u \leq v$ on $[\mathbf{0},\infty)$, and hence $u = v = u \circ \pi_z$. □

An important consequence of the comparison principle is the following approximate Hölder regularity result.

Lemma IV.3. *Let $u \in L^\infty_{loc}([\mathbf{0},\infty))$ be Pareto-monotone with $u = 0$ on Γ_h. Then for any $R > 0$ we have*

$$|u(x) - u(y)| \leq 2d^2 R^{\frac{d-1}{d}} \|S(h,\cdot,u)\|^{\frac{1}{d}}_{L^\infty((h,R]^d)} (|x-y|^{\frac{1}{d}} + h^{\frac{1}{d}}) \tag{4.7}$$

for all $x, y \in (h,R]^d$.

Proof. Let $R > 0$ and $x_0, y_0 \in (h,R]^d$. We first deal with the case where $x_0 \leqq y_0$. Set $\widehat{u}(x) = u(\pi_{x_0}(x))$ and define $\psi : \mathbb{R}^d \to \mathbb{R}$ by

$$\psi(x) = \begin{cases} d(x_1 \cdots x_d)^{\frac{1}{d}} & \text{if } x \in (\mathbf{0},\infty), \\ 0 & \text{otherwise.} \end{cases} \tag{4.8}$$

By the concavity of $t \mapsto t^{\frac{1}{d}}$ we have

$$\psi(x) - \psi(x - he_i) = d(x_1 \cdots x_d)^{\frac{1}{d}} x_i^{-\frac{1}{d}} (x_i^{\frac{1}{d}} - (x_i - h)^{\frac{1}{d}}) \geq x_i^{-1} (x_1 \cdots x_d)^{\frac{1}{d}} h,$$

for any $x \in (h\mathbf{1}, \infty)$ and hence

$$S(h, x, \psi) \geq 1 \quad \text{for all} \quad x \in (h\mathbf{1}, \infty). \tag{4.9}$$

By the translation invariance of S and (4.9) we have

$$S(h, x, \psi(\cdot - b)) \geq 1 \quad \text{for all } b \in [\mathbf{0}, \infty), \ x \in (b + h\mathbf{1}, \infty). \tag{4.10}$$

Set $b^i = (x_{0,i} - h)e_i \in \mathbb{R}^d$. For $x \in [\mathbf{0}, \infty)$ set

$$w(x) = \widehat{u}(x) + \|S(h, \cdot, u)\|^{\frac{1}{d}}_{L^\infty((h,R]^d)} \sum_{i=1}^{d} \psi(x - b^i),$$

and note that w is Pareto-monotone. Let $x \in (h\mathbf{1}, \infty) \setminus (h\mathbf{1}, x_0]$. Then for some k we have $x_k > x_{0,k}$, and hence $x > b^k + h\mathbf{1}$. We therefore have

$$\begin{aligned}
S(h, x, w) &\geq \frac{1}{h^d} \prod_{i=1}^{d} \Big(\widehat{u}(x) - \widehat{u}(x - he_i) \\
&\qquad + \|S(h, \cdot, u)\|^{\frac{1}{d}}_{L^\infty((h,R]^d)} (\psi(x - b^k) - \psi(x - b^k - he_i)) \Big) \\
&\geq S(h, x, \widehat{u}) + \|S(h, \cdot, u)\|_{L^\infty((h,R]^d)} S(h, x, \psi(\cdot - b^k)) \\
&\overset{(4.10)}{\geq} S(h, x, \widehat{u}) + \|S(h, \cdot, u)\|_{L^\infty((h,R]^d)} \\
&\geq S(h, x, u).
\end{aligned}$$

Suppose now that $x \in (h\mathbf{1}, x_0]$. Then since $u = \widehat{u}$ on $[\mathbf{0}, x_0]$ we have $S(h, x, \widehat{u}) = S(h, x, u)$ and hence $S(h, x, w) \geq S(h, x, u)$. Since $w \geq u = 0$ on $\Gamma_h \cap [0, R]^d$, we can

apply Lemma IV.1 to obtain $w \geq u$ on $[0,R]^d$, which yields

$$\begin{aligned} u(y_0) - \widehat{u}(y_0) &\leq \|S(h,\cdot,u)\|^{\frac{1}{d}}_{L^\infty((h,R]^d)} \sum_{i=1}^{d} \psi(y_0 - b^i) \\ &\leq dR^{\frac{d-1}{d}} \|S(h,\cdot,u)\|^{\frac{1}{d}}_{L^\infty((h,R]^d)} \sum_{i=1}^{d} (y_{0,i} - x_{0,i} + h)^{\frac{1}{d}} \\ &\leq d^2 R^{\frac{d-1}{d}} \|S(h,\cdot,u)\|^{\frac{1}{d}}_{L^\infty((h,R]^d)} (|x_0 - y_0|^{\frac{1}{d}} + h^{\frac{1}{d}}). \end{aligned} \tag{4.11}$$

Noting that $\pi_{x_0}(y_0) = x_0$ we have $\widehat{u}(y_0) = u(\pi_{x_0}(y_0)) = u(x_0)$, which completes the proof for the case that $x_0 \leqq y_0$.

Suppose now that $x_0, y_0 \in (h,R]^d$ such that $x_0 \not\leqq y_0$. Set

$$x = \pi_{x_0}(y_0) = \pi_{y_0}(x_0).$$

Then $|x_0 - x| \leq |x_0 - y_0|$, $|y_0 - x| \leq |x_0 - y_0|$, $x \leqq x_0$, and $x \leqq y_0$. It follows that

$$\begin{aligned} |u(x_0) - u(y_0)| &\leq |u(x_0) - u(x)| + |u(y_0) - u(x)| \\ &\leq 2d^2 R^{\frac{d-1}{d}} \|S(h,\cdot,u)\|^{\frac{1}{d}}_{L^\infty((h,R]^d)} (|x_0 - y_0|^{\frac{1}{d}} + h^{\frac{1}{d}}), \end{aligned}$$

which completes the proof. □

4.2.2 Main convergence result

Our main result is the following convergence statement for the scheme (S). Let us first recall the assumption (H0) on f from Chapter III:

(H0) There exists an open and bounded set $\Omega \subset \mathbb{R}^d_+$ with Lipschitz boundary such that f is non-negative and uniformly continuous on Ω and $f = 0$ on $\mathbb{R}^d \setminus \Omega$.

Theorem IV.4. *Let f be nonnegative and satisfy (H0). Let U be the unique Pareto-monotone viscosity solution of (P). For every $h > 0$ let $U_h : [\mathbf{0}, \infty) \to \mathbb{R}$ be the unique Pareto-monotone solution of (S). Then $U_h \to U$ uniformly on $[\mathbf{0}, \infty)$ as $h \to 0$.*

Proof. By (H0) we have that $f(x) = 0$ for $x \notin (0,1)^d$, and hence $\text{supp}(f(\lfloor \cdot \rfloor_h)) \subset [0,1]^d$. Therefore, by Proposition IV.2 and Lemma IV.3 we have

$$\|U_h\|_{L^\infty([\mathbf{0},\infty))} \leq C\|f\|^{\frac{1}{d}}_{L^\infty([\mathbf{0},\infty))}, \tag{4.12}$$

for all $h > 0$. Similarly, combining Proposition IV.2 with Lemma IV.3 we have

$$|U_h(x) - U_h(y)| \leq 2d^2\|f\|^{\frac{1}{d}}_{L^\infty([\mathbf{0},\infty))}(|x-y|^{\frac{1}{d}} + h^{\frac{1}{d}}) \text{ for all } x, y \in [\mathbf{0},\infty), \tag{4.13}$$

for every $h > 0$. The estimates in (4.12) and (4.13) show uniform boundedness, and a type of equicontinuity, respectively, for the sequence U_h. By an argument similar to the proof of the Arzelà-Ascoli Theorem (see the Appendix), there exists a subsequence $h_k \to 0$ and $u \in C^{0,\frac{1}{d}}([\mathbf{0},\infty))$ such that $U_{h_k} \to u$ uniformly on compact sets in $[\mathbf{0},\infty)$. By Proposition IV.2, we actually have $U_{h_k} \to u$ uniformly on $[\mathbf{0},\infty)$. Since the scheme (S) is monotone and consistent, it is a standard result that u is a viscosity solution of (P) [12]. Note that U_h is Pareto-monotone and $U_h = 0$ on Γ_h. Since $U_{h_k} \to u$ uniformly, it follows that u is Pareto-monotone and $u = 0$ on $\partial\mathbb{R}^d_+$. By Theorem III.19 we have $u = U$. Since we can apply the same argument to any subsequence of U_h, it follows that $U_h \to U$ uniformly on $[\mathbf{0},\infty)$. □

In Section 4.3, we observe that the numerical scheme provides a fairly consistent underestimate of the exact solution of (P). The following lemma shows that this is indeed the case whenever the solution U of (P) is concave.

Lemma IV.5. *Let f be nonnegative and satisfy (H0). Let U be the unique Pareto-monotone viscosity solution of (P). For every $h > 0$ let $U_h : [\mathbf{0},\infty) \to \mathbb{R}$ be the unique Pareto-monotone solution of (S). If U is concave on $[\mathbf{0},\infty)$ then $U_h \leq U$ for every $h > 0$.*

Proof. Fix $h > 0$. Since U is concave, it is differentiable almost everywhere.[1] Let $x \in (h\mathbf{1}, \infty)$ be a point at which U is differentiable and f is continuous. Since U is concave we have

$$U(x) - U(x - he_i) \geq hU_{x_i}(x) \text{ for all } i.$$

Since U is a viscosity solution of (P) and f is continuous at x we have

$$S(h, x, U) \geq U_{x_1}(x) \cdots U_{x_d}(x) = f(x).$$

Since $x \mapsto S(h, x, U)$ is continuous, we see that $S(h, x, U) \geq f_*(x) = f(x)$ for all $x \in (h\mathbf{1}, \infty]$. Now define $W_h(x) = U(\lfloor x \rfloor_h)$. Then we have

$$S(h, x, W_h) \geq f(\lfloor x \rfloor_h) \text{ for all } x \in (h\mathbf{1}, x],$$

and $W_h = 0$ on Γ_h. It follows from Lemma IV.1 that $U_h \leq W_h$. Since U is Pareto-monotone, we have $W_h \leq U$, which completes the proof. □

4.3 Numerical Results

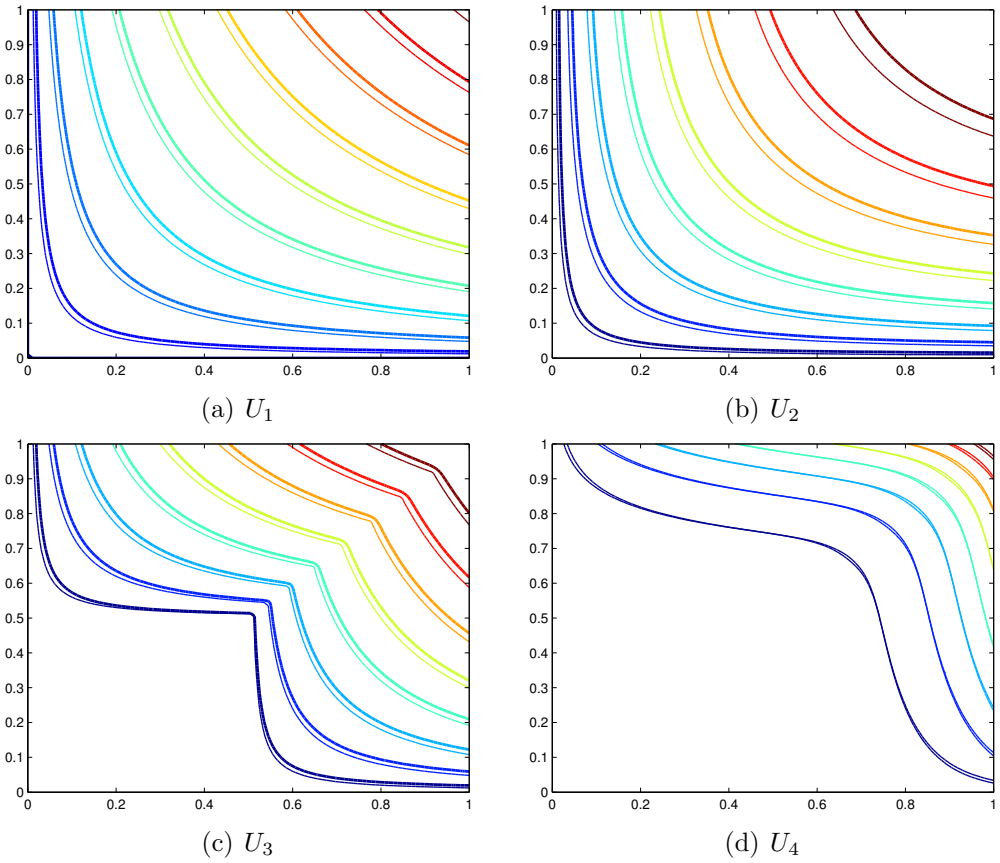
We now present some numerical results using the scheme (S) to approximate the viscosity solution of (P). We consider four special cases where the exact solution of (P) can be expressed in analytical form. Let $f_1(x) = 1$, $f_2(x) = \frac{2^d}{\pi^{d/2}} e^{-|x|^2}$,

$$f_3(x) = 1 - \chi_{[0,1/2]^d}(x) \quad \text{and} \quad f_4(x) = \left(\sum_{i=1}^d x_i^9 \right)^{1-d} \prod_{i=1}^d \left(9x_i^9 + \sum_{i=1}^d x_i^9 \right).$$

Here, χ_A denotes the characteristic function of the set A. The corresponding solutions of (P) are $U_1(x) = d(x_1 \cdots x_d)^{\frac{1}{d}}$, $U_2(x) = d\left(\prod_{i=1}^d \operatorname{erf}(x_i) \right)^{\frac{1}{d}}$, and

$$U_3(x) = d \max_{i \in \{1,\ldots,d\}} \left\{ \left(x_i - \frac{1}{2} \right)_+ \prod_{j \neq i} x_j \right\}^{\frac{1}{d}}, \quad U_4(x) = d \left(\prod_{i=1}^d x_i \cdot \sum_{i=1}^d x_i^9 \right)^{\frac{1}{d}},$$

[1] The fact that U is Pareto-monotone also implies differentiability almost everywhere.



(a) U_1 (b) U_2

(c) U_3 (d) U_4

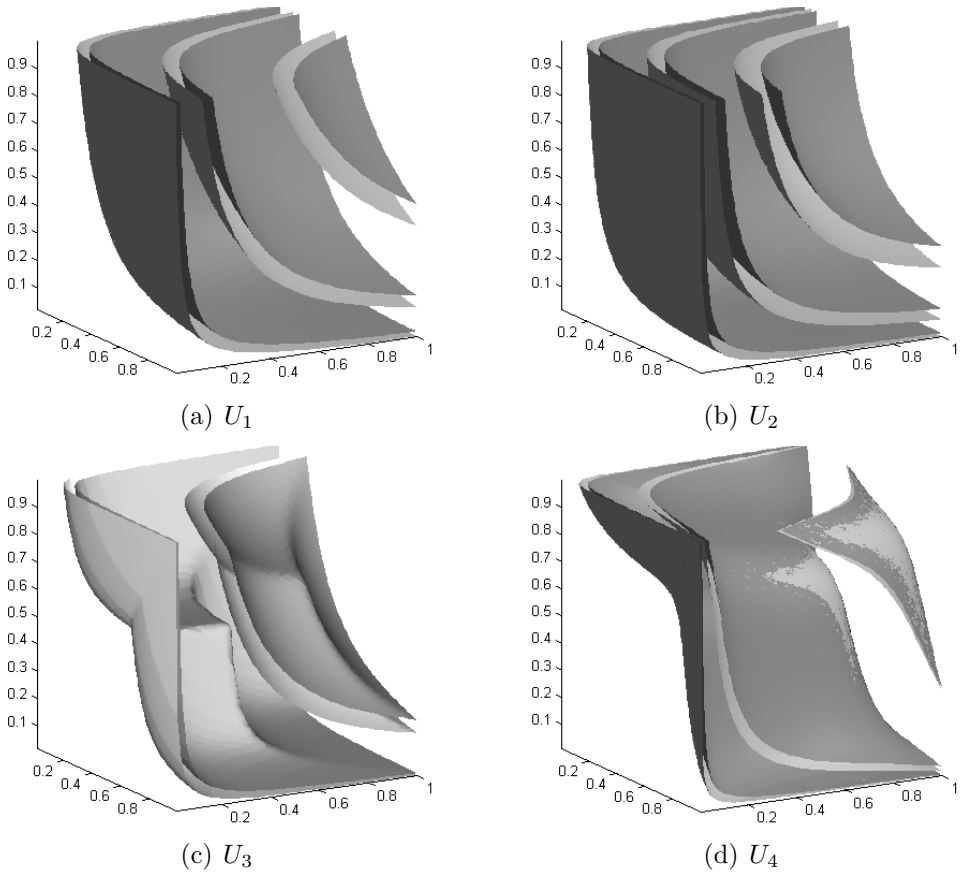
Figure 4.1: Comparison of numerical solutions and exact solutions of (P) for $d = 2$. The thin and thick lines represent the level sets of the exact and numerical solutions, respectively.

where $\mathrm{erf}\,(x)$ is the error function defined by $\mathrm{erf}\,(x) = 2/\sqrt{\pi}\int_0^x e^{-t^2}\,dt$, and $x_+ := \max(0, x)$. The solutions U_1 and U_2 are special cases of the formula

$$U(x) = d\left(\int_{[\mathbf{0},x]} f(y)\,dy\right)^{\frac{1}{d}}, \tag{4.14}$$

which holds when f is separable, i.e., $f(x) = f_1(x_1)\cdots f_d(x_d)$. The solution U_3 can be obtained by the method of characteristics. We chose to evaluate the proposed numerical scheme for U_4 because it has non-convex level sets, and then computed f_4 via (P). In the probabilistic interpretation of (P) as the continuum limit of non-dominated sorting, non-convex Pareto fronts play an important role [32].

We computed the numerical solutions for $d = 2$ and $d = 3$. For $d = 2$ we used a



(a) U_1 (b) U_2

(c) U_3 (d) U_4

Figure 4.2: Comparison of numerical solutions and exact solutions of (P) for $d = 3$. The light and dark surfaces represent the level sets of the exact and numerical solutions, respectively.

100×100 grid, and for $d = 3$, we used a $50 \times 50 \times 50$ grid and solved the scheme at each grid point via a binary search with precision $\varepsilon = 10^{-4}$. Figures 4.1 and 4.2 compare the level sets of the exact solutions to those of the numerical solutions for $d = 2$ and $d = 3$, respectively. In Figure 4.1, the thin lines correspond to the exact solution while the thick lines correspond to the numerical solutions, with the exception of 4.1(d) where both are thin lines for increased visibility. In Figure 4.2, the darker surfaces correspond to the numerical solution while the lighter surfaces represent the exact solution. For both $d = 2$ and $d = 3$, we can see that the level sets of the numerical solutions consistently overestimate the true solution, indicating

that the numerical solutions are converging from below to the exact solutions. We proved in Lemma IV.5 that $U_h \nearrow U$ whenever U is concave, so this observation is to be expected. Note however, that U_3 is not convex, yet the overestimation is still present, indicating that Lemma IV.5 may hold under more general hypotheses on U. We also observe that U_3 has a shock, which is resolved reasonably well for $d = 2$ and $d = 3$, given the grid sizes used.

4.3.1 Rate of convergence

We show here the results of some numerical experiments concerning the rate of convergence of $U_h \to U$ and $n^{-\frac{1}{d}} u_n \to c_d d^{-1} U$. Figure 4.3(a) shows $\|U_h - U\|_{L^1([0,1]^2)}$ and $\|U_h - U\|_{L^\infty(\mathbb{R}^d_+)}$ versus h for the density $f_3(x) = 1 - \chi_{[0,1/2]^d}(x)$ from the beginning of Section 4.3. Both norms appear to have convergence rates on the order of $O(h^\alpha)$, and a regression analysis yields $\alpha = 0.5006$ for the L^∞ norm and $\alpha = 0.8787$ for the L^1 norm. Thus, it is reasonable to suspect an L^∞ convergence rate of the form

$$\|U_h - U\|_{L^\infty(\mathbb{R}^d_+)} \leq C h^{\frac{1}{d}}, \tag{4.15}$$

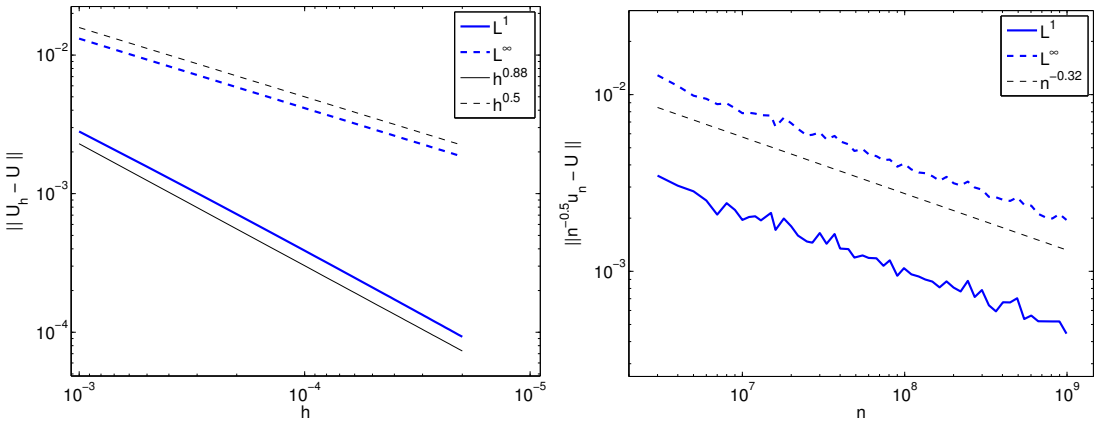
for some constant $C > 0$. It is quite natural that the convergence rate for the L^1 norm is substantially better than the L^∞ norm, due to the non-differentiability of U_3 at the boundary $\partial\mathbb{R}^2_+$. This induces a large error near $\partial\mathbb{R}^2_+$ which has a more significant impact on the L^∞ norm.

To measure the rate of convergence of $n^{-\frac{1}{d}} u_n \to c_d d^{-1} U$, we consider the following two norms

$$|n^{-\frac{1}{d}} u_n - c_d d^{-1} U|_{L^\infty} := \max_{1 \leq i \leq n} |n^{-\frac{1}{d}} u_n(X_i) - c_d d^{-1} U(X_i)| \tag{4.16}$$

and

$$|n^{-\frac{1}{d}} u_n - c_d d^{-1} U|_{L^1} := \frac{1}{n} \sum_{i=1}^{n} |n^{-\frac{1}{d}} u_n(X_i) - c_d d^{-1} U(X_i)| \tag{4.17}$$



(a) Convergence rate of scheme (S) (b) Stochastic convergence rates

Figure 4.3: Convergence rates for (a) the scheme (S) as a function of the grid resolution h, and (b) the stochastic convergence $n^{-\frac{1}{d}}u_n \to c_d d^{-1} U$ as a function of the number n of random samples.

Figure 4.3(b) shows (4.16) and (4.17) versus n for the same density f_3. For each n the values of (4.16) and (4.17) were computed by taking the average over 10 independent realizations. It appears that both norms decay on the order of $O(n^{-\alpha})$, and a regression analysis yields $\alpha = 0.3281$ for the L^1 norm (4.17) and $\alpha = 0.3144$ for the L^∞ norm (4.16). These results are in line with the known convergence rates for the longest chain problem with a uniform distribution on $[0,1]^d$ [16].

The results for the other densities f_1, f_2, and f_4 are similar. We demonstrated the convergence rates on f_3 due to the fact that it has many important features; namely, it is discontinuous, yields non-convex Pareto-fronts, and induces a shock in the viscosity solution U_3 of (P).

4.4 Fast approximate non-dominated sorting

We demonstrate now how the numerical scheme (S) can be used for fast approximate non-dominated sorting, and give a real-world application to anomaly detection in Section 4.4.4. We assume here that the given data $X_1, \ldots, X_n$ are drawn *i.i.d.* from

a reasonably smooth density function f, and that n is large enough so that $n^{-\frac{1}{d}}u_n$ is well approximated by $c_d d^{-1} U$. In this regime, it is reasonable to consider an approximate non-dominated sorting algorithm based on numerically solving (P). A natural algorithm is as follows.

Since the density f is rarely known in practice, the first step is to form an estimate $\widehat{f}$ of f using the samples $X_1, \ldots, X_n$. In the large sample regime, this can be done very accurately using, for example, a kernel density estimator [91] or a k-nearest neighbor estimator [64]. To keep the algorithm as simple as possible, we opt for a simple histogram to estimate f, aligned with the same grid used for numerically solving (P). When n is large, the estimation of f can be done with only a random subset of $X_1, \ldots, X_n$ of cardinality $k \ll n$, which avoids considering all n samples. The second step is to use the numerical scheme (S) to solve (P) on a fixed grid of size h, using the estimated density $\widehat{f}$ on the right hand side of (P). This yields an estimate $\widehat{U}_h$ of U, and the final step is to evaluate $\widehat{U}_h$ at each sample $X_1, \ldots, X_n$ to yield approximate Pareto ranks for each point. The final evaluation step can be viewed as an interpolation; we know the values of $\widehat{U}_h$ on each grid point and wish to evaluate $\widehat{U}_h$ at an arbitrary point. A simple linear interpolation is sufficient for this step. However, in the spirit of utilizing the PDE (P), we solve the scheme (S) at each point $X_1, \ldots, X_n$ using the values of $\widehat{U}_h$ at neighboring grid points, i.e., given $\widehat{U}_h(x - he_i)$ for all i, and $y \in [x - h\mathbf{1}, x]$, we compute $\widehat{U}_h(y)$ by solving

$$\prod_{i=1}^{d} (\widehat{U}_h(y) - \widehat{U}_h(y - h_i e_i)) = h_1 \cdots h_d \widehat{f}(x), \tag{4.18}$$

where $h_i = y_i - (x_i - h)$. In (4.18) we compute $\widehat{U}_h(y - h_i e_i)$ by linear interpolation using adjacent grid points. Figure 4.4 illustrates the grid used for computing $\widehat{U}_h(y)$.

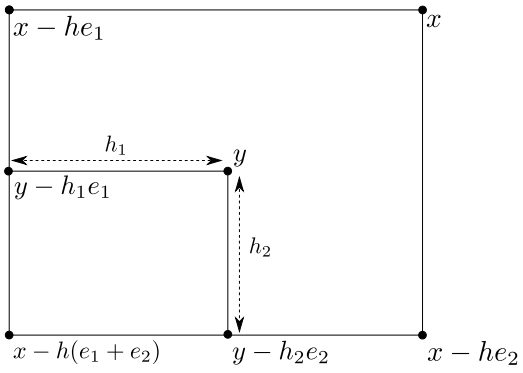


Figure 4.4: Depiction of the grid used for computing $\widehat{U}_h(y)$ according to (4.18). The values of $\widehat{U}_h(y - h_1 e_1)$ and $\widehat{U}_h(y - h_2 e_2)$ are computed by linear interpolation using adjacent grid points, i.e., $\widehat{U}_h(y - h_1 e_1)$ is computed via linearly interpolating between $\widehat{U}_h(x - he_1)$ and $\widehat{U}_h(x - h(e_1 + e_2))$.

The entire algorithm is summarized in Algorithm IV.1. For simplicity of dis-

---

Algorithm IV.1: Fast approximate non-dominated sorting

---

1. Select k points from $X_1, \ldots, X_n$ at random. Call them $Y_1, \ldots, Y_k$.
2. Select a grid spacing h for solving the PDE and estimate f with a histogram aligned to the grid $h\mathbb{N}_0^d$, i.e.,

$$\widehat{f}_h(x) = \frac{1}{kh^d} \cdot \#\left\{ Y_i \,:\, x \leqq Y_i \leqq x + h\mathbf{1} \right\} \text{ for } x \in h\mathbb{N}_0^d. \tag{4.19}$$

3. Compute the numerical solution $\widehat{U}_h$ on $h\mathbb{N}_0^d \cap [0,1]^d$ via (S).
4. Evaluate $\widehat{U}_h(X_i)$ for $i = 1, \ldots, n$ via interpolation.

---

cussion, we have assumed that $X_1, \ldots, X_n$ are drawn from $[0,1]^d$, but this is not essential as the scheme (S) can be easily adapted to any hypercube in $\mathbb{R}^d$, and this is in fact what we do in our implementation of Algorithm IV.1.

4.4.1 Convergence rates in Algorithm IV.1

It is important to understand how the parameters k and h in Algorithm IV.1 affect the accuracy of the estimate $\widehat{U}_h$. We first consider the estimate $\widehat{f}_h$. By (4.19),

we can write

$$h^d \widehat{f}_h(x) = \frac{1}{k} \sum_{i=1}^{k} \chi_{[x,x+h\mathbf{1}]}(Y_i).$$

Hence $h^d \widehat{f}_h(x)$ is the average of *i.i.d.* Bernoulli random variables with parameter

$$p = \int_{[x,x+h\mathbf{1}]} f(y)\, dy. \tag{4.20}$$

By the central limit theorem, the fluctuations of $\widehat{f}_h(x)$ about its mean satisfy

$$\left| \widehat{f}_h(x) - \frac{p}{h^d} \right| \leq C \frac{1}{\sqrt{k} h^d}, \tag{4.21}$$

with high probability.

Let us suppose now that f is globally Lipschitz. The following can be easily modified for f more or less regular, yielding similar results. Then by (4.20) we have

$$\left| f(x) - \frac{p}{h^d} \right| \leq C\sqrt{d} h.$$

Combining this with (4.21) we have

$$\| \widehat{f}_h - f \|_{L^\infty([0,1]^d \cap h\mathbb{N}^d)} \leq C \left(\frac{1}{\sqrt{k} h^d} + h \right), \tag{4.22}$$

with high probability. By the discrete comparison principle (Lemma IV.1) and (4.22) we have that

$$\| \widehat{U}_h - U_h \|_{L^\infty([0,1]^d)} \leq d \| \widehat{f}_h - f \|^{\frac{1}{d}}_{L^\infty([0,1]^d \cap h\mathbb{N}^d)} \leq C \left(k^{-\frac{1}{2d}} h^{-1} + h^{\frac{1}{d}} \right), \tag{4.23}$$

with high probability. Based on the numerical evidence presented in Section 4.3.1, it is reasonable to suspect that $\| U - U_h \|_{L^\infty([0,1]^d)} \leq C h^{\frac{1}{d}}$. If this is indeed the case, then in light of (4.23) we have

$$\| \widehat{U}_h - U \|_{L^\infty([0,1]^d)} \leq C \left(k^{-\frac{1}{2d}} h^{-1} + h^{\frac{1}{d}} \right), \tag{4.24}$$

with high probability.

The right side of the inequality (4.24) is composed of two competing additive terms. The first term $Ck^{-\frac{1}{2d}}h^{-1}$ captures the effect of random errors (variance) due to an insufficient number k of samples. The second term $Ch^{\frac{1}{d}}$ captures the effect of non-random errors (bias) due to insufficient resolution h of the proposed numerical scheme (S). This decomposition into random and non-random errors is analogous to the mean integrated squared error decomposition in the theory of non-parametric regression and image reconstruction [59]. Similarly to [59] we can use the bound in (4.24) to obtain rules of thumb on how to choose k and h. For example, we may first choose some value for k, and then choose h so as to equate the two competing terms in (4.24). This yields $h = k^{-\frac{1}{2(d+1)}}$ and (4.24) becomes

$$\|\widehat{U}_h - U\|_{L^\infty([0,1]^d)} \leq Ck^{-\frac{1}{2d(d+1)}} = Ch^{\frac{1}{d}}, \tag{4.25}$$

with high probability.

Notice that Steps 1-3 in Algorithm IV.1, i.e., computing $\widehat{U}_h$, require $O(kh^{-d})$ operations. If we choose the equalizing value $h = k^{-\frac{1}{2(d+1)}}$, then we find that computing $\widehat{U}_h$ has complexity $O\left(k^{\frac{3d+2}{2d+2}}\right)$. Thus Algorithm IV.1 is sublinear in the following sense. Given $\varepsilon > 0$, we can choose k large enough so that

$$\|\widehat{U}_h - U\|_{L^\infty([0,1]^d)} \leq \frac{\varepsilon}{2c_d},$$

with high probability. The L^1 sorting accuracy of using $\widehat{U}_h$ in place of u_n is then given by

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} |c_d\widehat{U}_h(X_i) - dn^{-\frac{1}{d}}u_n(X_i)| &\leq \frac{1}{n}\sum_{i=1}^{n}\Big(c_d|\widehat{U}_h(X_i) - U(X_i)| \\ &\qquad + |c_dU(X_i) - dn^{-\frac{1}{d}}u_n(X_i)|\Big) \\ &\leq \frac{\varepsilon}{2} + \frac{1}{n}\sum_{i=1}^{n}|c_dU(X_i) - dn^{-\frac{1}{d}}u_n(X_i)|,\end{aligned}$$

with high probability. By the stochastic convergence $dn^{-\frac{1}{d}}u_n \to c_d U$, and the rates presented in Section 4.3.1, there exists $N > 0$ such that for all $n \geq N$ we have

$$\frac{1}{n}\sum_{i=1}^{n} |c_d \widehat{U}_h(X_i) - dn^{-\frac{1}{d}}u_n(X_i)| \leq \varepsilon \tag{4.26}$$

with high probability. Thus, for any $\varepsilon > 0$ there exists N, k and h such that $\widehat{U}_h$ is an $O(\varepsilon)$ approximation of u_n for all $n \geq N$, and $\widehat{U}_h$ can be computed in constant time with respect to n. We emphasize that the sublinear nature of the algorithm lies in the computation of $\widehat{U}_h$. Ranking all samples, i.e., evaluating $\widehat{U}_h$ at each of $X_1, \ldots, X_n$, and computing the L^1 error in (4.26) of course requires $O(n)$ operations. In practice, it is often the case that one need not rank all n samples (e.g., in a streaming application [44]), and in such cases the entire algorithm is constant or sublinear in n in the sense described above.

4.4.2 Evaluation of Algorithm IV.1

We evaluated our proposed algorithm in dimension $d = 2$ for a uniform density and a mixture of Gaussians given by $f(x) = \frac{1}{4}\sum_{i=1}^{4} g_i(x)$, where each $g_i : \mathbb{R}^2 \to \mathbb{R}$ is a multivariate Gaussian density with covariance matrix Σ_i and mean μ_i. We write the covariance matrix in the form $\Sigma_i = R_{\theta_i}\text{diag}(\lambda_{i,1}, \lambda_{i,2})R_{\theta_i}^T$, where R_θ denotes a rotation matrix, and $\lambda_{i,1}$, $\lambda_{i,2}$ are the eigenvalues. The values for $\lambda_{i,j}, \mu_i$ and θ_i are given in Table 4.1, and the density is illustrated in Figure 4.5.

It is important to evaluate the accuracy of the approximate sorting obtained by Algorithm IV.1. In practice, the numerical ranks assigned to each point are largely irrelevant, provided the relative orderings between samples are correct. Hence a natural accuracy measure for a given ranking is the fraction of pairs (X_i, X_j) that are ordered correctly. Recalling that the true Pareto rank is given by $u_n(X_i)$, this

	$\lambda_{i,1}$	$\lambda_{i,2}$	θ_i	$(\mu_{i,1},\mu_{i,2})$
g_1	0.01	0.00025	$\frac{\pi}{3}$	(0.2,0.5)
g_2	0.0576	0.00064	0	(0.5,0.3)
g_3	0.04	0.00025	$-\frac{\pi}{6}$	(0.4,0.8)
g_4	0.01	0.01	0	(0.8,0.8)

Table 4.1: Parameter values for mixture of Gaussians density

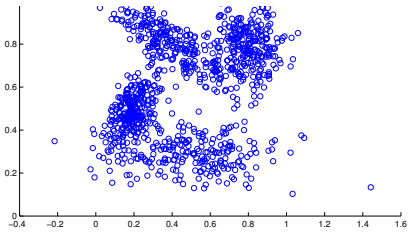


Figure 4.5: Depiction of random samples from the mixture of Gaussians density.

can be expressed as

$$\text{Accuracy} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \psi(u_n(X_i) - u_n(X_j), \widehat{U}_h(X_i) - \widehat{U}_h(X_j)), \tag{4.27}$$

where $\psi(x,y) = 1$ if $xy > 0$ and $\psi(x,y) = 0$ otherwise. It turns out that the accuracy scores (4.27) for our algorithm are often very close to 1. In order to make the plots easier to interpret visually, we have chosen to plot $-\log(1 - \text{Accuracy})$ instead of Accuracy in *all* plots.

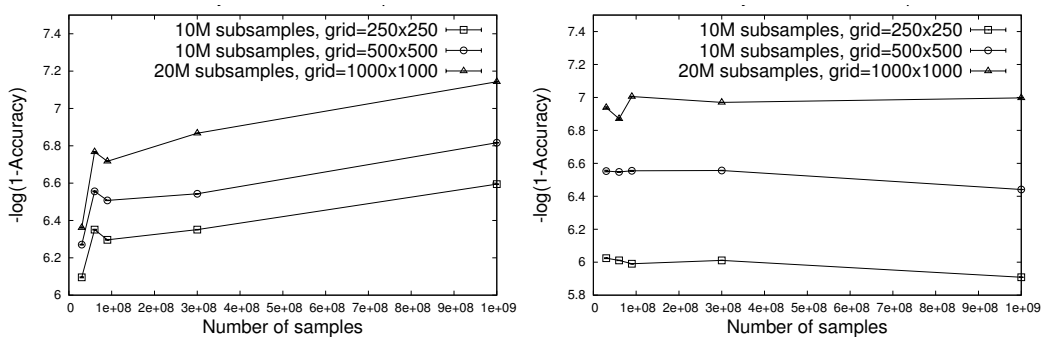
Unfortunately, the complexity of computing the accuracy score via (4.27) is $O(n^2)$, which is intractable for even moderate values of n. We note however that (4.27) is, at least formally, a Monte-Carlo approximation of

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \psi(U(x) - U(y), U_h(x) - U_h(y)) f(x) f(y)\, dxdy.$$

Hence it is natural to use a truncated Monte-Carlo approximation to estimate (4.27). This is done by selecting n pairs $(X_{i_1}, X_{j_1}), \ldots, (X_{i_n}, X_{j_n})$ at random and computing

$$\frac{1}{n} \sum_{k=1}^{n} \psi(u_n(X_{i_k}) - u_n(X_{j_k}), \widehat{U}_h(X_{i_k}) - \widehat{U}_h(X_{j_k})).$$

The complexity of the Monte-Carlo approximation is $O(n)$. In all plots in the paper, we computed the Monte-Carlo approximation 10 times and plotted means and error bars corresponding to a 95% confidence interval. In all of the figures, the confidence intervals are sufficiently small so that they are contained within the data point itself.



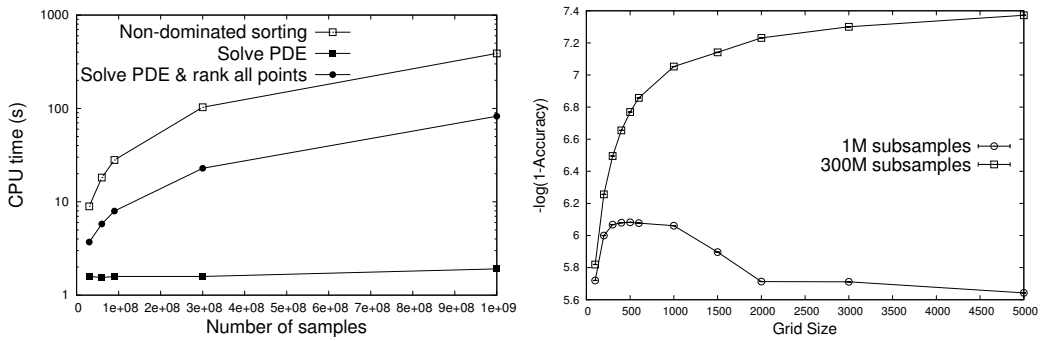
(a) Uniform distribution (b) Mixture of Gaussians

Figure 4.6: Comparison of accuracy versus number of samples for various grid sizes and number of subsamples k used to estimate f.

We can see in Figure 4.6 that we can achieve excellent accuracy while maintaining a fixed grid and subsample size as a function of n. We also see that, as expected, the accuracy increases when one uses more grid points for solving the PDE and/or more subsamples for estimating the density. We also see that the algorithm works better on uniformly distributed samples than on the mixture of Gaussians. Indeed, it is quite natural to expect the density estimation and numerical scheme to be less accurate when f changes rapidly.

We compared the performance of our algorithm against the fast two dimensional non-dominated sorting algorithm presented in [34], which takes $O(n \log n)$ operations to sort n points. The code for both algorithms was written in C++ and was compiled on the same architecture with the same compiler optimization flags. Figure 4.7(a) shows a comparison of the CPU time used by each algorithm. For our fast approximate sorting, we show the CPU time required to solve the PDE (Steps 1-3 in Algorithm IV.1) separately from the CPU time required to execute all of Algorithm IV.1, since the former is sublinear in n.

It is also interesting to consider the relationship between the grid size and the number of subsamples k. In Figure 4.7(b), we show accuracy versus grid size for



(a) CPU time (b) Accuracy vs grid size

Figure 4.7: (a) Comparison of CPU time versus number of samples for a grid size of 250×250 and $k = 10^7$ subsamples for estimating the density. (b) Comparison of accuracy versus grid size for $k = 10^6$ and $k = 3 \times 10^8$ subsamples for non-dominated sorting of $n = 3 \times 10^8$ points. Notice that when k is small compared to n it is not always beneficial to use a finer grid for solving the PDE and estimating the density.

$k = 10^6$ and $k = 3 \times 10^8$ subsamples for non-dominated sorting of $n = 3 \times 10^8$ points. Notice that for $k = 10^6$ subsamples, it is not beneficial to use a finer grid than approximately 500×500. This is quite natural in light of the error estimate on Algorithm IV.1 (4.24).

4.4.3 Subset ranking

There are certainly other ways one may think of to perform fast approximate sorting without invoking the PDE (P). One natural idea would be to perform non-dominated sorting on a random subset of $X_1, \ldots, X_n$, and then rank all n points via some form of interpolation. We will call such an algorithm *subset ranking* (in contrast to the PDE-based ranking we have proposed). Although such an approach is quite intuitive, it is important to note that there is, at present, no theoretical justification for such an approach. Nonetheless, it is important to compare the performance of our algorithm against such an algorithm.

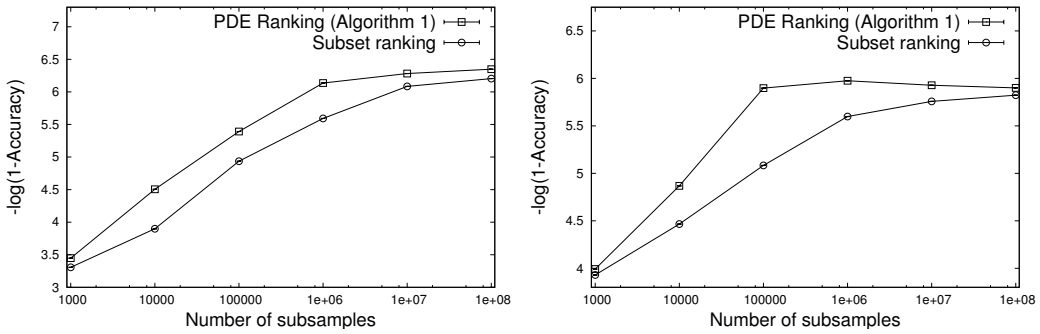
Let us describe how one might implement a subset ranking algorithm. As described above, the first step is to select a random subset of size k from $X_1, \ldots, X_n$.

Let us call the subset $Y_1, \ldots, Y_k$. We then apply non-dominated sorting to $Y_1, \ldots, Y_k$, which generates Pareto rankings $u_k(Y_i)$ for each Y_i. The final step is to rank $X_1, \ldots, X_n$ via interpolation. There are many ways one might approach this. In similar spirit to our PDE-based ranking (Algorithm IV.1), we use grid interpolation, using the same grid size as used to solve the PDE. We compute a ranking at each grid point by averaging the ranks of all samples from $Y_1, \ldots, Y_k$ that fall inside the corresponding grid cell. The ranking of an arbitrary sample X_i is then computed by linear interpolation using the ranks of neighboring grid points. In this way, the rank of X_i is an average of the ranks of nearby samples from $Y_1, \ldots, Y_k$, and there is a grid size parameter which allows a meaningful comparison with PDE-based ranking (Algorithm IV.1).

Figure 4.8 shows the accuracy scores for PDE-based ranking (Algorithm IV.1) and subset ranking of $n = 10^8$ samples drawn from the uniform and mixture of Gaussians distributions. A grid size of 250×250 was used for both algorithms, and we varied the number of subsamples from $k = 10^3$ to $k = 10^8$. Notice a consistent accuracy improvement when using PDE-based ranking versus subset ranking, when the number of subsamples is significantly less than n. It is somewhat surprising to note that subset ranking has much better than expected performance. As mentioned previously, to our knowledge there is no theoretical justification for such a performance when k is small.

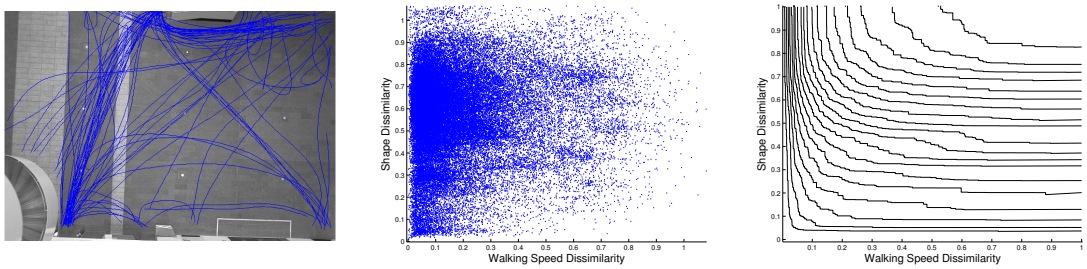
4.4.4 Application in anomaly detection

We now demonstrate Algorithm IV.1 on a large scale real data application of anomaly detection [51]. The data consists of thousands of pedestrian trajectories, captured from an overhead camera, and the goal is to differentiate nominal from



(a) Uniform distribution (b) Mixture of Gaussians

Figure 4.8: Comparison of PDE-based ranking (Algorithm IV.1) and naive subset interpolation ranking for sorting $n = 10^8$ samples. Accuracy scores are shown for various numbers of subsamples ranging from $k = 10^3$ to $k = 10^8$.



(a) Example trajectories (b) 50000 Pareto points (c) Pareto fronts

Figure 4.9: (a) Example pedestrian trajectories, (b) Plot of 50000 of the approximately 6×10^9 Pareto points, (c) 30 evenly spaced Pareto fronts computed from the 50000 points in (b).

anomalous pedestrian behavior in an unsupervised setting. The data is part of the Edinburgh Informatics Forum Pedestrian Database and was captured in the main building of the School of Informatics at the University of Edinburgh [68]. Figure 4.9(a) shows 100 of the over 100,000 trajectories captured from the overhead camera.

The approach to anomaly detection employed in [51] utilizes multiple criteria to measure the dissimilarity between trajectories, and combines the information using a Pareto-front method, and in particular, non-dominated sorting. The database consists of a collection of trajectories $\{\gamma_1, \ldots, \gamma_M\}$, where $M = 110035$, and the criteria used in [51] are a walking speed dissimilarity, and a trajectory shape dis-

similarity. Given two trajectories $\gamma_i, \gamma_j : [0,1] \to [0,1]^2$, the walking speed dissimilarity $c_{speed}(\gamma_i, \gamma_j)$ is the L^2 distance between velocity histograms of each trajectory, and the trajectory shape dissimilarity is the L^2 distance between the trajectories themselves, i.e., $c_{shape}(\gamma_i, \gamma_j) = \|\gamma_i - \gamma_j\|_{L^2(0,1)}$. There is then a Pareto point $X_{i,j} = (c_{speed}(\gamma_i, \gamma_j), c_{shape}(\gamma_i, \gamma_j))$ for every pair of trajectories (γ_i, γ_j), yielding $\binom{M}{2} \approx 6 \times 10^9$ Pareto points. Figure 4.9(b) shows an example of 50000 Pareto points and Figure 4.9(c) shows the respective Pareto fronts. In [51], only 1666 trajectories from one day were used, due to the computational complexity of computing the dissimilarities and non-dominated sorting.

The anomaly detection algorithm from [51] performs non-dominated sorting on the Pareto points $\{X_{i,j}\}_{1 \le i < j \le M}$, and uses this sorting to define an anomaly score for every trajectory γ_i. Let $n = \binom{M}{2}$ and let $u_n : \mathbb{R}^2 \to \mathbb{R}$ denote the longest chain function corresponding to this non-dominated sorting. The anomaly score for a particular trajectory γ_i is defined as

$$s_i = \frac{1}{M} \sum_{j=1}^{M} u_n(c_{speed}(\gamma_i, \gamma_j), c_{shape}(\gamma_i, \gamma_j)),$$

and trajectories with an anomaly score higher than a predefined threshold σ are deemed anomalous.

Using Algorithm IV.1, we can approximate u_n using only a small fraction of the Pareto points $\{X_{i,j}\}_{1 \le 1 < j \le M}$, thus alleviating the computational burden of computing all pairwise dissimilarities. Figure 4.10 shows the accuracy scores for Algorithm IV.1 and subset ranking versus the number of subsamples k used in each algorithm. Due to the memory requirements for non-dominated sorting, we cannot sort datasets significantly larger than than 10^9 points. Although there is no such limitation on Algorithm IV.1, it is important to have a ground truth sorting to compare against.

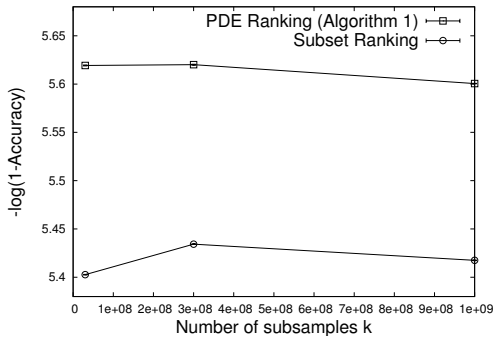


Figure 4.10: Accuracy scores for Algorithm IV.1 and subset ranking for sorting 10^9 Pareto points from the pedestrian anomaly detection problem versus the number of subsamples k.

Therefore we have used only 44722 out of 110035 trajectories, yielding approximately 10^9 Pareto points. For both algorithms, a 500×500 grid was used for solving the PDE and interpolation. Notice the accuracy scores are similar to those obtained for the test data in Figure 4.6. This is an intriguing observation in light of the fact that $\{X_{i,j}\}_{1 \leq i < j \leq M}$ are *not i.i.d.*, since they are elements of a Euclidean dissimilarity matrix.

4.5 Discussion

We have provided theory that demonstrates that, when $X_1, \ldots, X_n$ are *i.i.d.* in $\mathbb{R}^2$ with a nicely behaved density function f, the numerical scheme (S) for (P) can be utilized to perform fast approximate non-dominated sorting with a high degree of accuracy. We have also shown that in a real world example with non-*i.i.d.* data, the scheme (S) still obtains excellent sorting accuracy. We expect the same algorithm to be useful in dimensions $d = 3$ and $d = 4$ as well, but of course the complexity of solving (P) on a grid increases exponentially fast in d. In higher dimensions, one could explore other numerical techniques for solving (P) which do not utilize a fixed grid [23]. At present, there is also no good algorithm for non-dominated sorting in

high dimensions. The fastest known algorithm is $O(n(\log n)^{d-1})$ [57], which becomes intractable when n and d are large.

This algorithm has the potential to be particularly useful in the context of big data streaming problems [44], where it would be important to be able to construct an approximation of the Pareto depth function u_n without visiting all the datapoints $X_1, \ldots, X_n$, as they may be arriving in a data stream and it may be impossible to keep a history of all samples. In such a setting, one could slightly modify Algorithm IV.1 so that upon receiving a new sample, the estimate $\widehat{f}_h$ is updated, and every so often the scheme (S) is applied to recompute the estimate of $\widehat{U}_h$.

There are certainly many situations in practice where the samples $X_1, \ldots, X_n$ are not *i.i.d.*, or the density f is not nicely behaved. In these cases, there is no reason to expect our algorithm to have much success, and hence we make no claim of universal applicability. However, there are many cases of practical interest where these assumptions are valid, and hence this algorithm can be used to perform fast non-dominated sorting in these cases. Furthermore, as we have demonstrated in Section 4.4.4, there are situations in practice where the *i.i.d.* assumption is violated, yet our proposed algorithm maintains excellent accuracy and performance.

We proposed a simple *subset ranking* algorithm based on sorting a small subset of size k and then performing interpolation to rank all n samples. Although there is currently no theoretical basis for such an algorithm, we showed that subset ranking achieves surprisingly high accuracy scores and is only narrowly outperformed by our proposed PDE-based ranking. The simplicity of subset ranking makes it particularly appealing, but more research is needed to prove that it will always achieve such high accuracy scores for moderate values of k.

We should note that there are many obvious ways to improve our algorithm.

Histogram approximation to probability densities is quite literally the most basic density estimation algorithm, and one would expect to obtain better results with more sophisticated estimators. It would also be natural to perform some sort of histogram equalization to $X_1, \ldots, X_n$ prior to applying our algorithm in order to spread the samples out more uniformly and smooth out the effective density f. Provided such a transformation preserves the partial order $\leqq$ it would not affect the non-dominated sorting of $X_1, \ldots, X_n$. In the case that f is separable (a product density), one can perform histogram equalization on each coordinate independently to obtain uniformly distributed samples. We leave these and other potential improvements to future work; our purpose in this paper has been to demonstrate that one can obtain excellent results with a very basic algorithm.

Appendix

We use the following minor extension of the Arzelà-Ascoli Theorem in Section 4.2.2. It has recently come to our attention that a similar result is used in stochastic games to prove the convergence of value functions (see [69, Lemma 4.2]).

Let X be a compact metric space. We say that a sequence $\{f_n\}_{n=1}^{\infty}$ of real-valued functions on X is *approximately equicontinuous* if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\forall x, y \in X,\ |x - y| < \delta \implies |f_n(x) - f_n(y)| < \varepsilon + \frac{1}{n}, \tag{4.28}$$

for every $n \in \mathbb{N}$.

Theorem IV.6. *Let $\{f_n\}_{n=1}^{\infty}$ be approximately equicontinuous and uniformly bounded. Then there exists a subsequence of $\{f_n\}_{n=1}^{\infty}$ converging uniformly on X to a continuous function $f : X \to \mathbb{R}$.*

Proof. Let $\{x_i\}_{i=1}^{\infty}$ be a countably dense set in X. By a Cantor diagonal argument, we can extract a subsequence $\{f_{n_k}\}_{k=1}^{\infty}$ such that for all $i \in \mathbb{N}$, $\{f_{n_k}(x_i)\}_{k=1}^{\infty}$ is a convergent sequence.

Let $\varepsilon > 0$. Since $\{f_n\}_{n=1}^{\infty}$ is approximately equicontinuous there exists $\delta > 0$ such that for all n we have

$$|f_n(x) - f_n(y)| < \frac{\varepsilon}{4} + \frac{1}{n} \text{ for all } x, y \in X \text{ with } |x - y| < \delta. \tag{4.29}$$

The collection of open balls $\{B_{\delta/2}(z)\}_{z \in X}$ forms an open cover of X. Since X is compact, there exists a finite subcover $B_1, \ldots, B_M$ for some integer M. Without loss of generality we may assume that $x_i \in B_i$. Now let $x \in X$. By (4.29) we have

$$\begin{aligned} |f_{n_k}(x) - f_{n_j}(x)| &\leq |f_{n_k}(x) - f_{n_k}(x_i)| + |f_{n_k}(x_i) - f_{n_j}(x_i)| + |f_{n_j}(x_i) - f_{n_j}(x)| \\ &< \frac{\varepsilon}{2} + \frac{1}{n_k} + \frac{1}{n_j} + |f_{n_k}(x_i) - f_{n_j}(x_i)|, \end{aligned}$$

for some $i \in \{1, M\}$ and any k, j. Hence we have

$$\|f_{n_k} - f_{n_j}\|_{L^\infty(X)} \leq \frac{\varepsilon}{2} + \frac{1}{n_k} + \frac{1}{n_j} + \sup_{1 \leq i \leq M} |f_{n_k}(x_i) - f_{n_j}(x_i)|.$$

It follows that $\{f_{n_k}\}_{k=1}^{\infty}$ is Cauchy in L^∞, which completes the proof. □

CHAPTER V

Directed last passage percolation

Let us recall the directed last passage percolation (DLPP) problem formulated in Chapter I. Let $X(i,j)$ be nonnegative independent random variables defined on the lattice $\mathbb{N}^2$, and define the last passage time from $(1,1)$ to (M,N) by

$$L(M,N) = \max_{p \in \Pi_{M,N}} \sum_{(i,j)\in p} X(i,j), \tag{5.1}$$

where $\Pi_{M,N}$ denotes the set of up/right paths from $(1,1)$ to (M,N) in $\mathbb{N}^2$.

One quantity of interest in DLPP is the time constant, U, given by

$$U(x) := \lim_{N\to\infty} \frac{1}{N} L\left(\lfloor Nx \rfloor\right). \tag{5.2}$$

We study the time constant U for the DLPP problem with independent weights $X(i,j)$ that are either geometric or exponential, but not identically distributed. For exponential DLPP, we assume that $X(i,j)$ is exponentially distributed with mean $\lambda(iN^{-1}, jN^{-1})$ where $\lambda : [0,\infty)^2 \to [0,\infty)$, and we consider the aymptotics as $N \to \infty$. The setup is identical for geometric DLPP, except that the macroscopic inhomogeneity is in the parameter q of the geometric distribution. For directed polymers, this models a macroscopic (non-random) inhomogeneity in the strength of impurities; while for TASEP, it corresponds to an inhomogeneity in the rate at which particles move to the right. It turns out that one can employ ideas similar to

those in Chapter III to derive a Hamilton-Jacobi equation for U. This allows us to efficiently compute U via a numerical scheme related to the one studied in Chapter IV, and even compute the asymptotic shapes of optimal DLPP paths via dynamic programming. We summarize our main results in Section 5.1 and present the proofs thereafter.

5.1 Main result

Let us mention the conventions used in this paper. We say X is geometrically distributed with parameter q if

$$\mathbb{P}(X = k) = (1-q)^k q,$$

for $k \in \{0, 1, 2, 3, \dots\}$ and $0 < q \leq 1$, so that we have

$$\mathbb{E}(X) = \frac{1-q}{q} \quad \text{and} \quad \mathrm{Var}(X) = \frac{1-q}{q^2}. \tag{5.3}$$

We say that X is exponentially distributed with mean $\lambda \geq 0$ if

$$\mathbb{P}(X \in dx) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx \quad \text{for } x \in [0, \infty),$$

when $\lambda > 0$, and $X = 0$ with probability one when $\lambda = 0$. Here we have

$$\mathbb{E}(X) = \lambda \quad \text{and} \quad \mathrm{Var}(X) = \lambda^2. \tag{5.4}$$

In order to ensure that our results are applicable to both exponential and geometric DLPP, we parameterize these distributions instead by their mean μ. For the exponential distribution there is no change; we have $\lambda = \mu$. For the geometric distribution, we have by (5.3) that a geometric random variable with mean $\mu \geq 0$ has parameter

$$q = \frac{1}{1+\mu}. \tag{5.5}$$

For both cases, the variance is of course a function of the mean; in the exponential case we have $\sigma = \mu$, and in the geometric case we have $\sigma = \sqrt{\mu(1+\mu)}$.

Let us now present our main result. We consider the following two-sided DLPP model, similar to [7, 15, 26, 9, 19]. Let $X(i,j)$ be independent nonnegative random variables defined on the lattice $\mathbb{N}_0^2$, where $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. Let $L(M, N; Q, P)$ denote the last passage time from $(M, N) \in \mathbb{N}_0^2$ to $(Q, P) \in \mathbb{N}_0^2$, where $M \leq Q$ and $N \leq P$. This is defined as follows:

$$L(M,N;Q,P) = \max_{p \in \Pi_{(M,N),(Q,P)}} \sum_{(i,j)\in p} X(i,j), \tag{5.6}$$

where $\Pi_{(M,N),(Q,P)}$ denotes the set of up/right paths from (M, N) to (Q, P) in $\mathbb{N}_0^2$. The macroscopic inhomogeneity is described by functions $\mu : [0,\infty)^2 \to [0,\infty)$ and $\mu_s : \partial\mathbb{R}_+^2 \to [0,\infty)$, where $\mathbb{R}_+ = (0,\infty)$. Specifically, given a parameter N we make the following assumption:

The weights $X(i,j)$ are independent with mean

$$\mathbb{E}(X(i,j)) = \begin{cases} \mu(iN^{-1}, jN^{-1}), & \text{if } (i,j) \in \mathbb{N}^2, \\ \mu(iN^{-1}, jN^{-1}) + \mu_s(iN^{-1}, jN^{-1}), & \text{if } i = 0 \text{ or } j = 0. \end{cases} \tag{5.7}$$

The term μ corresponds to the macroscopic mean within the bulk $\mathbb{R}_+^2$, and the term μ_s corresponds to an additional source active only on the boundary $\partial\mathbb{R}_+^2$. We postpone the, somewhat technical, hypotheses on μ and μ_s to Section 5.1.1. Roughly speaking, we assume μ is piecewise Lipschitz continuous, with discontinuities restricted to a family of monotone increasing curves. Our main result is the following continuum limit:

Theorem V.1. *Let $\mu : [0,\infty)^2 \to [0,\infty)$ satisfy (F1) and (F3), and let $\mu_s : \partial\mathbb{R}_+^2 \to [0,\infty)$ satisfy (F2). Suppose that the weights $X(i,j)$ satisfy* (5.7) *and are either all*

exponential, or all geometric random variables. In the exponential case, set $\sigma = \mu$*, and in the geometric case, set* $\sigma = \sqrt{\mu(1+\mu)}$*. Then with probability one we have*

$$\frac{1}{N} L(0; \lfloor N \cdot \rfloor) \longrightarrow U \quad \textit{locally uniformly on } [0,\infty)^2, \tag{5.8}$$

where U *is the unique Pareto-monotone viscosity solution of*

$$\text{(P)} \begin{cases} (U_{x_1} - \mu)_+ (U_{x_2} - \mu)_+ = \sigma^2 & \textit{on } \mathbb{R}^2_+, \\ U = \varphi & \textit{on } \partial\mathbb{R}^2_+, \end{cases}$$

and $\varphi(x) = (x_1 + x_2) \int_0^1 \mu(tx) + \mu_s(tx)\, dt$*.*

Here, t_+ denotes the positive part of t given by $\max(t, 0)$. The hypotheses (F1), (F2) and (F3) are described in Section 5.1.1. From now on, (P) will refer to the Hamilton-Jacobi equation above, and not the equations given in Chapters III and IV. Note that (P) generalizes the Hamilton-Jacobi equations from the previous chapters; indeed, if we take $\mu \equiv 0$ and $\sigma^2 = f$, then we obtain the Hamilton-Jacobi equation for the continuum limit of non-dominated sorting.

Theorem V.1 is in many ways analogous to the continuum limit for non-dominated sorting and the longest chain problem given by Theorem III.2. Both of these results can be viewed as a type of stochastic homogenization [87], where the effective Hamiltonian is given in (P). A similar stochastic homogenization result has been obtained recently for first passage percolation [61], though in that case the exact form of the effective Hamiltonian is unknown. The Hamilton-Jacobi equation (P) is also closely related to the conservation law for the hydrodynamic limit of TASEP [43], and in Section 5.1.2 we show a formal equivalence between the two continuum limits.

We believe this new Hamilton-Jacobi equation will prove to be a useful tool for studying the DLPP problem, both theoretically and numerically. As an example, in Section 5.5.2 we show how to combine the numerical solution of this Hamilton-Jacobi

equation with dynamic programming to find the asymptotic shapes of optimal paths. We also believe that this work will provide a new perspective on the hydrodynamic limit of TASEP, and may be useful for studying the corresponding conservation law.

The rest of the chapter is organized as follows: We give the technical assumptions on μ and μ_s in Section 5.1.1, and in Section 5.1.2 we show formally that (P) is equivalent to the conservation law for the hydrodynamic limit of TASEP [43]. The proof of Theorem V.1 is given in Section 5.4 after some preliminary results. In particular, in Section 5.2 we present and analyze a variational problem for (P), and in Section 5.3, we prove a comparison principle for (P). In Section 5.5, we present a fast numerical scheme for computing the viscosity solution of (P), and we present the results of various numerical simulations in Section 5.5.1. Finally, in Section 5.5.2, we give an algorithm based on dynamic programming for finding the asymptotic shape of optimal DLPP paths, and in Section 5.6 we discuss possible directions for future work.

5.1.1 Hypotheses

Let us first introduce some notation. We say a curve Γ in $\mathbb{R}^2$ is continuous and strictly increasing if it can be parameterized in the form

$$\Gamma : t \mapsto (t, f(t)) \quad \text{for } t \in I,$$

where $f : I \to \mathbb{R}$ is continuous and strictly increasing, and I is an interval in $\mathbb{R}$. We make a similar definition for strictly decreasing. Notice that a continuous strictly increasing (resp. decreasing) curve can also be parameterized in the form $t \mapsto (f(t), t)$ where $f : I \to \mathbb{R}$ is continuous and strictly increasing (resp. decreasing). For simplicity, we will also use Γ to denote the locus of points that lie on the curve Γ.

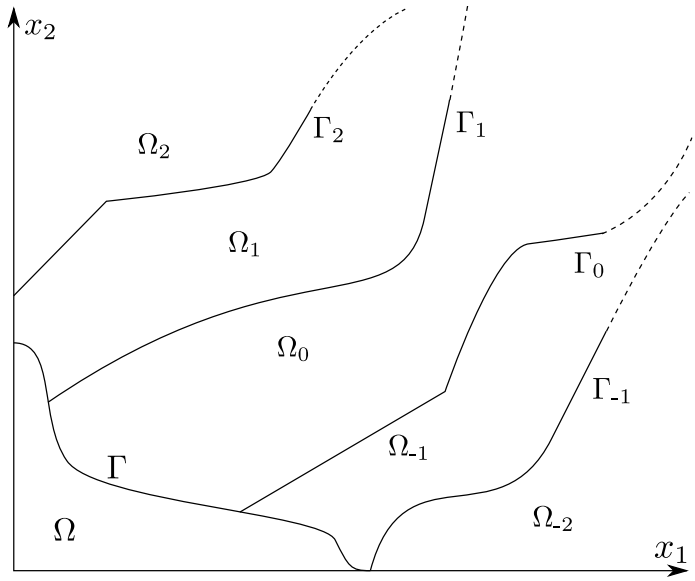


Figure 5.1: Depiction of quantities Ω_i and Γ_i. The function μ is assumed to be Lipschitz with constant C_{lip} when restricted to any Ω_i, and $\mu = 0$ on Ω.

We now describe the assumptions (F1)–(F3). Let Γ be a continuous strictly decreasing curve in $[0,1]^2$ with endpoints $(1,0)$ and $(0,1)$, and let $\Omega \subset [0,\infty)^2$ denote the bounded component of the complement of Γ in $[0,\infty)^2$. Let $\{\Gamma_i\}_{i\in\mathbb{Z}}$ be a locally finite non-intersecting collection of continuous strictly increasing curves. For each i we assume one endpoint of Γ_i is on $\partial([0,\infty)^2 \setminus \Omega)$ and the other endpoint is at ∞, i.e., the curve is unbounded. The complement of $\cup_i \Gamma_i$ in $[0,\infty)^2 \setminus \Omega$ therefore consists of a family $\{\Omega_i\}_{i\in\mathbb{Z}}$ of connected components. Each curve Γ_i is on the boundary of exactly two components, which we may assume are labeled Ω_i and Ω_{i-1}. See Figure 5.1 for an illustration of these quantities.

We place the following assumptions on μ and μ_s:

(F1) The function $\mu : [0,\infty)^2 \to [0,\infty)$ is bounded and upper semicontinuous, $\mu|_\Omega = 0$, and there exists a constant C_{lip} such that for every $i \in \mathbb{Z}$, $\mu|_{\Omega_i}$ is Lipschitz continuous with constant C_{lip}.

(F2) The source term $\mu_s : \partial\mathbb{R}^2_+ \to [0,\infty)$ is bounded and upper semicontinuous with

a locally finite set of discontinuities.

Throughout the paper we will regard μ_s as a function on $[0,\infty)^2$ by setting $\mu_s = 0$ on $\mathbb{R}^2_+$.

We also make the following technical assumption:

(F3) For every $i \in \mathbb{Z}$ and $x \in \Gamma_i$, there exists $\varepsilon > 0$ and $\zeta \in \{-1, +1\}$ such that for all $y \in B_\varepsilon(x) \cap \Gamma_i$ we have

$$\zeta \left(\lim_{\Omega_{i-1} \ni z \to y} \mu(z) - \lim_{\Omega_i \ni z \to y} \mu(z) \right) \geq 0. \tag{5.9}$$

The hypothesis (F3) is used to prove uniqueness of viscosity solutions of (P). Roughly speaking, (F3) requires that any jumps (discontinuities) in μ are locally in the 'same direction'.

In the exponential case, we have $\sigma^2 = \mu^2$, and in the geometric case, we have $\sigma^2 = \mu(1+\mu)$. Thus, if μ satisfies (F1), (F3), then so will σ^2, though possibly with a larger Lipschitz constant C_{lip}. Since it is convenient for the analysis, we will often regard μ and σ^2 as independent functions both satisfying (F1) and (F3). We will only need to recall the relationship between μ and σ at a few key points. In particular, the uniqueness proof for (P) (see Section 5.3) requires that μ and σ^2 satisfy (F3) simultaneously with the same choice of ζ. This is of course always true, since σ is a monotone increasing function of μ in both the exponential and geometric cases.

Let us briefly comment on the significance of Γ and Ω. The correspondence between exponential DLPP and TASEP (described in detail in Section 5.1.2) implies that the initial macroscopic density ρ_0 for TASEP is encoded into the curve Γ. If Γ and Ω are not present, then we have TASEP with the common step initial condition $\rho_0(s) = 1$ for $s \leq 0$ and $\rho_0(s) = 0$ for $s > 0$. Suppose now that Γ and Ω are present, and parameterize Γ by $t \mapsto (t, f(t))$ where f is continuous and strictly decreasing

with $f(0) = 1$ and $f(1) = 0$. Let us assume additionally that f is continuously differentiable. Based on the correspondence between TASEP and exponential DLPP, the initial density will be given by

$$\rho_0(s) = \begin{cases} 1, & \text{if } s \leq -1 \\ -f'(t_s)/(1 - f'(t_s)), & \text{if } s \in (-1, 1) \\ 0, & \text{if } s \geq 1. \end{cases}$$

where for $s \in (-1, 1)$, t_s is the unique $t \in (0, 1)$ satisfying $s = t - f(t)$. Thus by choosing f appropriately, one can obtain a large class of initial densities ρ_0 for TASEP with this setup.

5.1.2 Formal equivalence to hydrodynamic limit of TASEP

We show here a formal equivalence between (P) and the hydrodynamic limit of TASEP, given in [43]. TASEP is an interacting stochastic particle system on $\mathbb{Z}$ with state space $\{0, 1\}^{\mathbb{Z}}$, whose elements, η, represent particle configurations. If a particle is present at site $j \in \mathbb{Z}$, then $\eta_j = 1$, and if no particle is present, then $\eta_j = 0$. The process is exclusionary in the sense that at most one particle can occupy each site at a given time. The stochastic dynamics proceed as follows: a particle at site j jumps to site $j + 1$ after an exponential waiting time, provided the site $j + 1$ is empty. The exponential waiting times are independent and begin at the exact moment the right neighboring site is vacated. These dynamics, along with an initial condition $\eta(0) : \mathbb{Z} \to \{0, 1\}$, generate the stochastic process $\eta = \{\eta_i(t) : i \in \mathbb{Z}, t \in [0, \infty)\}$.

In the standard TASEP model, the exponential waiting times are independent with rate $c = 1$. As in [43], we allow the rates to have a macroscopic spatial (and temporal) dependence, i.e., the rate at position $j \in \mathbb{Z}$ and time $t \in [0, \infty)$ is $c(jN^{-1}, tN^{-1})$, where $c : \mathbb{R} \times [0, \infty) \to (0, \infty)$, and N is a parameter that we will

send to ∞. A central object of study is the macroscopic density $\rho(s,t)$, which is the almost sure limit (assuming it exists) of the discrete densities as follows:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=\lfloor Na\rfloor+1}^{\lfloor Nb\rfloor}\eta_i(Nt)=\int_a^b\rho(s,t)\,ds. \tag{5.10}$$

Georgiu et al. [43] showed that for

$$c(s,t)=c(s)=\begin{cases}c_1, & s\le 0\\ c_2, & s>0,\end{cases}$$

ρ can be identified as the unique entropy solution of the scalar conservation law

$$\rho_t+(c(s)\rho(1-\rho))_s=0,\quad \rho(s,0)=\rho_0(s), \tag{5.11}$$

where ρ_0 denotes the initial macroscopic density. We are using s for the spatial variable in (5.11) to avoid confusion with the spatial variables in (P). In what follows, we show formally that the conservation law (5.11) is equivalent to (P). For simplicity, we will ignore the initial condition ρ_0 and the boundary condition in (P), and restrict ourselves to showing that the (P) and (5.11) are equivalent in the bulk. We shall also assume that $\rho\in C^1$.

Consider now the exponential DLPP model with macroscopic mean $\lambda:[0,\infty)^2\to(0,\infty)$, i.e., $\mu=\sigma=\lambda$. Let L denote the last passage time given by (5.6), and let us write $L(m,n)=L(1,1;m,n)$ for convenience. Let U be the unique Pareto-monotone viscosity solution of (P), and let us assume that $U\in C^1$ and $\lambda>0$ so that $U_{x_1},U_{x_2}>\lambda>0$. By Theorem V.1 we have

$$\frac{1}{N}L(\lfloor Nx\rfloor)\longrightarrow U(x)\ \text{ with probability one.} \tag{5.12}$$

We also note that (P) can be rearranged as follows:

$$\frac{U_{x_1}(x)U_{x_2}(x)}{U_{x_1}(x)+U_{x_2}(x)}=\lambda(x). \tag{5.13}$$

Let us now describe in detail the correspondence between TASEP and DLPP, which can also be found here [78, 8]. We assign to a TASEP configuration η the site counter

$$I_j(t) = \#\text{ of particles that have jumped from site } j \text{ to } j+1 \text{ up to time } t. \tag{5.14}$$

and the height function

$$h_j(t) = \begin{cases} 2I_0(t) + \sum_{i=1}^{j} \big(1 - 2\eta_i(t)\big), & j \geq 1, \\ 2I_0(t), & j = 0, \\ 2I_0(t) + \sum_{i=j+1}^{0} \big(1 - 2\eta_i(t)\big), & j \leq -1. \end{cases} \tag{5.15}$$

Then we have $h_0(0) = 0$, and $h_j(t) - h_j(0) = 2I_j(t)$. The height function $h_j(t)$ is a stochastically growing interface, and is related to the corner growth model described earlier. Roughly speaking, the dynamical rule for the growth of $h_j(t)$ is that when a particle jumps to the right (from j to $j+1$), a valley $\diagdown\diagup$ turns into a mountain $\diagup\diagdown$, and the height at site j increases by 2.

Let us now define the random set

$$A(t) = \Big\{(m,n) \in \mathbb{Z}_+^2 \,:\, L(m,n) \leq t\Big\}.$$

Since L is non-decreasing in both arguments, it implicitly defines its own height function, $\widetilde{h}_j(t)$, which describes the boundary of $A(t)$ as follows:

$$A(t) = \Big\{(m,n) \in \mathbb{Z}_+^2 \,:\, \widetilde{h}_{m-n}(t) \geq m+n\Big\}.$$

The correspondence between TASEP and DLPP is the identification $\widetilde{h}_j(t) = h_j(t)$ in the sense of joint distributions. This connection is made rigorous by choosing appropriate boundary rates for DLPP [78]. Visually, the correspondence is obtained

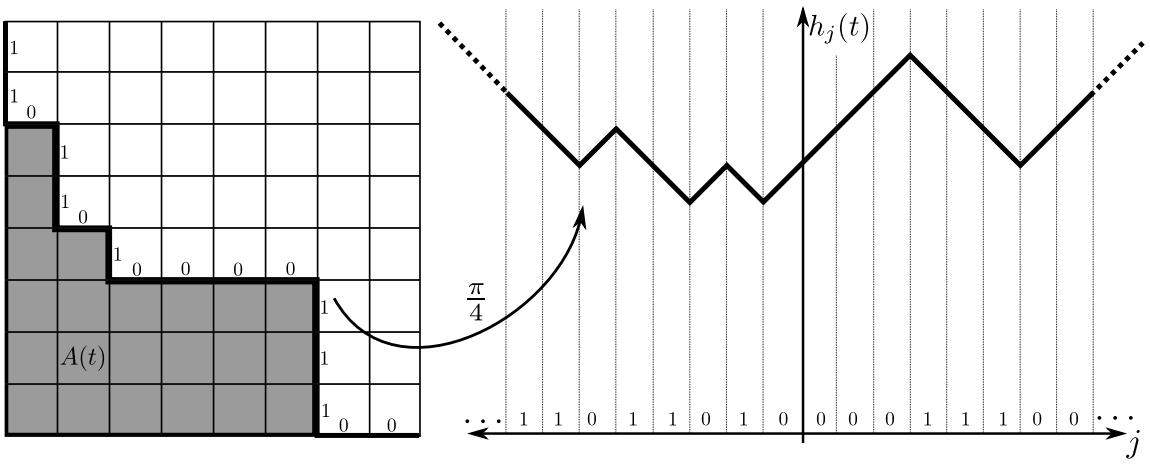


Figure 5.2: A visual depiction of the correspondence between TASEP and DLPP. On the left, the gray region is the set $A(t)$—the t sub-level set of L—and on the right we show the corresponding TASEP height function $h_j(t)$ obtained by rotating the boundary of $A(t)$ by $\pi/4$.

by rotating the boundary of $A(t)$ by $\pi/4$ to obtain the height function $h_j(t)$ (see Figure 5.2).

The correspondence between TASEP and DLPP says, at least formally, that

$$\Big\{(m,n) \in \mathbb{Z}_+^2 \,:\, L(m,n) \le t\Big\} = \Big\{(m,n) \in \mathbb{Z}_+^2 \,:\, h_{m-n}(t) \ge m+n\Big\}. \tag{5.16}$$

By (5.10) and (5.15), $h_j(t)$ has a macroscopic continuum limit, h^∞, such that

$$\frac{1}{N} h_{\lfloor sN \rfloor}(\lfloor tN \rfloor) \longrightarrow h^\infty(s,t) = g(t) + s - 2\int_0^s \rho(s',t)\, ds', \tag{5.17}$$

with probability one, where $g(t) := \lim_{N\to\infty} 2N^{-1} I_0(tN)$. It follows from (5.17) that

$$h_s^\infty(s,t) = 1 - 2\rho(s,t). \tag{5.18}$$

Combining (5.12), (5.16), and (5.17) we have that

$$\Big\{x \in \mathbb{R}_+^2 \,:\, U(x) = t\Big\} = \Big\{x \in \mathbb{R}_+^2 \,:\, h^\infty(x_1 - x_2, t) = x_1 + x_2\Big\}. \tag{5.19}$$

It follows from (5.19) that

$$h^\infty\big(x_1 - x_2, U(x)\big) = x_1 + x_2. \tag{5.20}$$

This is in some sense the "master equation" relating the continuum limits of TASEP and DLPP. Let us illustrate how to use (5.20) to derive the conservation law (5.11) from (P); deriving (P) from (5.11) follows in a similar fashion.

Differentiating (5.20) in both x_1 and x_2 we have

$$h_s^\infty(s,t) + h_t^\infty(s,t)U_{x_1}(x) = 1 \tag{5.21}$$

$$-h_s^\infty(s,t) + h_t^\infty(s,t)U_{x_2}(x) = 1. \tag{5.22}$$

where $t = U(x)$ and $s = x_1 - x_2$. Adding (5.21) and (5.22) we have

$$h_t^\infty(s,t) = \frac{2}{U_{x_1}(x) + U_{x_2}(x)}. \tag{5.23}$$

Similarly, by rearranging and dividing (5.21) by (5.22) we have

$$\frac{U_{x_1}(x)}{U_{x_2}(x)} = \frac{1 - h_s^\infty(s,t)}{1 + h_s^\infty(s,t)} \overset{(5.18)}{=} \frac{\rho(s,t)}{1-\rho(s,t)}. \tag{5.24}$$

This equality can also be obtained by noting that the slope of the level set $\{U(x) = t\}$ is given locally by the ratio of ones to zeros in the TASEP configuration.

Solving for ρ in (5.24) we have $\rho = U_{x_1}/(U_{x_1} + U_{x_2})$, which yields

$$\rho(s,t)\big(1-\rho(s,t)\big) = \frac{U_{x_1}(x)U_{x_2}(x)}{(U_{x_1}(x) + U_{x_2}(x))^2} \overset{(5.13)}{=} \frac{\lambda(x)}{U_{x_1}(x) + U_{x_2}(x)}, \tag{5.25}$$

where we invoked the Hamilton-Jacobi equation (P) in the second equality above. Since U is strictly monotone increasing in both x_1 and x_2, there is a one-to-one correspondence between the coordinates $x = (x_1, x_2)$ and $(s,t) = (x_1 - x_2, U(x))$. Let us write $c(s,t) := \lambda(x)^{-1}$. Since λ is the exponential mean, c is the exponential rate for TASEP. Then combining (5.25) with (5.23) we have

$$h_t^\infty(s,t) = 2c(s,t)\rho(s,t)\big(1-\rho(s,t)\big). \tag{5.26}$$

Differentiating with respect to s on both sides of (5.26) and applying (5.18) we have

$$-2\rho_t(s,t) = 2\big(c(s,t)\rho(s,t)(1-\rho(s,t))\big)_s, \tag{5.27}$$

which is precisely the conservation law (5.11). Furthermore, by combining (5.26) and (5.18), we have the following Hamilton-Jacobi equation for h^∞:

$$h_t^\infty(s,t) = \frac{c(s,t)}{2}\left(1 - h_s^\infty(s,t)^2\right). \tag{5.28}$$

5.2 Variational problem

In this section we give a variational interpretation for U and analyze its relevant properties. This variational problem first appeared in [81], in a different form, for exponential DLPP with a continuous macroscopic rate λ, and is similar to the well-known variational problem for the longest chain problem [31].

Let $\mathcal{A}$ denote the set of C^1 monotone curves, given by

$$\mathcal{A} = \Big\{\gamma \in C^1([0,1];[0,\infty)^2) \,:\, \gamma'(t) \geqq 0 \text{ for all } t \in [0,1]\Big\}. \tag{5.29}$$

We write $\gamma(t) = (\gamma_1(t), \gamma_2(t))$ to denote the components of γ. For $\mu, \sigma : [0,\infty)^2 \to \mathbb{R}$, let us define $\ell_{\mu,\sigma} : [0,\infty)^2 \times [0,\infty)^2 \to [0,\infty)$ by

$$\ell_{\mu,\sigma}(x,p) = \mu(x)(p_1+p_2) + 2\sigma(x)\sqrt{p_1 p_2}, \tag{5.30}$$

and for $\gamma \in \mathcal{A}$ we set

$$J_{\mu,\sigma}(\gamma) = \int_0^1 \ell_{\mu,\sigma}(\gamma(t), \gamma'(t))\, dt. \tag{5.31}$$

We finally define

$$U_{\mu,\sigma}(x) = \sup\Big\{J_{\mu,\sigma}(\gamma) \,:\, \gamma \in \mathcal{A},\ \gamma(0) = 0,\ \text{and } \gamma(1) = x\Big\}, \tag{5.32}$$

for $x \in [0,\infty)^2$. Borrowing language from optimal control theory [10], we will call $U_{\mu,\sigma}$ the value function for this variational problem. We will often write J, ℓ and U in place of $J_{\mu,\sigma}, \ell_{\mu,\sigma}$ and $U_{\mu,\sigma}$, respectively, when it is clear from the context what μ

and σ are. Notice that when $x \in \partial\mathbb{R}^2_+$ with $x_2 = 0$ we have

$$U(x) = \int_0^{x_1} \mu(t, 0)\, dt. \tag{5.33}$$

A similar formula holds when $x \in \partial\mathbb{R}^2_+$ with $x_1 = 0$, and in general we can write

$$U(x) = (x_1 + x_2) \int_0^1 \mu(tx)\, dt, \tag{5.34}$$

for $x \in \partial\mathbb{R}^2_+$.

We also define

$$W_{\mu,\sigma}(x, y) = \sup\Big\{ J_{\mu,\sigma}(\gamma) \,:\, \gamma \in \mathcal{A},\ \gamma(0) = x,\ \text{and } \gamma(1) = y \Big\}, \tag{5.35}$$

for $x, y \in [0, \infty)^2$ with $x \leqq y$. As before, we will often drop the subscripts on $W_{\mu,\sigma}$ when convenient. Similar to (5.33)–(5.34), when $x, y \in [0, \infty)^2$ with $x \leqq y$ and $x_2 = y_2$, we can write

$$W(x, y) = \int_{x_1}^{y_1} \mu(t, x_2)\, dt, \tag{5.36}$$

with a similar formula holding when $x_1 = y_1$. In general, whenever $x \leqq y$ but $x_i = y_i$ for some i we can write

$$W(x, y) = (y_1 - x_1 + y_2 - x_2) \int_0^1 \mu(x + (y - x)t)\, dt. \tag{5.37}$$

The remainder of this section is organized as follows. In Section 5.2.1 we prove that U and W are uniformly continuous, under assumptions on μ and σ that are similar to (F1) and (F3), but slightly weaker. Then in Section 5.2.2, we show that $U_{\mu+\mu_s,\sigma}$ is a viscosity solution of (P), and prove a similar result for $W_{\mu,\sigma}$. This result, Theorem V.8, follows from classical optimal control theory [10], and (P) is exactly the Hamilton-Jacobi-Bellman equation for the variational (optimal-control) problem (5.32). For more information on Hamilton-Jacobi equations and optimal control, we refer the reader to [10].

5.2.1 Regularity

Hölder or Lipschitz regularity of the value function in optimal control theory is a standard classical result [10]. However, it is typically assumed that $x \mapsto \ell_{\mu,\sigma}(x,p)$ is uniformly continuous, which is not compatible with (F1). We show here that the specific form of $\ell_{\mu,\sigma}$ allows us to show that $U_{\mu+\mu_s,\sigma}$ and $W_{\mu,\sigma}$ are uniformly continuous, provided the discontinuities in μ occur along monotone increasing curves.

Since it is useful later, we will slightly weaken the hypothesis (F1), and allow μ to be "badly behaved" within a narrow tube of the monotone curves Γ_i. This weakened hypothesis is specifically designed so that the regularity result applies to inf- and sup-convolutions of functions satisfying (F1). Inf- and sup-convolutions (defined in Remark V.18) are commonly used for regularization in the theory of viscosity solutions [10, 27].

The weakened hypothesis requires the following notation; for $\theta \geq 0$ define

$$\Gamma_{i,\theta} = \Big\{ x \in [0,\infty)^2 \,:\, \operatorname{dist}(x,\Gamma_i) \leq \theta \Big\}, \tag{5.38}$$

$$\Omega_{i,\theta} = \Big\{ x \in \Omega_i \,:\, \operatorname{dist}(x,\Gamma_i) > \theta \ \text{ and } \ \operatorname{dist}(x,\Gamma_{i+1}) > \theta \Big\}, \tag{5.39}$$

$$\Gamma_{\theta} = \Big\{ x \in [0,\infty)^2 \,:\, \operatorname{dist}(x,\Gamma) \leq \theta \Big\}, \tag{5.40}$$

$$\Omega_{\theta} = \Big\{ x \in \Omega \,:\, \operatorname{dist}(x,\Gamma) > \theta \Big\}. \tag{5.41}$$

The weakened version of (F1) is the following:

(F1*) The function $\mu : [0,\infty)^2 \to [0,\infty)$ is bounded and upper semicontinuous, $\mu|_{\Omega_\theta} = 0$, and there exists a constant C_{lip} such that for every $i \in \mathbb{Z}$, $\mu|_{\Omega_{i,\theta}}$ is Lipschitz continuous with constant C_{lip}.

We now give the regularity result for W.

Theorem V.2. *Suppose that μ satisfies (F1*) for $\theta \geq 0$, and suppose that $\sigma : [0,\infty)^2 \to [0,\infty)$ is bounded and Borel-measurable. Then for every $R > 0$ there exist a modulus of continuity ω, and a constant $C = C(C_{lip}, \|\mu\|_\infty, \|\sigma\|_\infty, R) > 0$ such that*

$$|W_{\mu,\sigma}(z,x) - W_{\mu,\sigma}(z,y)| \leq C\left(\sqrt{|x-y|} + \omega(|x-y|) + \omega(\theta)\right), \tag{5.42}$$

for all $x, y, z \in [0,R]^2$ with $x, y \geqq z$. Furthermore, ω depends only on $\Gamma, \{\Gamma_i\}_{i\in\mathbb{Z}}$ and $R > 0$.

Proof. Let $R > 0$. We will prove the result for $z = 0$; the case of $z \neq 0$ is very similar. For simplicity of notation, let us set $V(x) = W_{\mu,\sigma}(0,x)$. Notice that we can reduce the proof to the case where $x, y \in [0,R]^2$ with $x \leqq y$. Indeed, let $x, y \in [0,R]^2$ and set $x' = (\min(x_1,y_1), \min(x_2,y_2))$. Then we have

$$|U(x) - U(y)| \leq |U(x) - U(x')| + |U(y) - U(x')|,$$

and $x' \leqq x$ and $x' \leqq y$.

Thus let us assume that $x \leqq y$. Let $\varepsilon > 0$ and let $\gamma \in \mathcal{A}$ such that $\gamma(0) = 0$, $\gamma(1) = y$, and $V(y) \leq J(\gamma) + \varepsilon$. Define

$$s_1 = \sup\left\{t > 0 \,:\, \gamma(t) \leqq x\right\} \quad \text{and} \quad s_2 = \inf\left\{t > 0 \,:\, \gamma(t) \geqq x\right\}.$$

Without loss of generality, we may assume that $\gamma_2(s_2) = x_2$. Define

$$\overline{\gamma}(t) = \Big(\min\big(x_1, \gamma_1(t)\big), \gamma_2(t)\Big) \quad \text{for } t \in [0, s_2].$$

The proof is split into two steps now.

1. We claim that

$$|V(x) - V(y)| \leq \int_{s_1}^{s_2} |\mu(\gamma(t)) - \mu(\overline{\gamma}(t))|\, \gamma_2'(t)\, dt + C\sqrt{|x-y|} + \varepsilon. \tag{5.43}$$

where $C = C(\|\mu\|_\infty, \|\sigma\|_\infty, R)$.

To see this: First note that $\gamma(s_2) \geqq x$ and $\gamma(1) = y$. It follows that

$$\begin{aligned}
\int_{s_2}^1 \ell(\gamma(t), \gamma'(t))\, dt &\leq \|\mu\|_\infty \int_{s_2}^1 \gamma_1'(t) + \gamma_2'(t)\, dt + 2\|\sigma\|_\infty \int_{s_2}^1 \sqrt{\gamma_1'(t)\gamma_2'(t)}\, dt \\
&\leq 2\|\mu\|_\infty |x - y| + 2\|\sigma\|_\infty \left(\int_{s_2}^1 \gamma_1'(t)\, dt \int_{s_2}^1 \gamma_2'(t)\, dt \right)^{\frac{1}{2}} \\
&\leq 2\|\mu\|_\infty |x - y| + 2\|\sigma\|_\infty |x - y| \\
&= 2(\|\mu\|_\infty + \|\sigma\|_\infty)|x - y|, \qquad (5.44)
\end{aligned}$$

where the second line follows from Hölder's inequality. We claim now that $\gamma_1(s_1) = x_1$. To see this: suppose to the contrary that $\gamma_1(s_1) < x_1$, which implies that $s_1 < s_2$. By the definition of s_1 we must have $\gamma_2(s_1) = x_2$ and $\gamma_2(s) > x_2$ for $s > s_1$. This contradicts our assumption that $\gamma_2(s_2) = x_2$. Hence $\gamma_1(s_1) = x_1$.

Now we have

$$\int_{s_1}^{s_2} \gamma_1'(t)\, dt = \gamma_1(s_2) - \gamma_1(s_1) \leq y_1 - x_1 \leq |x - y|. \qquad (5.45)$$

Since $\gamma = \overline{\gamma}$ on $[0, s_1]$ and $\overline{\gamma}(s_2) = x$ we have

$$\begin{aligned}
V(y) - V(x) &\leq J(\gamma) + \varepsilon - \int_0^{s_2} \ell(\overline{\gamma}(t), \overline{\gamma}'(t))\, dt \\
&= \int_{s_1}^{s_2} \ell(\gamma(t), \gamma'(t)) - \ell(\overline{\gamma}(t), \overline{\gamma}'(t))\, dt + \int_{s_2}^1 \ell(\gamma(t), \gamma'(t))\, dt + \varepsilon \\
&\overset{(5.44)}{\leq} \underbrace{\int_{s_1}^{s_2} \ell(\gamma(t), \gamma'(t)) - \ell(\overline{\gamma}(t), \overline{\gamma}'(t))\, dt}_{A} + C(\|\mu\|_\infty, \|\sigma\|_\infty)|x - y| + \varepsilon. \qquad (5.46)
\end{aligned}$$

If $s_1 = s_2$ then the claim (5.43) follows from (5.46). So suppose that $s_1 < s_2$. Since

$\overline{\gamma}_1'(t) = 0$ and $\overline{\gamma}_2'(t) = \gamma_2'(t)$ for $t \in (s_1, s_2)$, we have

$$\begin{aligned} A &= \int_{s_1}^{s_2} \left(\mu(\gamma(t)) - \mu(\overline{\gamma}(t))\right) \gamma_2'(t) + \mu(\gamma(t))\gamma_1'(t) + 2\sigma(\gamma(t))\sqrt{\gamma_1'(t)\gamma_2'(t)}\, dt \\ &\leq \int_{s_1}^{s_2} |\mu(\gamma(t)) - \mu(\overline{\gamma}(t))|\, \gamma_2'(t)\, dt + \|\mu\|_\infty \int_{s_1}^{s_2} \gamma_1'(t)\, dt \\ &\qquad + 2\|\sigma\|_\infty \left(\int_{s_1}^{s_2} \gamma_1'(t)\, dt \int_{s_1}^{s_2} \gamma_2'(t)\, dt\right)^{\frac{1}{2}} \\ &\overset{(5.45)}{\leq} \int_{s_1}^{s_2} |\mu(\gamma(t)) - \mu(\overline{\gamma}(t))|\, \gamma_2'(t)\, dt + C(\|\mu\|_\infty, \|\sigma\|_\infty, R)\sqrt{|x-y|} \end{aligned} \tag{5.47}$$

which establishes (5.43).

2. We claim that

$$\int_{s_1}^{s_2} |\mu(\gamma(t)) - \mu(\overline{\gamma}(t))|\, \gamma_2'(t)\, dt \leq C\left(\sqrt{|x-y|} + \omega(|x-y|) + \omega(\theta) + \theta\right), \tag{5.48}$$

where $C = C(C_{lip}, R, \|\mu\|_\infty, \|\sigma\|_\infty)$. Notice that once (5.48) is established, the proof is completed by combining (5.48) with (5.43) and sending $\varepsilon \to 0$.

Since the collection of curves $\{\Gamma_i\}_{i=-\infty}^{\infty}$ is locally finite, we may assume that $\Gamma_{1,\theta}, \ldots, \Gamma_{M,\theta}$ are the only tubular neighborhoods that have a non-empty with intersection $[0, R]^2$. Since Γ_i is continuous and strictly increasing, we can parameterize the portion of Γ_i that intersects $[0, R]^2$ as follows:

$$\Gamma_i : t \mapsto (t, f_i(t)), \quad t \in I_i,$$

where $f_i : I_i \to [0, \infty)$ is continuous and strictly increasing, and I_i is a closed interval in $[0, R]$. Similarly we can parameterize Γ as

$$\Gamma : t \mapsto (t, f(t)), \quad t \in [0, 1],$$

where $f : [0, 1] \to [0, 1]$ is continuous and strictly decreasing. Note that the functions $f_1, \ldots, f_M, f$ share a common modulus of continuity ω, by virtue of their compact domains. We also note that ω and M depend only on $\Gamma, \{\Gamma_i\}_{i\in\mathbb{Z}}$, and $R > 0$.

To prove (5.48), first set $c = \omega(\theta) + \theta$. A simple computation shows that

$$\text{dist}((t, f_i(t) + c), \Gamma_i) > \theta \quad \text{and} \quad \text{dist}((t, f_i(t) - c), \Gamma_i) > \theta, \tag{5.49}$$

for any $t \in I_i$. A similar statement holds for Γ and f. For each $i \in \{1, \dots, M\}$, we define

$$m_i^+ = \sup_{I_i \cap [x_1, y_1]} f_i, \quad \text{and} \quad m_i^- = \inf_{I_i \cap [x_1, y_1]} f_i,$$

and

$$K_i = \Big\{ t \in (s_1, s_2) \,:\, (x_1, m_i^- - c) \leqq \gamma(t) \leqq (y_1, m_i^+ + c) \Big\}. \tag{5.50}$$

Similarly we set

$$m^+ = \sup_{[0,1] \cap [x_1, y_1]} f, \quad \text{and} \quad m^- = \inf_{[0,1] \cap [x_1, y_1]} f,$$

$$K = \Big\{ t \in (s_1, s_2) \,:\, (x_1, m^- - c) \leqq \gamma(t) \leqq (y_1, m^+ + c) \Big\}. \tag{5.51}$$

and

$$\mathcal{H} = (s_1, s_2) \setminus (K \cup K_1 \cup \cdots \cup K_M) \tag{5.52}$$

By the definition of $m_i^\pm$ and $m^\pm$ we have

$$m_i^+ - m_i^- \leq \omega(y_1 - x_1) \quad \text{and} \quad m^+ - m^- \leq \omega(y_1 - x_1). \tag{5.53}$$

It follows from (5.49)–(5.52), (F1*), and the fact that γ is monotone, that whenever $t \in \mathcal{H}$ we have either $\mu(\gamma(t)) = \mu(\overline{\gamma}(t)) = 0$ or

$$\mu(\gamma(t)) - \mu(\overline{\gamma}(t)) = \mu_{i,\theta}(\gamma(t)) - \mu_{i,\theta}(\overline{\gamma}(t)),$$

for some $i \in \{0, \dots, M\}$. Thus, invoking (F1*), we have

$$|\mu(\gamma(t)) - \mu(\overline{\gamma}(t))| \leq C_{lip} |\gamma(t) - \overline{\gamma}(t)| \leq C_{lip} |x - y|, \tag{5.54}$$

for all $t \in \mathcal{H}$.

For any $i \in \{1, \dots, M\}$, we have

$$\begin{aligned}\int_{K_i} |\mu(\gamma(t)) - \mu(\overline{\gamma}(t))| \gamma_2'(t)\, dt &\leq 2\|\mu\|_\infty \int_{K_i} \gamma_2'(t)\, dt \\ &\leq 2\|\mu\|_\infty |m_i^+ - m_i^- + 2c| \\ &\overset{(5.53)}{\leq} 2\|\mu\|_\infty \omega(|x-y|) + 4\|\mu\|_\infty(\omega(\theta)+\theta). \end{aligned} \tag{5.55}$$

We have an identical estimate when K_i is replaced by K. Combining (5.54) with (5.55) we have

$$\begin{aligned}&\int_{s_1}^{s_2} |\mu(\gamma(t)) - \mu(\overline{\gamma}(t))|\, \gamma_2'(t)\, dt \\ &\quad = \int_{\mathcal{H}} |\mu(\gamma(t)) - \mu(\overline{\gamma}(t))|\, \gamma_2'(t)\, dt + \int_{K \cup K_1 \cup \cdots \cup K_M} |\mu(\gamma(t)) - \mu(\overline{\gamma}(t))|\, \gamma_2'(t)\, dt \\ &\quad \leq C_{lip}|x-y| \int_{s_1}^{s_2} \gamma_2'(t)\, dt + \sum_{i=1}^{M} \int_{K_i} |\mu(\gamma(t)) - \mu(\overline{\gamma}(t))|\, \gamma_2'(t)\, dt \\ &\qquad\qquad + \int_K |\mu(\gamma(t)) - \mu(\overline{\gamma}(t))|\, \gamma_2'(t)\, dt \\ &\quad \leq C_{lip} R|x-y| + 2(M+1)\|\mu\|_\infty \omega(|x-y|) + 4(M+1)\|\mu\|_\infty(\omega(\theta)+\theta), \end{aligned} \tag{5.56}$$

which establishes (5.48) and completes the proof. □

Corollary V.3. *Suppose that μ satisfies (F1*) for $\theta \geq 0$, and suppose that σ is bounded and Borel-measurable. Then for every $R > 0$ there exist a modulus of continuity ω, and a constant $C = C(C_{lip}, \|\mu\|_\infty, \|\sigma\|_\infty, R) > 0$ such that*

$$|W_{\mu,\sigma}(x,z) - W_{\mu,\sigma}(y,z)| \leq C\left(\sqrt{|x-y|} + \omega(|x-y|) + \omega(\theta)\right), \tag{5.57}$$

for all $x, y, z \in [0,R]^2$ with $x, y \leqq z$. As in Theorem V.2, ω depends only on $\Gamma, \{\Gamma_i\}_{i\in\mathbb{Z}}$ and $R > 0$.

Proof. The proof follows from Theorem V.2 by symmetry. □

Remark V.4. Notice in Theorem V.2 that if $\theta = 0$ then we have the estimate

$$|W_{\mu,\sigma}(z,x) - W_{\mu,\sigma}(z,y)| \leq C\left(\sqrt{|x-y|} + \omega(|x-y|)\right), \tag{5.58}$$

for all $x, y, z \in [0,R]^2$ with $x, y \geq z$. Inspecting the proof of Theorem V.2, we see that ω is the modulus of continuity of the curves $\Gamma, \{\Gamma_i\}_{i\in\mathbb{Z}}$ as functions over both coordinate axes. Thus, the regularity of W is inherited from the regularity of the curves $\Gamma, \{\Gamma_i\}_{i\in\mathbb{Z}}$. For example, if the curves $\Gamma, \{\Gamma_i\}_{i\in\mathbb{Z}}$ are Hölder-continuous with exponent $\alpha \leq 1/2$ as functions over both coordinate axes, then we have that $W(z,\cdot) \in C^{0,\alpha}([z_1,R]\times[z_2,R])$ for every $R > 0$ and its Hölder seminorm depends only on $\|\mu\|_\infty$, $\|\sigma\|_\infty$, R, and C_{lip}. The same remark holds for Corollary V.3 and (5.57).

We now plan to use Theorem V.2 to prove a similar regularity result for $U_{\mu+\mu_s,\sigma}$. To do this, we relate W and U via the following dynamic programming principle:

Proposition V.5. *Suppose that μ satisfies (F1*) for $\theta \geq 0$, μ_s satisfies (F2), and suppose that σ is bounded and Borel-measurable. Then for any $y \in [0,\infty)^2$ we have*

$$U_{\mu+\mu_s,\sigma}(y) = \max_{x\in\partial\mathbb{R}^2_+ \,:\, x\leqq y}\Big\{U_{\mu+\mu_s,\sigma}(x) + W_{\mu,\sigma}(x,y)\Big\}. \tag{5.59}$$

Notice that the boundary source μ_s is absent in the term $W_{\mu,\sigma}$ in (5.59). This allows us to concentrate much of our analysis on $W_{\mu,\sigma}$, which involves only the macroscopic inhomogeneities in the bulk $\mathbb{R}^2_+$, and then extend our results to hold for $U_{\mu+\mu_s,\sigma}$ via the dynamic programming principle (5.59).

Proof. We first note that the maximum in (5.59) is indeed attained, due to the continuity of $U_{\mu+\mu_s,\sigma}$ restricted to $\partial\mathbb{R}^2_+$ and Corollary V.3.

If $y \in \partial\mathbb{R}^2_+$, then in light of (5.34), (5.37) and the fact that $\mu_s \geq 0$, the maximum in (5.59) is attained at $x = y$ and the validity of (5.59) is trivial.

Suppose now that $y \in \mathbb{R}^2_+$ and let $v(y)$ denote the right hand side in (5.59), and set $U = U_{\mu+\mu_s,\sigma}$. We first show that $U \leq v$. Let $\varepsilon > 0$ and $\gamma \in \mathcal{A}$ such that $\gamma(0) = 0$, $\gamma(1) = y$ and $J_{\mu+\mu_s,\sigma}(\gamma) \geq U(y) - \varepsilon$. Let

$$s = \sup\big\{t \in [0,1] \,:\, \gamma(t) \in \partial\mathbb{R}^2_+\big\}.$$

Then we have $0 \leq s < 1$. Set $x = \gamma(s)$ and

$$\gamma^1(t) = \gamma(st) \;\text{ and }\; \gamma^2(t) = \gamma\big(s + t(1-s)\big),$$

for $t \in [0,1]$. Then we have

$$U(y) \leq J_{\mu+\mu_s,\sigma}(\gamma) + \varepsilon = J_{\mu+\mu_s,\sigma}(\gamma^1) + J_{\mu,\sigma}(\gamma^2) + \varepsilon \leq U(x) + W(x,y) + \varepsilon.$$

Sending $\varepsilon \to 0$ we have $U \leq v$.

We now show that $v \leq U$. Let $x \in \partial\mathbb{R}^2_+$ be a point at which the maximum is attained in (5.59) and let $\varepsilon > 0$. Let $\gamma^1 \in \mathcal{A}$ with $\gamma^1(0) = 0$, $\gamma^1(1) = x$ such that $U(x) \leq J_{\mu+\mu_s,\sigma}(\gamma^1) + \frac{\varepsilon}{3}$. Let $z \in [x,y]$ such that $z \in \mathbb{R}^2_+$ and $W(x,y) \leq W(z,y) + \frac{\varepsilon}{3}$. Let $\gamma^2 \in \mathcal{A}$ with $\gamma^2(0) = z$, $\gamma^2(1) = y$ such that $W(z,y) \leq J_{\mu,\sigma}(\gamma^2) + \frac{\varepsilon}{3}$. We can stitch together γ^1 and γ^2 as follows

$$\gamma(t) = \begin{cases} \gamma^1(3t), & \text{if } 0 \leq t < \frac{1}{3}, \\ x + (3t-1)(z-x), & \text{if } \frac{1}{3} \leq t < \frac{2}{3}, \\ \gamma^2(3t-2), & \text{if } \frac{2}{3} \leq t \leq 1. \end{cases}$$

Then we have

$$\begin{aligned} v(y) = U(x) + W(x,y) &\leq U(x) + W(z,y) + \frac{\varepsilon}{3} \\ &\leq J_{\mu+\mu_s,\sigma}(\gamma^1) + J_{\mu,\sigma}(\gamma^2) + \varepsilon \leq J_{\mu+\mu_s,\sigma}(\gamma) + \varepsilon \leq U(y) + \varepsilon, \end{aligned}$$

where we used the fact that $\gamma^2(t) \in \mathbb{R}^2_+$ for all t, hence $J_{\mu,\sigma}(\gamma^2) = J_{\mu+\mu_s,\sigma}(\gamma^2)$. Sending $\varepsilon \to 0$ we have $v \leq U$. $\square$

Before continuing with the regularity result for $U_{\mu+\mu_s,\sigma}$, let us introduce a bit of notation. For $\xi \in \mathbb{R}^d_+$, let $\pi_\xi : \mathbb{R}^d \to [0,\xi]$ denote the projection mapping $\mathbb{R}^d$ onto the convex set $[0,\xi]$. For $x \in [0,\infty)^d$, π_ξ is given explicitly by

$$\pi_\xi(x) = \Big(\min(x_1,\xi_1), \dots, \min(x_d,\xi_d) \Big). \tag{5.60}$$

Corollary V.6. *Suppose that μ satisfies (F1*) for $\theta \geq 0$, μ_s satisfies (F2), and suppose that σ is bounded and Borel-measurable. Then for every $R > 0$ there exists a modulus of continuity ω, and a constant $C = C(C_{lip}, \|\mu\|_\infty, \|\sigma\|_\infty, \|\mu_s\|_\infty, R) > 0$ such that*

$$|U_{\mu+\mu_s,\sigma}(x) - U_{\mu+\mu_s,\sigma}(y)| \leq C\left(\sqrt{|x-y|} + \omega(|x-y|) + \omega(\theta)\right), \tag{5.61}$$

for all $x, y \in [0,R]^2$. As in Theorem V.2, ω depends only on $\Gamma, \{\Gamma_i\}_{i\in\mathbb{Z}}$ and $R > 0$.

Proof. Let $x, y \in [0,R]^2$ and set $U = U_{\mu+\mu_s,\sigma}$ and $W = W_{\mu,\sigma}$. As in Theorem V.2 we may assume that $x \leqq y$. By Proposition V.5, there exists $y' \in \partial\mathbb{R}^2_+$ with $y' \leqq y$ such that

$$U(y) = U(y') + W(y', y). \tag{5.62}$$

Set $x' = \pi_x(y')$. Then since $x' \in \partial\mathbb{R}^2_+$ and $x' \leqq x$, we have by Proposition V.5 that

$$U(x) \geq U(x') + W(x', x). \tag{5.63}$$

By subtracting (5.63) from (5.62) and recalling (5.34) we have

$$\begin{aligned} |U(x) - U(y)| &= U(y) - U(x) \\ &\leq U(y') - U(x') + W(y', y) - W(x', x) \\ &\leq \|\mu + \mu_s\|_\infty |x' - y'| + |W(y', y) - W(x', y)| + |W(x', y) - W(x', x)|. \end{aligned}$$

The proof is completed by applying Theorem V.2 and Corollary V.3 and noting that $|x' - y'| \leq |x - y|$. □

Of course, Remark V.4 holds with obvious modifications for U and (5.61).

Remark V.7. The hypothesis that the curves Γ_i are continuous and strictly increasing cannot in general be weakened to continuous and non-decreasing. For example, consider the case where $\mu = \sigma = 1$ on $[0.5, 1] \times [0, 1]$ and $\mu = \sigma = 0$ on $[0, 0.5) \times [0, 1]$. Then we have

$$U_{\mu,\sigma}(x) = \begin{cases} 0, & \text{if } x \in [0, 0.5) \times [0, 1], \\ x_1 + x_2 - 0.5 + 2\sqrt{(x_1 - 0.5)x_2}, & \text{if } x \in [0.5, 1] \times [0, 1], \end{cases}$$

which has a discontinuity along the vertical line $\{x_1 = 0.5\}$, which would correspond to one of the curves Γ_i on which μ is discontinuous.

5.2.2 Hamilton-Jacobi-Bellman equation

In this section we show in Theorem V.8 that $U_{\mu+\mu_s,\sigma}$ is a viscosity solution of (P). In fact, (P) is the Hamilton-Jacobi-Bellman equation for the simple optimal control problem [10] defined by $U_{\mu+\mu_s,\sigma}$. For more information on the connection between Hamilton-Jacobi equations and optimal control problems, we refer the reader to [10].

Theorem V.8. *Suppose that $\mu, \sigma : [0, \infty)^2 \to [0, \infty)$ are Borel-measurable and bounded. Let $z \in [0, \infty)^2$ and set $V(x) = W_{\mu,\sigma}(z, x)$ for $x \in [z, \infty)$. If V is continuous then V satisfies*

$$\begin{cases} (V_{x_1} - \mu)_+ (V_{x_2} - \mu)_+ = \sigma^2 & \text{on } (z, \infty), \\ \min(V_{x_1}, V_{x_2}) \geq \mu & \text{on } (z, \infty), \end{cases} \tag{5.64}$$

in the viscosity sense.

Proof. The proof is based on a standard technique from optimal control theory for relating variational problems to Hamilton-Jacobi equations [10]. The proof is very similar to Theorem III.10. We will only sketch parts of the proof here.

The proof is based on the following dynamic programming principle

$$V(y) = \sup_{x \in \partial B_r(y) \,:\, x \leqq y} \Big\{ V(x) + W(x,y) \Big\}, \tag{5.65}$$

which holds for $y \in (z, \infty)$ and $r > 0$ small enough so that $B_r(y) \subset (z, \infty)$. The proof of (5.65) is very similar to the proof of Proposition V.5.

Let us now show that V is a viscosity supersolution of (5.64). Let $y \in (z, \infty)$ and let $p \in D^- V(y)$. As in the proof of Theorem III.10, we can use the dynamic programing principle to obtain

$$\sup_{a \in \mathbb{R}^2_+} \Big\{ - \big\langle p - \mu_*(y)(1,1), a \big\rangle + 2\sigma_*(y)\sqrt{a_1 a_2} \Big\} \leq 0. \tag{5.66}$$

Suppose now that $\sigma_*(y) = 0$. It follows from (5.66) that $\min(p_1, p_2) \geq \mu_*(y)$, and hence U is a viscosity supersolution of (5.64) at y. Consider now $\sigma_*(y) > 0$. Setting $a_1 = 1$ in (5.66) we have

$$\sup_{a_2 > 0} \Big\{ - (p_1 - \mu_*(y)) - (p_2 - \mu_*(y)) a_2 + 2\sigma_*(y)\sqrt{a_2} \Big\} \leq 0.$$

It follows that $p_2 > \mu_*(y)$. By a similar argument we have $p_1 > \mu_*(y)$. Now set

$$a_1 = \sqrt{\frac{p_2 - \mu_*(y)}{p_1 - \mu_*(y)}} \quad \text{and} \quad a_2 = \sqrt{\frac{p_1 - \mu_*(y)}{p_2 - \mu_*(y)}} \tag{5.67}$$

in (5.66) and simplify to find that

$$(p_1 - \mu_*(y))(p_2 - \mu_*(y)) \geq {\sigma^*}^2(y).$$

Therefore V is a viscosity supersolution of (5.64).

Let us now show that V is a viscosity subsolution of (5.64). Let $y \in (z, \infty)$ and let $p \in D^+ U(y)$. Utilizing the dynamic programing principle (5.65) again we have

$$\sup_{a \in \mathbb{R}^2_+ \,:\, a_1 a_2 = 1} \Big\{ - \big\langle p - \mu^*(y)(1,1), a \big\rangle + 2\sigma^*(y) \Big\} \geq 0. \tag{5.68}$$

If $\min(p_1, p_2) \leq \mu^*(y)$ then we have

$$(p_1 - \mu^*(y))_+(p_2 - \mu^*(y))_+ \leq 0 \leq {\sigma^*}^2,$$

and hence U is a viscosity subsolution of (5.64). Thus we may assume that $\min(p_1, p_2) > \mu^*(y)$, and hence

$$\limsup_{|a|\to\infty\,:\,a\in\mathbb{R}^2_+} -\big\langle p - \mu^*(y)(1,1), a\big\rangle + 2\sigma^*(y) = -\infty.$$

It follows that the supremum in (5.68) is attained at some $a^* \in \mathbb{R}^2_+$. Introducing a Lagrange multiplier $\lambda > 0$, the necessary conditions for a^* to be a maximizer of the constrained maximization problem (5.68) are

$$a_1^* = \lambda(p_2 - \mu^*(y)),\ a_2^* = \lambda(p_1 - \mu^*(y)), \quad \text{and } a_1^* a_2^* = 1.$$

It follows that $\lambda = (p_1 - \mu^*(y))^{-\frac{1}{2}}(p_2 - \mu^*(y))^{\frac{1}{2}}$ and a^* is given by (5.67). Substituting this into (5.68) we find that

$$(p_1 - \mu^*(y))(p_2 - \mu^*(y)) \leq {\sigma^*}^2(y),$$

and hence V is a viscosity subsolution of (5.64). □

Remark V.9. It follows from Theorem V.8 that $U = U_{\mu+\mu_s,\sigma}$ is a viscosity solution of (P) and satisfies

$$\min(U_{x_1}, U_{x_2}) \geq \mu \ \text{ on } \mathbb{R}^2_+ \tag{5.69}$$

in the viscosity sense. Indeed, we can simply apply Theorem V.8 with $\mu + \mu_s$ in place of μ and $z = 0$, in which case we have $U(x) = W_{\mu+\mu_s,\sigma}(0, x)$.

5.3 Comparison Principle

We study here the general Hamilton-Jacobi equation

$$\begin{cases} H(x, Du) = 0 & \text{on } (z, \infty), \\ u = \varphi & \text{on } \partial(z, \infty). \end{cases} \tag{5.70}$$

Here, $z \in [0, \infty)^d$, $\varphi : \partial(z, \infty) \to \mathbb{R}$ is continuous and Pareto-monotone, $H : \mathbb{R}^d_+ \times \mathbb{R}^d \to \mathbb{R}$ is the Hamiltonian, and $u : [z, \infty) \to \mathbb{R}$ is the unknown function. For simplicity of notation, we will set $z = 0$ throughout much of this section. The case where $z \neq 0$ follows by a simple translation argument.

Let us recall the assumptions from the comparison principles in Chapter III:

(H1) For every $x \in \mathbb{R}^d_+$, the mapping $H(x, \cdot) : \mathbb{R}^d \to \mathbb{R}$ is Pareto-monotone.

(H2) There exists a modulus of continuity m such that

$$H(x, p) - H(y, p) \leq m(|p||x - y| + |x - y|) \tag{5.71}$$

for all $p \in [0, \infty)^d$ and $x, y \in \mathbb{R}^d_+$.

The assumption (H1) is clearly satisfied by (P). We first show that the value function W is truncatable, as per Definition III.14.

Proposition V.10. *Suppose that $\mu, \sigma : [0, \infty)^2 \to [0, \infty)$ are Borel-measurable and bounded. Let $z \in [0, \infty)^2$ and define $V(x) = W_{\mu,\sigma}(z, x)$ for $x \in [z, \infty)$. If V is continuous then V is a truncatable viscosity solution of*

$$(V_{x_1} - \mu)_+(V_{x_2} - \mu)_+ = \sigma^2 \quad \text{on } (z, \infty). \tag{5.72}$$

Proof. It follows from Theorem V.8 that V is a viscosity solution of (5.72). We need only show that V is truncatable. Let $\xi \in (z, \infty)$, let $\chi : [0, \infty)^2 \to \{0, 1\}$ denote

the characteristic function of $[z,\xi]$, and set $\overline{V} = W_{\chi\cdot\mu,\chi\cdot\sigma}(z,\cdot)$. By the definition of $\overline{V}$ and χ we have $\overline{V}(x) = V(x) = V^\xi(x)$ for any $x \in [z,\xi]$. Let $x \in [z,\infty) \setminus [z,\xi]$, $\varepsilon > 0$, and let $\gamma \in \mathcal{A}$ with $\gamma(0) = z$, $\gamma(1) = x$ such that $\overline{V}(x) \le J_{\chi\cdot\mu,\chi\cdot\sigma}(\gamma) + \varepsilon$. Let γ^1 denote the portion of γ inside $[z,\xi]$, let γ^2 denote the remaining portion of γ, and reparametrize γ^1 and γ^2 so that $\gamma^1, \gamma^2 : [0,1] \to \mathbb{R}^2$. Letting $y = \gamma^1(1) \in [z,\xi]$ we have

$$\overline{V}(x) \le J_{\chi\cdot\mu,\chi\cdot\sigma}(\gamma^1) + J_{\chi\cdot\mu,\chi\cdot\sigma}(\gamma^2) + \varepsilon = J_{\mu,\sigma}(\gamma^1) + \varepsilon \le V(y) + \varepsilon.$$

Since $y \leqq x$ and $y \in [z,\xi]$, we also have $\overline{V}(x) \ge \overline{V}(y) = V(y)$. It follows that

$$\overline{V}(x) = \sup_{y\in[z,\xi]\,:\,y\leqq x} V(y).$$

By continuity of V, the supremum above is attained, and the maximizing argument of y is exactly $y = \pi_\xi(x)$—the projection of x onto $[0,\xi]$. Therefore we have $\overline{V}(x) = V(\pi_\xi(x))$. Since x is arbitrary, we see that $\overline{V} = V \circ \pi_\xi = V^\xi$, the ξ-truncation of V.

Since $V^\xi = V \circ \pi_\xi$ is continuous, it follows from Theorem V.8 that V^ξ is a viscosity solution of

$$(V_{x_1} - \chi\mu)_+(V_{x_2} - \chi\mu)_+ \le \chi\sigma^2 \ \text{ on } (z,\infty).$$

Since $0 \le \chi \le 1$ and $t \mapsto (p_1 - t)_+(p_2 - t)_+$ is monotone decreasing, it follows that V^ξ is viscosity subsolution of (5.72), which completes the proof. □

We now show that truncatability enjoys a useful L^∞-stability property.

Proposition V.11. *Let $z \in \mathbb{R}^2_+$ and for each $k \ge 1$ suppose that $u_k \in C([z,\infty))$ is a truncatable viscosity solution of*

$$H_k(x, Du_k) \le 0 \quad \text{on } (z,\infty). \tag{5.73}$$

If $u_k \to u$ locally uniformly, for some $u \in C([z,\infty))$, then u is a truncatable viscosity solution of

$$\underline{H}(x, Du) \leq 0 \quad \text{on } (z,\infty), \tag{5.74}$$

where

$$\underline{H}(x,p) := \liminf_{\substack{k\to\infty \\ y\to x}} H_k(y,p).$$

Proof. It is a standard result (see [27, Remark 6.3]) that u is a viscosity solution of (5.74). To see that u is truncatable: Fix $\xi \in (z,\infty)$, let u^ξ be the ξ-truncation of u, and let u_k^ξ be the ξ-truncation of u_k. Since u_k is truncatable, we have that u_k^ξ is a viscosity solution of (5.73) for every k. Furthermore, we have $u_k^\xi \to u^\xi$ locally uniformly, and therefore u^ξ is a viscosity solution of (5.74). Thus u is truncatable. □

Note that Proposition V.11 was used implicitly in the proof of Lemma III.18.

For the remainder of the section we set

$$H(x,p) = (p_1 - \mu(x))_+(p_2 - \mu(x))_+ - \sigma^2(x). \tag{5.75}$$

Our aim now is to apply the comparison principles from Chapter III—namely Theorems III.11 and III.16—to obtain a comparison principle, and a perturbation result, for the Hamilton-Jacobi equation (P). First we need to show that (H2) and $(\text{H3})_{\mathcal{O}}$ are satisfied by H given in (5.75).

Proposition V.12. *Suppose that $\mu, \sigma : [0,\infty)^2 \to [0,\infty)$, and let H be given by (5.75). Then for any $x, y \in \mathbb{R}^2_+$*

$$H(y,p) - H(x,p) \leq 2|p|(\mu(x) - \mu(y))_+ + \sigma^2(x) - \sigma^2(y). \tag{5.76}$$

Proof. Let $p \in [0,\infty)^2$, and set $h(t) = (p_1 - t)_+(p_2 - t)_+$ so that

$$H(x,p) = h(\mu(x)) - \sigma^2(x).$$

Suppose first that $\mu(y) < \min(p_1, p_2)$. Since h is convex, we have

$$h(\mu(x)) - h(\mu(y)) \geq h'(\mu(y))(\mu(x) - \mu(y)) = -(p_1 + p_2 - 2\mu(y))(\mu(x) - \mu(y)).$$

Since $p_1 + p_2 - 2\mu(y) \geq 0$ we have

$$\begin{aligned} h(\mu(y)) - h(\mu(x)) &\leq (p_1 + p_2 - 2\mu(y))(\mu(x) - \mu(y)) \\ &\leq (p_1 + p_2 - 2\mu(y))(\mu(x) - \mu(y))_+ \\ &\leq (p_1 + p_2)(\mu(x) - \mu(y))_+. \end{aligned}$$

Therefore we have

$$h(\mu(y)) - h(\mu(x)) \leq 2|p|(\mu(x) - \mu(y))_+. \tag{5.77}$$

If $\mu(y) \geq \min(p_1, p_2)$ then we have $h(\mu(y)) = 0 \leq h(\mu(x))$, and hence (5.77) holds.

□

Remark V.13. It follows from Proposition V.12 that H satisfies (H2) if μ and σ^2 are globally Lipschitz continuous on $\mathbb{R}^2_+$.

Corollary V.14. *Suppose that μ and σ^2 are non-negative and globally Lipschitz continuous on $\mathbb{R}^2_+$. Let $u \in USC([0,\infty)^2)$ be a viscosity solution of*

$$(u_{x_1} - \mu)_+(u_{x_2} - \mu)_+ \leq \sigma^2 \quad \text{on } \mathbb{R}^2_+, \tag{5.78}$$

and let $v \in LSC([0,\infty)^2)$ be a Pareto-monotone viscosity solution of

$$(v_{x_1} - \mu)_+(v_{x_2} - \mu)_+ \geq \sigma^2 \quad \text{on } \mathbb{R}^2_+. \tag{5.79}$$

Furthermore, suppose that

$$\big\{x \in \mathbb{R}^2_+ \,:\, \mu(x) = 0\big\} \supset \big\{x \in \mathbb{R}^2_+ \,:\, \sigma(x) = 0\big\}. \tag{5.80}$$

Then $u \leq v$ on $\partial\mathbb{R}^2_+$ implies $u \leq v$ on $\mathbb{R}^2_+$.

Proof. We claim that

$$\min(v_{x_1}, v_{x_2}) \geq \mu \quad \text{on } \mathbb{R}^2_+, \tag{5.81}$$

in the viscosity sense. To see this, let $x \in \mathbb{R}^2_+$ and let $p \in D^- v(x)$. Then we have

$$(p_1 - \mu(x))_+ (p_2 - \mu(x))_+ \geq \sigma(x)^2.$$

If $\sigma(x) > 0$, then we must have $\min(p_1, p_2) \geq \mu(x)$ as desired. If $\sigma(x) = 0$, then by (5.80) we have $\mu(x) = 0$, and we have $\min(p_1, p_2) \geq 0 = \mu(x)$ by virtue of the monotonicity of v.

Let $a > 0$ and set $\overline{v}(x) = v(x) + \sqrt{a}(x_1 + x_2)$. By (5.79) and (5.81) we see that $\overline{v}$ is a viscosity solution of

$$(\overline{v}_{x_1} - \mu)_+ (\overline{v}_{x_2} - \mu)_+ \geq \sigma^2 + a \quad \text{on } \mathbb{R}^2_+.$$

By Proposition V.12 and Remark V.13 we see that (H1) and (H2) are satisfied. Therefore we can apply Theorem III.11 to find that $u \leq \overline{v}$. Sending $a \to 0$ completes the proof. □

Recall that μ and σ^2 are not independent functions in the DLPP problem, even though we have treated them as such for much of the analysis. From this point on, we will need to recall their relationship, as it is important for proving uniqueness in (P). Specifically, we need to assume that μ and σ^2 satisfy (F3) for the *same* choice of ζ at each $x \in \Gamma_i$. When this holds, we say that μ and σ^2 *simultaneously* satisfy (F3). Since $\sigma = \mu$ for exponential DLPP and $\sigma = \sqrt{\mu(1+\mu)}$ for geometric DLPP, σ is always a monotone increasing function of μ, and hence μ and σ^2 simultaneously satisfy (F3) in both cases.

Proposition V.15. *Let μ and σ^2 simultaneously satisfy (F1) and (F3). Then H given by* (5.75) *satisfies (H3)$_\mathcal{O}$ with $\mathcal{O} = \mathbb{R}^2_+ \setminus \overline{\Omega}$.*

Proof. Let $\xi \in \mathcal{O}$. If $\xi \in \Omega_i$, then we can choose ε_ξ small enough so that $B_{2\varepsilon_\xi}(\xi) \subset \Omega_i$. By Proposition V.12 we see that any choice for $\mathbf{v}_\xi$ will suffice since μ and σ^2 are Lipschitz with constant C_{lip} when restricted to Ω_i.

If $\xi \in \Gamma_i$ for some i, then let ζ be as given in (F3). Assume for now that $\zeta = -1$, and set $\mathbf{v}_\xi = (1,-1)/\sqrt{2}$. Let $\varepsilon_\xi > 0$ be less than half the value of ε from (F3), and then choose $\varepsilon_\xi > 0$ smaller, if necessary, so that $B_{2\varepsilon_\xi}(\xi)$ has an empty intersection with Γ and all other Γ_j, and $\varepsilon_\xi \leq 1/2$. Let μ_i and σ_i^2 denote the Lipschitz extensions of $\mu|_{\Omega_i}$ and $\sigma^2|_{\Omega_i}$ to $\overline{\Omega_i}$, respectively, and make the same definitions for μ_{i-1} and σ_{i-1}^2. Then (F3) implies that $\mu_i \geq \mu_{i-1}$ and $\sigma_i^2 \geq \sigma_{i-1}^2$ on $B_{2\varepsilon_\xi}(\xi) \cap \Gamma_i$. Furthermore, since μ and σ^2 are upper semicontinuous, we have $\mu = \mu_i$ and $\sigma = \sigma_i$ on $B_{2\varepsilon_\xi}(\xi) \cap \Gamma_i$.

Let $y \in B_{\varepsilon_\xi}(\xi)$, $\varepsilon < \varepsilon_\xi$, $p \in \mathbb{R}^2$, and $\mathbf{v} \in \mathbb{S}^{d-1}$ with $|\mathbf{v} - \mathbf{v}_\xi| < \varepsilon_\xi$. If $y + \varepsilon\mathbf{v} \in \overline{\Omega_i}$, then since Γ_i is monotone, $|\mathbf{v} - \mathbf{v}_\xi| \leq \frac{1}{2}$, and $y \in B_{2\varepsilon_\xi}(\xi)$, we must have that $y \in \Omega_i$. Since μ_i and σ_i^2 are Lipschitz on $\overline{\Omega_i} \cap B_{2\varepsilon_\xi}(\xi)$, we can invoke Proposition V.12 to show that $(\text{H3})_\mathcal{O}$ holds.

Now suppose that $y + \varepsilon\mathbf{v} \in \Omega_{i-1}$. If $y \in \Omega_{i-1}$, then $(\text{H3})_\mathcal{O}$ holds as before, so assume that $y \in \overline{\Omega_i}$. Let $\varepsilon' > 0$ such that $y + \varepsilon'\mathbf{v} \in \Gamma_i$. Then we have

$$\begin{aligned}
\mu(y+\varepsilon\mathbf{v}) - \mu(y) &= \mu_{i-1}(y+\varepsilon\mathbf{v}) - \mu_i(y+\varepsilon'\mathbf{v}) + \mu_i(y+\varepsilon'\mathbf{v}) - \mu_i(y) \\
&\leq \mu_{i-1}(y+\varepsilon\mathbf{v}) - \mu_{i-1}(y+\varepsilon'\mathbf{v}) + \mu_i(y+\varepsilon'\mathbf{v}) - \mu_i(y) \\
&\leq 2C_{lip}\varepsilon,
\end{aligned}$$

where we used the fact that $\mu_i \geq \mu_{i-1}$ on $\Gamma_i \cap B_{2\varepsilon_\xi}(\xi)$. We have an identical estimate for σ^2, and the proof is completed by invoking Proposition V.12. □

Corollary V.16. *Let μ and σ^2 simultaneously satisfy (F1) and (F3). Let $u \in C([0,\infty)^2)$ be a truncatable viscosity solution of* (5.78)*, let $v \in C([0,\infty)^2)$ be a Pareto-monotone viscosity solution of* (5.79)*, and suppose that* (5.80) *holds. Then*

$u \le v$ on $\Omega \cup \partial\mathbb{R}^2_+$ implies $u \le v$ on $\mathbb{R}^2_+$.

The proof of Corollary V.16 is similar to Corollary V.14.

We now prove an important perturbation result. Roughly speaking, it says that if we smooth out the macroscopic mean μ and variance σ (i.e., remove the discontinuities), then the resulting change in the value function W is uniformly small. This result is used in the proof of our main result, Theorem V.1. The proof relies on the uniqueness of truncatable viscosity solutions of (P) (Theorem III.16 and Corollary V.16), and the result can then be used to prove a comparison principle for (P) without the truncatability assumption (see Theorem V.20).

Theorem V.17. *Let μ and σ^2 satisfy* (5.80) *and simultaneously satisfy (F1), (F3). Let $\mu_k, \sigma_k^2 \in C^{0,1}([0,\infty)^2)$ satisfy (F1*) with $\theta = \frac{1}{k}$. Furthermore suppose that*

$$\mu_*(x) \le \liminf_{\substack{k\to\infty \\ y\to x}} \mu_k(y), \quad \mu^*(x) \ge \limsup_{\substack{k\to\infty \\ y\to x}} \mu_k(y), \tag{5.82}$$

and

$$\sigma_*(x) \le \liminf_{\substack{k\to\infty \\ y\to x}} \sigma_k(y), \quad \sigma^*(x) \ge \limsup_{\substack{k\to\infty \\ y\to x}} \sigma_k(y), \tag{5.83}$$

for all $x \in \mathbb{R}^2_+$. Then for every $z \in [0,\infty)^2$ we have

$$W_{\mu_k,\sigma_k}(z,\cdot) \longrightarrow W_{\mu,\sigma}(z,\cdot) \quad \textit{locally uniformly on } [z,\infty).$$

Proof. For simplicity, let us set $V_k(x) = W_{\mu_k,\sigma_k}(z,x)$ and $V(x) = W_{\mu,\sigma}(z,x)$ for $x \in [z,\infty)$. Since $\mu_k, \sigma_k^2 \in C^{0,1}([0,\infty)^2)$, we can apply Theorem V.2 with $\theta = 0$ to find that V_k is continuous on $[z,\infty)$. We can apply Theorem V.2 again with $\theta = 1/k$ to show that for every $R > \max(z_1, z_2)$, there exists $C = C(C_{lip}, \|\mu\|_\infty, \|\sigma\|_\infty, R)$ and a modulus of continuity ω such that

(5.84)

$$|V_k(x) - V_k(y)| \le C(\sqrt{|x-y|} + \omega(|x-y|) + \omega(k^{-1})) \text{ for all } x, y \in [z_1, R] \times [z_2, R].$$

This approximate Hölder estimate is sufficient to apply a slightly modified version of the Arzelà-Ascoli theorem (see Theorem IV.6). Therefore, by passing to a subsequence if necessary, there exists $v \in C([z,\infty))$ such that $V_k \to v$ locally uniformly on $[z,\infty)$. By Proposition V.10, V_k is a Pareto-monotone truncatable viscosity solution of

$$(V_{k,x_1} - \mu_k)_+(V_{k,x_2} - \mu_k)_+ = {\sigma_k}^2 \text{ on } (z,\infty). \tag{5.85}$$

Since $V_k \to v$ locally uniformly and (5.82)-(5.83) hold, we can apply Proposition V.11, and classical results from the viscosity solution theore [27], to find that v is a Pareto-monotone truncatable viscosity solution of

$$(v_{x_1} - \mu)_+(v_{x_2} - \mu)_+ = \sigma^2 \text{ on } (z,\infty). \tag{5.86}$$

We claim that $v = V$ on $\partial(z,\infty)$. To see this: Let $x \in \partial(z,\infty)$, hence $x_i = z_i$ for some i. Without loss of generality, assume that $x_1 = z_1$. Then by (5.36) and Fatou's lemma we have

$$\begin{aligned} v(x) = \lim_{k\to\infty} V_k(x) &= \lim_{k\to\infty} \int_{z_2}^{x_2} \mu_k(z_1,t)\,dt \\ &\le \int_{z_2}^{x_2} \limsup_{k\to\infty} \mu_k(z_1,t)\,dt \\ &\le \int_{z_2}^{x_2} \mu(z_1,t)\,dt = V(x), \end{aligned}$$

where the last line follows from (5.82) and the fact that μ is upper semicontinuous. By a similar argument with Fatou's lemma we have

$$v(x) \ge \int_{z_2}^{x_2} \mu_*(z_1,t)\,dt. \tag{5.87}$$

Notice that (F1) implies that $\mu_* = \mu$ on Ω_i for all i and on Ω. Hence, all the points $x \in [0,\infty)^2$ for which $\mu_*(x) \neq \mu(x)$ are contained in $\cup_{i\in\mathbb{Z}}\Gamma_i \cup \Gamma$. Since the curves Γ_i

are strictly increasing and Γ is strictly decreasing, the curve $t \mapsto (z_1, t)$ for $t \in [z_2, x_2]$ has a finite number of intersections with $\cup_{i\in\mathbb{Z}}\Gamma_i \cup \Gamma$. It follows that

$$v(x) \overset{(5.87)}{\geq} \int_{z_2}^{x_2} \mu_*(z_1, t)\, dt = \int_{z_2}^{x_2} \mu(z_1, t)\, dt = V(x),$$

and hence $v(x) = V(x)$, which establishes the claim.

By Proposition V.15, H given by (5.75) satisfies $(\mathrm{H3})_{\mathcal{O}}$ for $\mathcal{O} = \mathbb{R}^2_+ \setminus \overline{\Omega}$. By (F1*) and (5.35) we have $V_k(x) = 0$ for $x \in \overline{\Omega_\theta} \cap [z, \infty)$, and hence $v(x) = 0$ for $x \in \overline{\Omega} \cap [z, \infty)$. Similarly, we have that $V(x) = 0$ for $x \in \overline{\Omega} \cap [z, \infty)$. It follows that $v = V$ on $[z, \infty) \setminus \mathcal{O}$. Therefore we can apply a translated form of Corollary V.16 to find that $v = V$ on $[z, \infty)^2$. □

Remark V.18. Sequences generated by inf- and sup-convolutions of μ and σ^2 satisfy the hypotheses of Theorem V.17. Recall that the sup-convolution of $\mu : [0, \infty)^2 \to \mathbb{R}$ is defined by

$$\mu^k(x) = \sup_{y\in[0,\infty)^2} \Big\{\mu(y) - k|x - y|\Big\}, \tag{5.88}$$

and the inf-convolution by $\mu_k := -(-\mu)^k$.

Corollary V.19. *Let μ and σ^2 simultaneously satisfy (F1), (F3) and* (5.80)*, let $\mu_k, \sigma_k^2 \in C^{0,1}([0, \infty)^2)$ satisfy (F1*) with $\theta = \frac{1}{k}$, and let μ_s satisfy (F2). If* (5.82)–(5.83) *hold for all $x \in \mathbb{R}^2_+$ then*

$$U_{\mu_k+\mu_s,\sigma_k} \longrightarrow U_{\mu+\mu_s,\sigma} \quad \textit{locally uniformly on } [0, \infty)^2.$$

Proof. Fix $y \in [0, \infty)^2$. By Proposition V.5 we have

$$U_{\mu_k+\mu_s,\sigma_k}(y) = \max_{x\in\partial\mathbb{R}^2_+ \,:\, x\leqq y} \Big\{U_{\mu_k+\mu_s,\sigma_k}(x) + W_{\mu_k,\sigma_k}(x, y)\Big\}, \tag{5.89}$$

and

$$U_{\mu+\mu_s,\sigma}(y) = \max_{x\in\partial\mathbb{R}^2_+ \,:\, x\leqq y} \Big\{U_{\mu+\mu_s,\sigma}(x) + W_{\mu,\sigma}(x, y)\Big\}. \tag{5.90}$$

Arguing by symmetry, it follows from Theorem V.17 that

$$W_{\mu_k,\sigma_k}(\cdot,y) \longrightarrow W(\cdot,y) \text{ uniformly on } [0,y]. \tag{5.91}$$

It follows from (5.33) and a similar argument as in Theorem V.17 that $U_{\mu_k+\mu_s,\sigma_k}(x) \to U_{\mu+\mu_s,\sigma}(x)$ for any $x \in \partial\mathbb{R}^2_+$. By the Arzelà-Ascoli Theorem we find that

$$U_{\mu_k+\mu_s,\sigma_k} \longrightarrow U_{\mu+\mu_s,\sigma} \text{ uniformly on } [0,y] \cap \partial\mathbb{R}^2_+. \tag{5.92}$$

Combining (5.89)–(5.92), we have that $U_{\mu_k+\mu_s,\sigma_k}(y) \to U_{\mu+\mu_s,\sigma}(y)$. Locally uniform convergence follows again from the Arzelà-Ascoli Theorem. □

Theorem V.20. *Let μ and σ^2 simultaneously satisfy (F1), (F3) and* (5.80)*, and let μ_s satisfy (F2). Let $u \in C([0,\infty)^2)$ be a viscosity solution of* (5.78) *and let $v \in C([0,\infty)^2)$ be a Pareto-monotone viscosity solution of* (5.79)*. Then if $u \le \varphi \le v$ on $\partial\mathbb{R}^2_+$, where φ is given in the statement of Theorem V.1, then $u \le v$ on $\mathbb{R}^2_+$.*

Proof. Let $\mu^k, \sigma^{2,k}$ and μ_k, σ_k^2 be the sup- and inf-convolutions of μ and σ^2 as defined in Remark V.18, respectively. To simplify notation, let us write $U^k := U_{\mu^k+\mu_s,\sigma^k}$, $U_k := U_{\mu_k+\mu_s,\sigma_k}$, and $U := U_{\mu+\mu_s,\sigma}$. By definition we have $U_k \le U \le U^k$, and by Corollary V.19 and Remark V.18 we have $U_k, U^k \to U$ locally uniformly on $[0,\infty)^2$ as $k \to \infty$.

Since $\mu_k \le \mu$ and $\sigma_k \le \sigma$ we have that v is a viscosity solution of

$$(v_{x_1} - \mu_k)_+(v_{x_2} - \mu_k)_+ \ge {\sigma_k}^2 \text{ on } \mathbb{R}^2_+.$$

By Theorem V.8, U_k is a viscosity solution of

$$(U_{k,x_1} - \mu_k)_+(U_{k,x_2} - \mu_k)_+ = {\sigma_k}^2 \text{ on } \mathbb{R}^2_+.$$

Furthermore, we have $U_k = \varphi_k \le \varphi \le v$ on $\partial\mathbb{R}^2_+$ where $\varphi_k(x) = (x_1+x_2)\int_0^1 \mu_k(tx) + \mu_s(tx)\,dt$. Since μ_k and σ_k^2 are globally Lipschitz we can apply Corollary V.14 to

obtain $U_k \leq v$. Sending $k \to \infty$ we have $U \leq v$. By a similar argument we can prove that $u \leq U$, which completes the proof. □

5.4 Proof of main result

In this section we give the proof of our main result, Theorem V.1. We first have a preliminary convergence result on the interior $(0,\infty)^2$, which we later adapt to account for the boundary source μ_s. For $N \geq 1$ we define

$$w_N(x,y) := L\Big(\lfloor Nx \rfloor + \mathbf{1}_x; \lfloor Ny \rfloor\Big), \tag{5.93}$$

where

$$\mathbf{1}_x = \big(\chi_{\{x_1=0\}}, \chi_{\{x_2=0\}}\big), \tag{5.94}$$

and L is defined in (5.6).

Lemma V.21. *Assume μ satisfies (F1) and (F3). Suppose that the weights $X(i,j)$ satisfy* (5.7) *and are either all exponential, or all geometric random variables. In the exponential case, set $\sigma = \mu$, and in the geometric case, set $\sigma = \sqrt{\mu(1+\mu)}$. Then for every $y \in (0,\infty)^2$ we have*

$$\frac{1}{N} w_N(\cdot, y) \longrightarrow W_{\mu,\sigma}(\cdot, y) \quad \textit{uniformly on } [0,y],$$

with probability one.

Proof. Let $y \in (0,\infty)^2$. Let μ^k and μ_k be the sup- and inf-convolutions of μ, defined in Remark V.18. In the exponential case, set $\sigma^k = \mu^k$ and $\sigma_k = \mu_k$, and in the geometric case, set $\sigma^k = \sqrt{\mu^k(1+\mu^k)}$ and $\sigma_k = \sqrt{\mu_k(1+\mu_k)}$. To simplify notation, let us also set $W^k := W_{\mu^k,\sigma^k}$, $W_k := W_{\mu_k,\sigma_k}$, and $W := W_{\mu,\sigma}$, and note that $W_k \leq$

$W \leq W^k$. Notice that by the definition of σ, we have that (5.80) holds for both the exponential and geometric cases. We can therefore invoke Theorem V.17 to find that

$$W_k(x,y) \longrightarrow W(x,y) \text{ and } W^k(x,y) \longrightarrow W(x,y) \text{ for all } x \in [0,y]. \tag{5.95}$$

Let $N \geq 1$. In the exponential case, for $(i,j) \in \mathbb{N}^2$ let $X^k(i,j)$ be independent and exponentially distributed with parameter $\lambda = \mu^k(iN^{-1}, jN^{-1})$, and let $X_k(i,j)$ be independent and exponentially distributed with parameter $\lambda = \mu_k(iN^{-1}, jN^{-1})$. In the geometric case, for $(i,j) \in \mathbb{N}^2$ let $X^k(i,j)$ be independent and geometrically distributed with parameter $q = (1+\mu^k(iN^{-1}, jN^{-1}))^{-1}$, and let $X_k(i,j)$ be independent and geometrically distributed with parameter $q = (1+\mu_k(iN^{-1}, jN^{-1}))^{-1}$. In either case, set

$$L_k(M,N;Q,P) = \max_{p \in \Pi_{(M,N),(Q,P)}} \sum_{(i,j)\in p} X_k(i,j), \tag{5.96}$$

$$L^k(M,N;Q,P) = \max_{p \in \Pi_{(M,N),(Q,P)}} \sum_{(i,j)\in p} X^k(i,j), \tag{5.97}$$

and set

(5.98)

$$w_{k,N}(x,y) := L_k\Big(\lfloor Nx \rfloor + \mathbf{1}_x; \lfloor Ny \rfloor\Big), \quad \text{and} \quad w^k_N(x,y) := L^k\Big(\lfloor Nx \rfloor + \mathbf{1}_x; \lfloor Ny \rfloor\Big).$$

We can define $X_k(i,j)$ and $X^k(i,j)$ on the same probability space as $X(i,j)$ in such a way that $X_k(i,j) \leq X(i,j) \leq X^k(i,j)$ for all $(i,j) \in \mathbb{N}^2$ with probability one. We therefore have $w_{k,N} \leq w_N \leq w^k_N$ with probability one. Since μ_k, σ_k, μ^k, and σ^k are continuous on $[0,\infty)^2$, we can invoke Theorem [81, Theorem 1] to find that

$$\frac{1}{N} w_{k,N}(x,y) \longrightarrow W_k(x,y) \quad \text{and} \quad \frac{1}{N} w^k_N(x,y) \longrightarrow W^k(x,y),$$

with probability one, for fixed $x \in [0,y]$. We should note that [81, Theorem 1] as stated applies only to exponential DLPP, but one can check that the proof holds for

geometric DLPP as well, with minor modifications. It follows that for every $k \geq 1$ we have

$$W_k(x,y) \leq \liminf_{N\to\infty} \frac{1}{N} w_N(x,y) \leq \limsup_{N\to\infty} \frac{1}{N} w_N(x,y) \leq W^k(x,y),$$

with probability one. Sending $k \to \infty$ and recalling (5.95) we have for every $x \in [0,y]$ that

$$\frac{1}{N} w_N(x,y) \longrightarrow W(x,y) \quad \text{with probability one.} \tag{5.99}$$

Uniform convergence follows from the fact that $x \mapsto w_N(x,y)$ and $x \mapsto W(x,y)$ are monotone decreasing (in the Pareto sense) and $x \mapsto W(x,y)$ is uniformly continuous on $[0,y]$; the proof is similar to Theorem III.2. □

We now have the proof of Theorem V.1.

Proof. By the law of large numbers we have

$$\frac{1}{N} L(0; \lfloor N\cdot \rfloor) \longrightarrow U = \varphi \tag{5.100}$$

locally uniformly on $\partial\mathbb{R}^2_+$ with probability one.

Let $y \in \mathbb{R}^2_+$. From the definition of L we have the following dynamic programming principle

$$L(0; \lfloor Ny \rfloor) = \max_{x \in \mathbb{R}^2_+ \,:\, x \leqq y} \Big\{ L(0; \lfloor Nx \rfloor) + w_N(x,y) \Big\}. \tag{5.101}$$

Combining Lemma V.21, Proposition V.5, and (5.100), we can pass to the limit in (5.101) to obtain

$$\frac{1}{N} L(0; \lfloor Ny \rfloor) \longrightarrow \max_{x \in \mathbb{R}^2_+ \,:\, x \leqq y} \Big\{ U(x) + W(x,y) \Big\} = U(y),$$

with probability one. As in Lemma V.21, locally uniform convergence follows from the monotonicity of U and $x \mapsto N^{-1} L(0; \lfloor Nx \rfloor)$, along with the uniform continuity given by Theorem V.2; the proof is similar to Theorem III.2. □

5.5 Numerical scheme

We present here a fast numerical scheme for computing the viscosity solution U of (P). The scheme is a minor modification of the scheme used in Chapter IV. Since information propagates along coordinate axes in the definition of the variational problem (5.35) for U, it is natural to consider using backward difference quotients to approximate (P). Letting $U^h_{i,j}$ denote the numerical solution on the grid $h\mathbb{N}_0^2$ of spacing h, we have

$$\Big(U^h_{i,j} - U^h_{i-1,j} - h\mu_{i,j}\Big)_+ \Big(U^h_{i,j} - U^h_{i,j-1} - h\mu_{i,j}\Big)_+ = h^2\sigma^2_{i,j}, \tag{5.102}$$

where $\mu_{i,j} = \mu(hi, hj) + \mu_s(hi, hj)$ and $\sigma_{i,j} = \sigma(hi, hj)$. Given $U^h_{i-1,j}$ and $U^h_{i,j-1}$, we can solve (5.102) for $U^h_{i,j} \geq \max(U^h_{i-1,j} + h\mu_{i,j}, U^h_{i,j-1} + h\mu_{i,j})$ via the quadratic formula to obtain

$$U^h_{i,j} = \frac{1}{2}\Big(U^h_{i-1,j} + U^h_{i,j-1}\Big) + h\mu_{i,j} + \frac{1}{2}\sqrt{\Big(U^h_{i-1,j} - U^h_{i,j-1}\Big)^2 + 4h^2\sigma^2_{i,j}}, \tag{5.103}$$

for $i, j \geq 1$. The choice of the positive root in (5.103) reflects the monotonicity of the scheme, and ensures that it captures the viscosity solution of (P). When $i = 0$ or $j = 0$, we recall the boundary condition (5.33) to obtain

$$U^h_{0,j} = U^h_{0,j-1} + h\mu_{0,j} \quad \text{and} \quad U^h_{i,0} = U^h_{i-1,0} + h\mu_{i,0}. \tag{5.104}$$

Notice that when $i = 0$, if we set $U^h_{-1,j} = 0$ and $\sigma_{i,j} = 0$ in (5.103), then (5.103) and (5.104) are equivalent. In fact, even when $\sigma_{i,j} \neq 0$, (5.103) and (5.104) are asymptotically equivalent as $h \to 0$ provided $U^h_{0,j} \gg h$. The same observations hold when $j = 0$ if we set $U^h_{i,-1} = 0$. Thus, to account for the boundary condition in (P), we can simply set

$$U^h_{i,j} = 0 \quad \text{for } (i, j) \notin \mathbb{N}_0^2, \tag{5.105}$$

and compute $U^h_{i,j}$ via (5.103) for all $(i,j) \in \mathbb{N}_0^2 \cap [0,R]^2$, for any $R > 0$. In summary, we propose the following numerical scheme for approximating viscosity solutions of (P):

$$\text{(S)} \begin{cases} U^h_{i,j} = \frac{1}{2}\big(U^h_{i-1,j} + U^h_{i,j-1}\big) + h\mu_{i,j} + \frac{1}{2}\sqrt{\big(U^h_{i-1,j} - U^h_{i,j-1}\big)^2 + 4h^2\sigma^2_{i,j}}, & \text{if } (i,j) \in \mathbb{N}_0^2 \\ U^h_{i,j} = 0, & \text{otherwise.} \end{cases}$$

Note that we can visit the grid points in any sweeping pattern that visits $(i-1,j)$ and $(i,j-1)$ before (i,j), which reflects the cone of influence in the percolation problem. This scheme requires visiting each grid point exactly once and hence has linear complexity.

Our first result guarantees that the simple boundary condition in (S) agrees with the boundary condition in (P) as $h \to 0$.

Lemma V.22. *Let $U^h_{i,j}$ satisfy the scheme (S) and suppose that $\sigma_{i,j}$ is bounded by M for all $(i,j) \in \mathbb{N}_0^2 \cap \partial\mathbb{R}^2_+$. If $i,j \le h^{-1}R$ then there exists a constant $C > 0$ such that*

$$\left| U^h_{i,0} - h\sum_{k=0}^{i} \mu_{k,0} \right|, \left| U^h_{0,j} - h\sum_{k=0}^{j} \mu_{0,k} \right| \le C(1 + RM^2)\sqrt{h}. \tag{5.106}$$

Proof. Let us give the proof for $i = 0$. The case of $j = 0$ is similar. Define

$$J := \sup\Big\{ j \ge 0 \,:\, U^h_{0,j} \le \sqrt{h} \Big\}.$$

For $j \ge J$ it follows from the scheme (S) and a Taylor expansion that

$$U^h_{0,j} = \frac{1}{2}U^h_{0,j-1} + h\mu_{0,j} + \frac{1}{2}U^h_{0,j-1} + O\Big(h^{\frac{3}{2}}M^2\Big) = U^h_{0,j-1} + h\mu_{0,j} + O\Big(h^{\frac{3}{2}}M^2\Big). \tag{5.107}$$

Iterating (5.107) we have

$$U^h_{0,j} = h\left(\sum_{k=J+1}^{j} \mu_{0,k}\right) + U^h_{0,J} + O\Big(h^{\frac{3}{2}}jM^2\Big) = h\left(\sum_{k=J+1}^{j} \mu_{0,k}\right) + O\Big(\sqrt{h} + jh^{\frac{3}{2}}M^2\Big).$$

Since $j \leq h^{-1}R$ we have

$$U^h_{0,j} \leq h\left(\sum_{k=0}^{j} \mu_{0,k}\right) + O\Big(\big(1 + RM^2\big)\sqrt{h}\Big). \tag{5.108}$$

Noting the equivalence of (5.103) and (5.104) when $\sigma_{0,j} = 0$, we can set $\sigma_{0,j} = 0$ in (5.103) and iterate as before to obtain

$$U^h_{0,j} \geq h\left(\sum_{k=0}^{j} \mu_{0,k}\right).$$

Combining this with (5.108) completes the proof. □

Theorem V.23. *Suppose that μ and σ^2 are non-negative, globally Lipschitz continuous on $[0,\infty)^2$ and satisfy* (5.80)*, and let μ_s satisfy (F2). For $h > 0$ let $U^h(x) = U_{\lfloor h^{-1}x_1\rfloor,\lfloor h^{-1}x_2\rfloor}$ denote the extension of the numerical solution $U^h_{i,j}$ of (S) to $[0,\infty)^2$. Then we have*

$$U^h \longrightarrow U \quad \textit{locally uniformly on } [0,\infty)^2, \tag{5.109}$$

where U is the unique Pareto-monotone viscosity solution of (P).

Proof. The proof follows the standard framework outlined by Barles and Souganidis [12]. This general theory guarantees convergence of any scheme that is monotone, stable, and consistent, provided the PDE enjoys strong uniqueness—a comparison principle for semicontinuous sub- and supersolutions. Corollary V.14 is the required strong uniqueness result, and it is easy to see that the scheme (5.102) is both monotone and consistent. Indeed, for any $\psi \in C^1([0,\infty)^2)$ we have

$$\frac{1}{h^2}\Big(\psi(x) - \psi(x - he_1) - h\mu(x)\Big)_+\Big(\psi(x) - \psi(x - he_2) - h\mu(x)\Big)_+ \\ \longrightarrow \Big(\psi_{x_1}(x) - \mu(x)\Big)_+\Big(\psi_{x_2}(x) - \mu(x)\Big)_+,$$

as $h \to 0$, which is the required consistency. To show monotonicity, let $u, v : [0,\infty)^2$ such that $u(x) = v(x)$ and $u \le v$. Then we have

$$\begin{aligned}\Big(v(x) - v(x - he_1) - h\mu(x)\Big)_+ &\Big(v(x) - v(x - he_2) - h\mu(x)\Big)_+ \\ &= \Big(u(x) - v(x - he_1) - h\mu(x)\Big)_+ \Big(u(x) - v(x - he_2) - h\mu(x)\Big)_+ \\ &\le \Big(u(x) - u(x - he_1) - h\mu(x)\Big)_+ \Big(u(x) - u(x - he_2) - h\mu(x)\Big)_+,\end{aligned}$$

where the last line follows from the monotonicity of $t \mapsto (p_1 - t)_+(p_2 - t)_+$.

Therefore, to complete the proof, we need to show that the scheme is stable, and that the boundary condition is satisfied. Stability refers to a bound on U^h, independent of h. By Lemma V.22, (F2), and the continuity of μ, we have that

$$U^h \longrightarrow \varphi \quad \text{locally uniformly on } \partial\mathbb{R}^2_+ \ \text{ as } h \to 0, \tag{5.110}$$

where $\varphi(x) = (x_1 + x_2)\int_0^1 \mu(tx) + \mu_s(tx)\,dt$, which verifies the boundary condition.

Stability follows from a comparison principle for (S), and is similar to Lemma IV.3. We give the argument here for completeness. Let

$$V(x) = \|\mu + \mu_s\|_\infty(x_1 + x_2) + 2\|\sigma\|_\infty\sqrt{x_1 x_2} + 1.$$

We claim that $U^h(x) \le V(x)$. To see this, suppose to the contrary that $U^h(x) > V(x)$ for some $x \in [0, R]^2$, $R > 0$. First note that

$$\varphi(x) \le (x_1 + x_2)\|\mu + \mu_s\|_\infty = V(x) - 1,$$

for $x \in \partial\mathbb{R}^2_+$. Therefore, by (5.110), we have that $U^h \le V - \frac{1}{2}$ on $[0, R]^2 \cap \partial\mathbb{R}^2_+$ for h small enough. Therefore, there exists $z \in [h, R]^2$ such that

$$U^h(z) > V(z) \ \text{ and } \ U^h(z - he_i) \le V(z - he_i) \ \text{ for } i = 1, 2. \tag{5.111}$$

Note that by the concavity of $t \mapsto \sqrt{t}$ we have that

$$V(z) - V(z - he_i) \geq h\|\mu + \mu_s\|_\infty + h\|\sigma\|_\infty \frac{\sqrt{z_1 z_2}}{z_i}.$$

It follows that

$$\Big(V(z) - V(z - he_1) - h\|\mu + \mu_s\|_\infty\Big)\Big(V(z) - V(z - he_2) - h\|\mu + \mu_s\|_\infty\Big) \geq h^2\|\sigma\|_\infty^2.$$

By monotonicity of $t \mapsto (p_1 - t)_+(p_2 - t)_+$ we therefore have

$$\begin{aligned}\Big(V(z) - V(z - he_1) - h\mu(z)\Big)\Big(V(z) - V(z - he_2) - h\mu(z)\Big) &\geq h^2\|\sigma\|_\infty^2 \\ &\geq \Big(U^h(z) - U^h(z - he_1) - h\mu(z)\Big)\Big(U^h(z) - U^h(z - he_2) - h\mu(z)\Big).\end{aligned} \tag{5.112}$$

This contradicts (5.111), hence $U^h \leq V$. The proof is completed by invoking [12, Theorem 2.1]. □

We now extend the numerical convergence result to μ and σ^2 satisfying (F1) and (F3).

Corollary V.24. *Suppose that μ and σ^2 simultaneously satisfy (F1), (F3) and* (5.80)*, and let μ_s satisfy (F2). Define U^h as in Theorem V.23. Then we have*

$$U^h \longrightarrow U \quad \text{locally uniformly on } [0,\infty)^2, \tag{5.113}$$

where U is the unique Pareto-monotone viscosity solution of (P).

Proof. Define $\mu^k, \sigma^k, \mu_k, \sigma_k, U_k$ and U^k as in the proof of Theorem V.20. By definition we have $U_k \leq U \leq U^k$, and by Corollary V.19 and Remark V.18 we have $U_k, U^k \to U$ locally uniformly on $[0,\infty)^2$ as $k \to \infty$.

Let U_k^h and $U^{k,h}$ denote the numerical solutions defined by (S) for $\mu_k + \mu_s, \sigma_k$ and $\mu^k + \mu_s, \sigma^k$, respectively, extended to $[0,\infty)^2$ as in Theorem V.23. Since $\mu^k, \sigma^{2,k}, \mu_k,$

and σ_k^2 are Lipschitz continuous and μ_s satisfies (F2), we can apply Theorem V.23 to show that

$$U_k^h \longrightarrow U_k \quad \text{and} \quad U^{k,h} \longrightarrow U^k, \tag{5.114}$$

locally uniformly on $[0,\infty)^2$ as $h \to 0$. Since $\mu_k \le \mu \le \mu^k$ and $\sigma_k \le \sigma \le \sigma^k$, we can make an argument, as in Theorem V.23, based on a comparison principle for (S), to show that $U_k^h \le U^h \le U^{k,h}$ for all h, k. The proof is completed by combining this with (5.114) and the locally uniform convergence $U_k, U^k \to U$. □

5.5.1 Numerical simulations

We present here some numerical simulations comparing the numerical solutions of (P), computed by (S), to realizations of directed last passage percolation (DLPP). We restrict our attention to the box $[0,1]^2$ for simplicity. For the case of exponential DLPP, we consider three macroscopic means, λ_1, λ_2, and λ_3 given by

$$\lambda_1(x) = \begin{cases} 1, & \text{if } x_1 \ge 0.5 \text{ or } x_2 \ge 0.5, \\ 0, & \text{otherwise,} \end{cases} \tag{5.115}$$

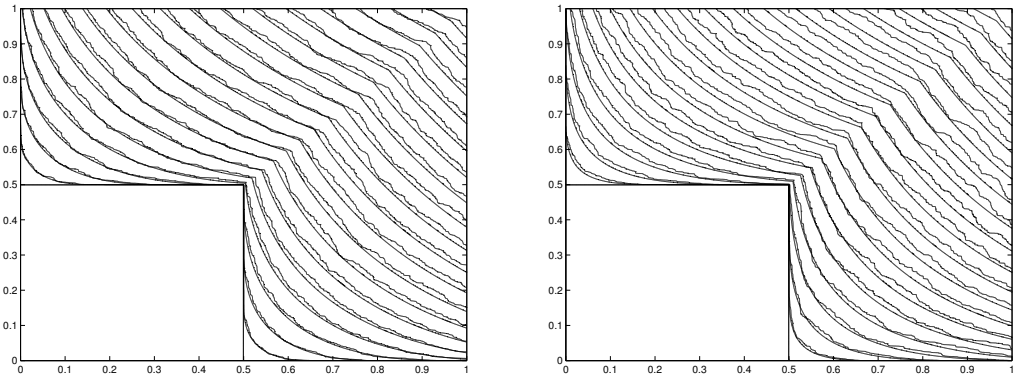
$$\lambda_2(x) = \exp\left(-10\,|x-(0.25,0.75)|^2\right) + \exp\left(-10\,|x-(0.75,0.25)|^2\right), \tag{5.116}$$

and

$$\lambda_3(x) = \begin{cases} 0.5, & \text{if } |x-(1,0)|^2 \le 0.49 \text{ or } |x-(0,1)|^2 \le 0.49, \\ 1, & \text{otherwise.} \end{cases} \tag{5.117}$$

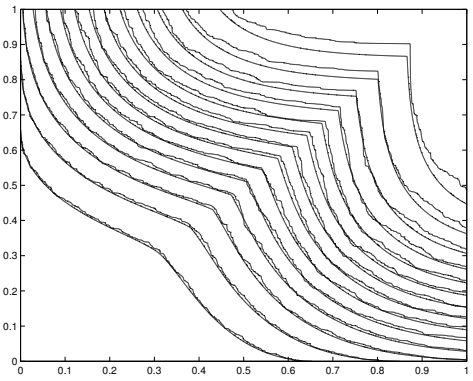
Since the results are very similar for geometric DLPP, we consider only one macroscopic parameter q given by

$$q(x) = \begin{cases} 0.5, & \text{if } x_1 \ge 0.5 \text{ or } x_2 \ge 0.5, \\ 1, & \text{otherwise.} \end{cases} \tag{5.118}$$

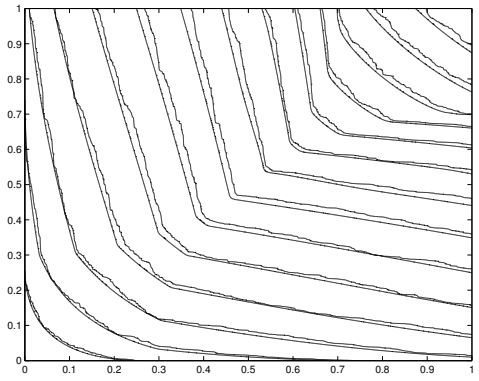


(a) Exponential DLPP with mean λ_1

(b) Geometric DLPP with parameter q



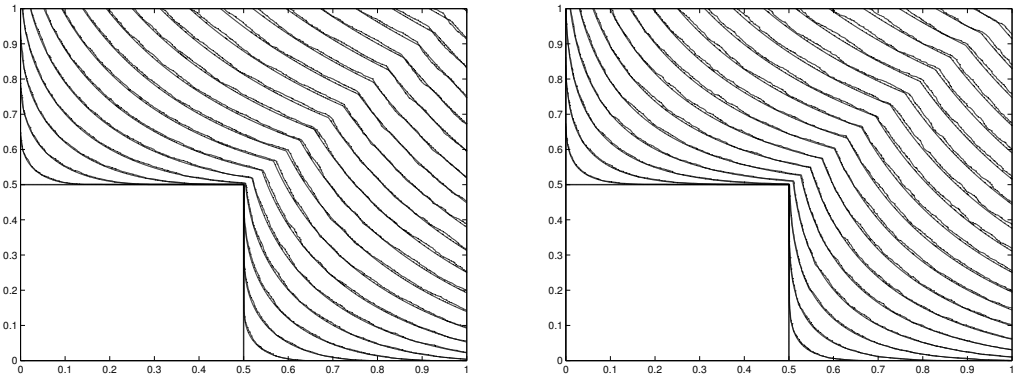
(c) Exponential DLPP with mean λ_2



(d) Exponential DLPP with mean λ_3

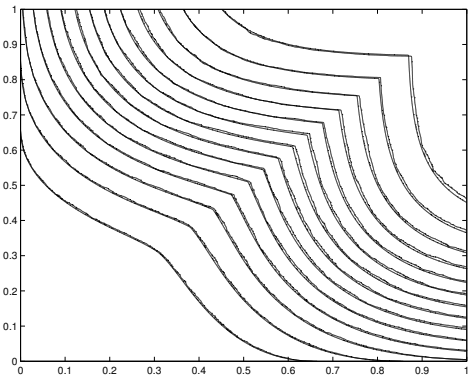
Figure 5.3: Comparisons of the level sets of numerical solutions of (P), computed via (S), and the level sets of exponential/geometric DLPP simulations on a 1000×1000 grid. The smooth lines correspond to the numerical solutions of (P), while the rough lines correspond to the DLPP simulations.

Figure 5.3 compares the level sets of the numerical solutions of (P) with simulations of exponential/geometric DLPP on a 1000×1000 grid. The smooth curves correspond to the level sets of the numerical solution of (P) while the rough curves correspond to the level sets of the last passage time from the DLPP simulation. Figure 5.4 shows the same comparison, except for DLPP simulations on a 5000×5000 grid. In both cases, the numerical solutions of (P) were computed on a 1000×1000 grid. To give an idea of the computational complexity, it takes approximately a quarter of a second to numerically solve the PDE on this grid in MATLAB on an average laptop.

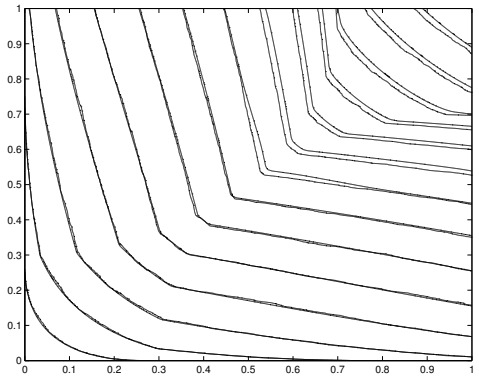


(a) Exponential DLPP with mean λ_1

(b) Geometric DLPP with parameter q



(c) Exponential DLPP with mean λ_2



(d) Exponential DLPP with mean λ_3

Figure 5.4: Comparisons of the level sets of numerical solutions of (P), computed via (S), and the level sets of exponential/geometric DLPP simulations on a 5000×5000 grid.

5.5.2 Finding maximal curves

We now propose an algorithm based on dynamic programming for finding maximizing curves, and we prove in Theorem V.25 and Corollary V.26 that the curve produced by our algorithm is approximately optimal for the variational problem (5.32) defining U. Other approaches to finding maximizing curves, such as the method of characteristics [33], or solving the Euler-Lagrange equations [81], are not guaranteed to produce optimal curves, due to crossing characteristics, and the possibility of local minima. Our method is related to the method of synthesis in optimal control theory for computing optimal controls from solutions of Hamilton-Jacobi-Bellman

equations [10].

Our algorithm has a parameter $\varepsilon > 0$ and a starting point $x \in \mathbb{R}^2_+$, and computes a curve γ_ε with $\gamma_\varepsilon(0) = 0$ and $\gamma_\varepsilon(1) = x$ that nearly maximizes J. The algorithm works by starting at x and tracing our way back to the origin by solving a series of dynamic programming problems. We set $x_0 = x$, and generate $x_1, \ldots, x_k, \ldots$ as follows: Given we are at step $k \geq 0$, we use a dynamic programming principle (similar to Proposition V.5) to write

$$U(x_k) = \max_{s \in [0,1]} \Big\{ U(y(s)) + W(y(s), x_k) \Big\}, \tag{5.119}$$

where $y(s) = x_k - (1-s, s)\varepsilon$. An application of Hölder's inequality yields

$$J(\gamma) \leq \mu^*(x_k)\varepsilon + 2\sigma^*(x_k)\varepsilon\sqrt{s(1-s)} + o(\varepsilon), \tag{5.120}$$

for any $\gamma \in \mathcal{A}$ with $\gamma(0) = y(s)$ and $\gamma(1) = x_k$. When μ and σ are continuous, this upper bound can be attained (in the limit as $\varepsilon \to 0$) by the diagonal curve $\gamma(t) = (1-t)y(s) + x_k t$. Thus we are justified in making the following approximation

$$W(y(s), x_k) = \sup_{\gamma \in \mathcal{A} \,:\, \gamma(0) = y(s), \gamma(1) = x_k} J(\gamma) \approx \mu(x_k)\varepsilon + 2\sigma(x_k)\varepsilon\sqrt{s(1-s)}. \tag{5.121}$$

Substituting (5.121) into (5.119) we find that

$$U(x_k) \approx \mu(x_k)\varepsilon + \max_{s \in [0,1]} \left\{ U(y(s)) + 2\sigma(x_k)\varepsilon\sqrt{s(1-s)} \right\}. \tag{5.122}$$

We then define

$$x_{k+1} := y(s_k^*)_+ = (x_k - (1 - s_k^*, s_k^*)\varepsilon)_+, \tag{5.123}$$

where $s_k^* \in [0,1]$ is the maximizing argument in (5.122) and $x_+ = (\max(x_1, 0), \max(x_2, 0))$. The algorithm is terminated as soon as $x_k \in \partial\mathbb{R}^2_+$ and we append the final terminal point $x_{k+1} = 0$. In (5.122), we set $U(y(s)) = 0$ whenever $y(s) \notin [0, \infty)^2$. The algorithm is summarized in Algorithm V.1. Notice that the boundary source μ_s does

not appear explicitly in Algorithm V.1, though it does appear implicitly through the solution U of (P).

Algorithm V.1: Find nearly maximizing path

Given a step size $\varepsilon > 0$ and $x_0 \in \mathbb{R}^2_+$, we generate $x_1, \ldots, x_k, \ldots$ as follows:
$k = 0$;
while $x_k \in \mathbb{R}^2_+$ **do**
  $s_k^* = \text{argmax}_{s\in[0,1]} \left\{ U(x_k - (1-s, s)\varepsilon) + 2\sigma(x_k)\varepsilon\sqrt{s(1-s)} \right\}$;
  $x_{k+1} = (x_k - (1 - s_k^*, s_k^*)\varepsilon)_+$;
end
$x_{k+1} = 0$;

Each step of the algorithm moves a distance of at least $\varepsilon/2$ in the direction $(-1, 0)$ or $(0, -1)$. If $x_0 \in [0, R]^2$, then the algorithm will terminate in at most $4R/\varepsilon$ steps. Furthermore, when μ and σ^2 are Lipschitz, we can show that the polygonal curve γ_ε generated by Algorithm V.1 has energy within $O(\varepsilon)$ of the maximizing curve. This is summarized in the following result.

Theorem V.25. *Let $R > 0$, suppose that μ and σ^2 are non-negative, globally Lipschitz continuous on $[0, R]^2$ with constant $C_{lip} > 0$, and suppose that μ_s satisfies (F2). Let $\varepsilon > 0$, $x_0 \in (0, R]^2$, and let $x_1, \ldots, x_K$ be the points generated by Algorithm V.1. Let $\gamma_\varepsilon : [0, 1] \to [0, \infty)^2$ be the monotone polygonal curve passing through $x_K, x_{K-1}, \ldots, x_1, x_0$. Then there exists a constant $C = C(\|\mu\|_\infty, \|\sigma\|_\infty) > 0$ such that*

$$U_{\mu+\mu_s,\sigma}(x_0) \le J_{\mu+\mu_s,\sigma}(\gamma_\varepsilon) + C(1 + C_{lip}R)\varepsilon. \tag{5.124}$$

Proof. For convenience, we set $U = U_{\mu+\mu_s,\sigma}$, $J = J_{\mu+\mu_s,\sigma}$, and we extend μ, σ and U to functions on $\mathbb{R}^2$ by setting $\mu(x) = \sigma(x) = U(x) = 0$ for $x \notin [0, \infty)^2$. Writing $\Delta t = 1/K$ and $t_j = j\Delta t$ for $j = 0, \ldots, K$, we can parameterize γ_ε so that

$$\gamma_\varepsilon'(t) = \frac{1}{\Delta t}(x_{K-j} - x_{K-j+1}) = \frac{\varepsilon}{\Delta t}(1 - s_{K-j}^*, s_{K-j}^*), \tag{5.125}$$

for $t \in (t_{j-1}, t_j)$ and $j \geq 3$. It follows that

$$\begin{aligned}
\int_{t_2}^{1} \ell(\gamma_\varepsilon(t), \gamma_\varepsilon'(t))\, dt &= \sum_{j=3}^{K} \int_{t_{j-1}}^{t_j} \ell(\gamma_\varepsilon(t), \gamma_\varepsilon'(t))\, dt \\
&= \varepsilon \sum_{j=3}^{K} \frac{1}{\Delta t} \int_{t_{j-1}}^{t_j} \mu(\gamma_\varepsilon(t)) + 2\sigma(\gamma_\varepsilon(t)) \sqrt{(1 - s^*_{K-j}) s^*_{K-j}}\, dt \\
&\geq \varepsilon \sum_{j=3}^{K} \left(\frac{1}{\Delta t} \int_{t_{j-1}}^{t_j} \mu(x_{K-j}) + 2\sigma(x_{K-j}) \sqrt{(1 - s^*_{K-j}) s^*_{K-j}}\, dt - 3C_{lip}\varepsilon \right) \\
&= \left(\sum_{j=3}^{K} \mu(x_{K-j}) + 2\sigma(x_{K-j}) \sqrt{(1 - s^*_{K-j}) s^*_{K-j}} \right) \varepsilon - 3KC_{lip}\varepsilon^2. \qquad (5.126)
\end{aligned}$$

An application of Hölder's inequality gives

$$J(\gamma) \leq \left(\mu(x_{K-j}) + 2\sigma(x_{K-j}) \sqrt{s(1-s)} \right) \varepsilon + 3C_{lip}\varepsilon^2, \qquad (5.127)$$

for $j \geq 2$ and any $\gamma \in \mathcal{A}$ with $\gamma(0) = y(s)$ and $\gamma(1) = x_{K-j}$. Combining this with the dynamic programming principle (5.119) we have

$$U(x_{K-j}) \leq \mu(x_{K-j})\varepsilon + \max_{s \in [0,1]} \left\{ U(y(s)) + 2\sigma(x_{K-j})\varepsilon \sqrt{s(1-s)}) \right\} + 3C_{lip}\varepsilon^2, \qquad (5.128)$$

for all $j \geq 2$. By the definition of s^*_{K-j} we have

(5.129)

$$U(x_{K-j}) \leq U(x_{K-j+1}) + \varepsilon \left(\mu(x_{K-j}) + 2\sigma(x_{K-j}) \sqrt{(1 - s^*_{K-j}) s^*_{K-j}} \right) + 3C_{lip}\varepsilon^2,$$

for $j \geq 3$. By iterating this inequality for $j = K, \ldots, 3$ we have

$$\begin{aligned}
U(x_0) &\leq U(x_{K-2}) + \left(\sum_{j=3}^{K} \mu(x_{K-j}) + 2\sigma(x_{K-j}) \sqrt{(1 - s^*_{K-j}) s^*_{K-j}} \right) \varepsilon + 3KC_{lip}\varepsilon^2 \\
&\overset{(5.126)}{\leq} U(x_{K-2}) + \int_{t_2}^{1} \ell(\gamma_\varepsilon(t), \gamma_\varepsilon'(t))\, dt + 6KC_{lip}\varepsilon^2. \qquad (5.130)
\end{aligned}$$

We have two cases now. Suppose first that $y(s^*_{K-2}) \notin [0, \infty)^2$. Then $U(y(s^*_{K-2})) = 0$ and by (5.128) we have that $U(x_{K-2}) \leq C\varepsilon$. Combining this with (5.130) we have

$$U(x_0) \leq J(\gamma_\varepsilon) + C\varepsilon + 6KC_{lip}\varepsilon^2. \qquad (5.131)$$

The proof is completed by noting that $K \leq 4R/\varepsilon$.

Suppose now that $y(s^*_{K-2}) \in [0,\infty)^2$. Then (5.129) holds for $j = 2$ and combining this with (5.130) we have

$$U(x_0) \leq U(x_{K-1}) + \int_{t_2}^{1} \ell(\gamma_\varepsilon(t), \gamma'_\varepsilon(t))\, dt + 6(K+1)C_{lip}\varepsilon^2. \tag{5.132}$$

Since $x_K = 0$ we must have $x_{K-1} \in \partial\mathbb{R}^2_+$. It follows that

$$\int_0^{t_1} \ell(\gamma_\varepsilon(t), \gamma'_\varepsilon(t))\, dt = U(x_{K-1}).$$

Inserting this into (5.132) we see that

$$U(x_0) \leq J(\gamma_\varepsilon) + 6(K+1)C_{lip}\varepsilon^2. \qquad \square$$

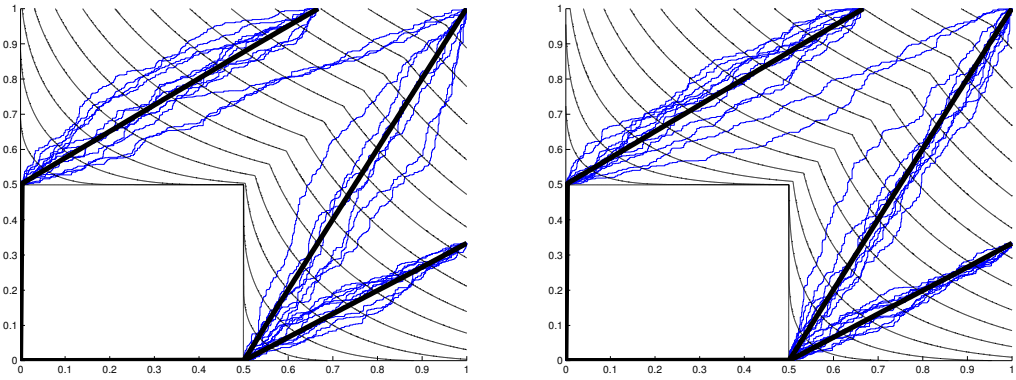
If μ and σ^2 are not globally Lipschitz continuous, then Algorithm V.1 is not guaranteed to yield optimal curves. However, it can be easily modified to give an algorithm that does.

Corollary V.26. *Suppose that μ and σ^2 simultaneously satisfy (F1), (F3) and* (5.80)*, and let μ_s satisfy (F2). Let μ_k and σ_k be sequences of functions such that μ_k and σ_k^2 are Lipschitz with constant k, $\mu_k \leq \mu$, $\sigma_k \leq \sigma$ and $U_{\mu_k+\mu_s,\sigma_k} \to U_{\mu+\mu_s,\sigma}$ locally uniformly. Let $x_0 \in (0,R]^2$ and let $\gamma_k : [0,1] \to [0,\infty)^2$ be the monotone polygonal curve generated by applying Algorithm V.1 to x_0, μ_k, σ_k and U_k with $\varepsilon = k^{-1}(U_{\mu+\mu_s,\sigma}(x_0) - U_{\mu_k+\mu_s,\sigma_k}(x_0))$. Then we have*

$$U(x_0) \leq J(\gamma_k) + o(1) \quad \text{as } k \to \infty. \tag{5.133}$$

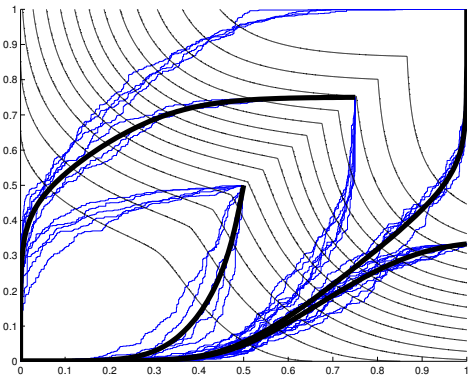
Proof. Let us set $J_k = J_{\mu_k+\mu_s,\sigma_k}$, $J = J_{\mu+\mu_s,\sigma}$, $U_k = U_{\mu_k+\mu_s,\sigma_k}$, and $U = U_{\mu+\mu_s,\sigma}$. By Theorem V.25 there exists a constant $C = C(\|\mu\|_\infty, \|\sigma\|_\infty) > 0$ such that

$$U_k(x_0) \leq J_k(\gamma_k) + C(1+kR)k^{-1}(U(x_0) - U_k(x_0)) \leq J(\gamma_k) + C(1+R)(U(x_0) - U_k(x_0)).$$

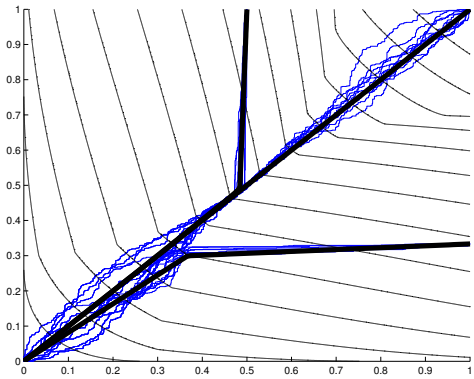


(a) Exponential DLPP with mean λ_1

(b) Geometric DLPP with parameter q



(c) Exponential DLPP with mean λ_2



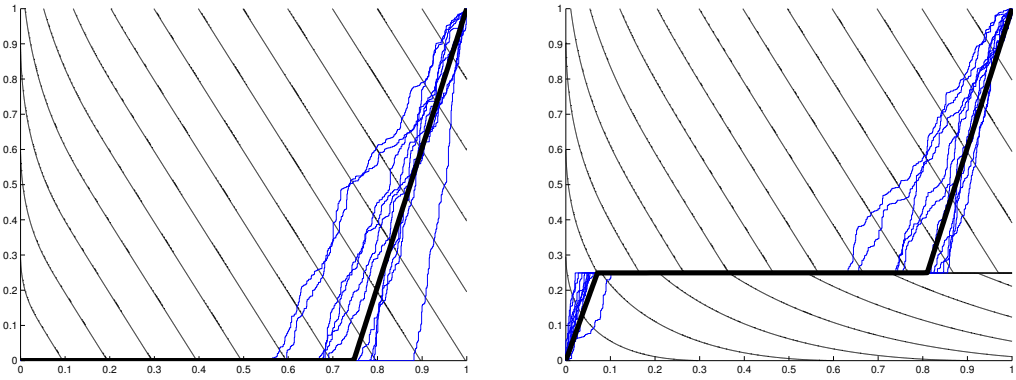
(d) Exponential DLPP with mean λ_3

Figure 5.5: Comparisons of the curve γ_ε ($\varepsilon = 0.01$) generated by Algorithm V.1 to the optimal paths from 10 realizations of DLPP for the macroscopic weights considered in Section 5.5.1. In each experiment, we show the curve γ_ε and optimal paths for several different terminal points $x_0 \in (0,1)^2$. Notice that in a), b) and c), there are multiple optimizing curves, and Algorithm V.1 finds only one curve, depending on the choice one makes when there are multiple maximizing arguments for s_k^*. The DLPP simulations were performed on a 1000×1000 grid, s_k^* was computed via an exhaustive search with a grid size of 0.01.

It follows that

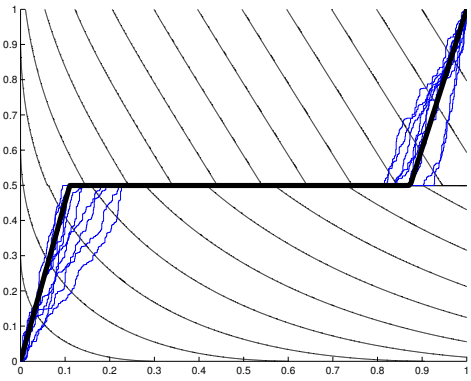
$$U(x_0) \leq J(\gamma_k) + C(2+R)(U(x_0) - U_k(x_0)). \qquad \square$$

We now show some simulation results using Algorithm V.1 to compute approximately optimal curves for the exponential/geometric DLPP simulations presented in Section 5.5.1. Figure 5.5 shows the curves generated by Algorithm V.1 along with optimal paths for 10 realizations of DLPP on a 1000×1000 grid. We also show the

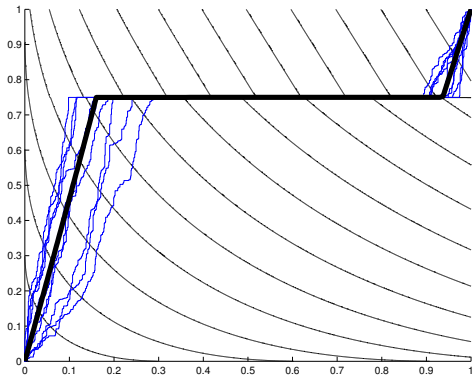


(a) Source at $\{x_2 = 0\}$

(b) Source at $\{x_2 = 0.25\}$



(c) Source at $\{x_2 = 0.5\}$



(d) Source at $\{x_2 = 0.75\}$

Figure 5.6: Comparisons of the optimal curve γ_ε ($\varepsilon = 0.01$) generated by Algorithm V.1 to the optimal paths from 10 realizations of exponential DLPP. The macroscopic weight functions are constant $\mu = 1$ on $[0, 1]^2$ plus a source term $\mu_s = 2$ concentrated on a horizontal line. The simulations were performed on a 1000×1000 grid.

level sets of the numerical solutions of (P) to give points of reference. In all cases, we used a step size of $\varepsilon = 0.01$ and computed s_k^* in Algorithm V.1 by an exhaustive search with a grid size of 0.01. With these choices of parameters, Algorithm V.1 runs in approximately a quarter of a second, assuming the numerical solution U is already available. Note also that we implemented Algorithm V.1 exactly as written, even when μ and σ are discontinuous, and do not substitute continuous versions as in Corollary V.26.

As in [81], it is expected that the optimal paths for DLPP will asymptotically concentrate around optimal curves for the variational problem, and this is clearly

reflected in the simulations in Figure 5.5. Notice that for exponential DLPP with means λ_1, λ_2 and geometric DLPP with parameter q, there are multiple maximizing curves for any terminal point x along the diagonal $\{x_1 = x_2\}$. We see that some of the DLPP realizations concentrate around one optimal path, while the remaining realizations concentrate around the other. Algorithm V.1 will of course only find one of the maximizing curves, depending on the choice one makes when there are multiple maximizing arguments in the definition of s_k^*.

We now show some simulations with a source term μ_s. Here we consider exponential DLPP with mean $\lambda = 1$ on $[0,1] \times (0,1]$ and $\lambda = 3$ on $[0,1] \times \{0\}$. Figure 5.6(a) shows the optimal curve generated by Algorithm V.1, along with the level sets of the numerical solution of (P) and the optimal paths from 10 realizations of exponential DLPP on a 1000×1000 grid.

Although our assumptions only allow sources on the boundary $\partial\mathbb{R}^2_+$, many of the results in the paper can be shown to hold for sources along horizontal or vertical lines in $\mathbb{R}^2_+$. The idea is to find the appropriate dynamic programming principle that plays the role of Proposition V.5, so that the effect of the weights in the bulk is separated from the source. In the case of a source along the line $\{x_2 = \alpha\}$ for $\alpha \in (0,1)$, and assuming no boundary sources, the dynamic programming principle would be

$$U(y) = \max_{0 \le x_1 \le x_1' \le y_1} \left\{ W(0,(x_1,\alpha)) + \int_{x_1}^{x_1'} \mu(t,\alpha) + \mu_s(t,\alpha)\, dt + W((x_1',\alpha), y) \right\},$$

where $U = U_{\mu+\mu_s,\sigma}$, $W = W_{\mu,\sigma}$, and μ and σ^2 are, say, Lipschitz on $[0,\infty)^2$, and μ_s represents the source, which is nonzero only on the line $\{x_2 = \alpha\}$. We can then use this dynamic programming principle and its discrete version (similar to (5.101)) in the proof of Theorem V.1. The one caveat is that U is in general discontinuous along the line containing the source, though U remains locally uniformly continuous on each

of the components of $\mathbb{R}^2_+$ obtained by removing the source line. Thus, U can only be identified via the variational problem (5.35), since we have not proven uniqueness of discontinuous viscosity solutions of (P). However, our numerical results suggest that either uniqueness holds for (P) in some special cases where U is discontinuous, or at the very least our numerical scheme for (P) selects the "correct" viscosity solution for the percolation problem.

Figure 5.6(b), 5.6(c), and 5.6(d) show the optimal curve generated by Algorithm V.1, along with DLPP simulations for sources on the horizontal lines $\{x_2 = 0.25\}$, $\{x_2 = 0.5\}$, and $\{x_2 = 0.75\}$, respectively.

5.5.3 TASEP with slow bond rate

Finally, we consider the totally asymmetric simple exclusion process (TASEP) with a slow bond rate at the origin. This model was originally introduced by Janowsky and Lebowitz [56], and some partial results were obtained more recently by Seppäläinen [84]. The process of interest is the usual TASEP with exponential rates of 1 at all locations in $\mathbb{Z}$ except for the origin, which has a slower rate of $r \in (0, 1]$. One can think of this as modeling traffic flow on a road with a single toll both that every car must pass through.

Through the correspondence with DLPP, the slow bond rate corresponds to a source on the diagonal $\{x_1 = x_2\}$. In the context of our paper, we would have

$$\mu(x) = \begin{cases} 1/r, & \text{if } x_1 = x_2, \\ 1, & \text{otherwise.} \end{cases} \tag{5.134}$$

Notice that μ does not satisfy the assumptions of Theorem V.1, and we do not expect the continuum limit (P) to hold in this case.

A quantity of interest is

$$\kappa(r) := \lim_{N\to\infty} \frac{1}{N} L(N,N) \quad \text{for } r \leq 1,$$

which corresponds to the reciprocal of the maximum TASEP current [84]. It is known that $\kappa(1) = 4$ and Seppäläinen [84] proved the following bounds:

$$\max\left\{4, \frac{r^2 + 2(1+r)}{2r(1+r)}\right\} \leq \kappa(r) \leq 3 + \frac{1}{r}. \tag{5.135}$$

It is an open problem to determine $\kappa(r)$ for $r < 1$. In particular, one is interested in whether $\kappa(r) > 4$ for all $r < 1$, or if there are some values of r close to $r = 1$ for which the inverse current $\kappa(r)$ remains unchanged.

Even though we do not expect our continuum limit Hamilton-Jacobi equation to hold for the slow bond rate problem, it is nevertheless interesting to see what our results would say about this open problem were they to hold. It is easy to see that $U_{\mu,\sigma}(1,1) = 4/r$ for $\mu = \sigma$ given by (5.134). Indeed, one can see that the optimal curve in the variational problem (5.32) must lie on the diagonal $\{x_1 = x_2\}$, which gives the energy $4/r$. This would suggest that

$$\kappa(r) = \lim_{N\to\infty} \frac{1}{N} L(N,N) = U_{\mu,\sigma}(1,1) = \frac{4}{r}.$$

Notice that this violates the bounds in (5.135), which indicates that the Hamilton-Jacobi equation continuum limit (Theorem V.1) does *not* hold for sources along diagonal lines.

5.6 Discussion and future work

In this chapter, we identified a Hamilton-Jacobi equation for the continuum limit of a macroscopic two-sided directed last passage percolation (DLPP) problem. We rigorously proved the continuum limit when the macroscopic rates are discontinuous.

Furthermore, we presented a numerical scheme for solving the Hamilton-Jacobi equation, and an algorithm for finding optimal curves based on a dynamic programming principle. Below we make some remarks, discuss simple extensions of this work, and ideas for future work.

- Regularity of μ, σ: There are many simple modifications of (F1) under which one can prove Theorem V.1. For example, the existence of the set Ω bounded by the strictly decreasing curve Γ and $\partial\mathbb{R}^2_+$ on which $\mu = \sigma = 0$ is not necessary, and one can check that the proofs hold without this assumption. This would correspond to a TASEP model with step initial condition. The curves Γ_i on which μ and σ may admit discontinuities can all be chosen to be strictly decreasing instead of increasing, with appropriate modifications in the proofs. In fact, we can even allow the curves to switch from strictly increasing to strictly decreasing, provided the critical point is isolated, and we make an additional cone condition assumption at this point. However, the curves Γ_i cannot have any positive measure flat regions, as this can induce discontinuities in U, as shown in Remark V.7.

- Discontinuous viscosity solutions: The regularity assumption (F1) was chosen to ensure that U is locally uniformly continuous. This is essential for invoking the Arzelà-Ascoli Theorem in the proof of Theorem V.17, and in the proof of the comparison principle for (P) (Theorem III.16). We believe that Theorem V.1 holds under far more general assumptions on μ, allowing U to be discontinuous. Presently, we do not know how to prove this. The largest obstacle seems to be proving uniqueness of viscosity solutions of (P) when the solutions U and the macroscopic weights μ are discontinuous. Our numerical

results seem to support this conjecture, as the numerical scheme is able to very accurately capture discontinuities in U.

- HYDRODYNAMIC LIMIT OF TASEP: As we showed in Section 5.1.2, the Hamilton-Jacobi equation (P) is formally equivalent to the conservation law governing the hydrodynamic limit of TASEP [43, 83]. It would be very interesting to make this connection rigorous.

- HIGHER DIMENSIONS: The main obstacle in generalizing the Hamilton-Jacobi equation (P), and the results in this paper, to dimensions $d \geq 3$, is the fact that the exact form of the time constant (5.2) for *i.i.d.* random variables $X(i,j)$ is unknown. If an exact form for the time constant U were to be discovered for $d \geq 3$, then we anticipate no problems in generalizing the results in this paper to higher dimensions. We should note that although the exact form of U is unknown for $d \geq 3$, it is known that U is continuous, 1-homogeneous, symmetric in all variables, and superadditive, under fairly broad assumptions on the distribution of $X(i,j)$ [71]. This is enough to show that U is the viscosity solution of some Hamilton-Jacobi equation, but the explicit form of the equation is unknown.

CHAPTER VI

Conclusion

In this dissertation we discovered and rigorously proved continuum limits for some sorting and percolation problems of interest in mathematical and scientific contexts. The first problem we addressed was non-dominated sorting, which is fundamental in multi-objective optimization problems. The sorting is akin to arranging points in Euclidean space into fronts according to a partial order. We proved that the fronts converge in the (random) large sample size limit to the level sets of a function that satisfies a Hamilton-Jacobi equation in the viscosity sense. As an application of this, we proved that non-dominated sorting is stable under bounded random perturbations. We then proposed, and proved convergence for, a linear complexity numerical scheme for solving this Hamilton-Jacobi equation. This allowed us to design a fast, potentially sublinear, algorithm for approximate non-dominated sorting of massive datasets. We evaluated our algorithm on a set of synthetic data, and on a massive dataset from a real-world anomaly detection problem. Our results indicate that the algorithm can significantly reduce the computational complexity of non-dominated sorting while obtaining excellent accuracy in the asymptotic regime. We believe this algorithm may be particularly useful in the big-data streaming context, where the sorting would have to be efficiently updated upon the arrival of each new sample

without keeping a history of previous samples.

The second problem we studied was directed last passage percolation (DLPP), which is a stochastic growth model with many applications in scientific and mathematical contexts. The DLPP model is equivalent to a model of zero-temperature directed polymer growth, to stochastic interacting particle systems, such as the totally asymmetric simple exclusion process (TASEP), and to queuing systems, among others. The DLPP model is also closely related to non-dominated sorting via the longest chain problem, and we showed that by applying similar techniques, we could derive and rigorously prove a Hamilton-Jacobi equation continuum limit for the time-constant in a macroscopic two-sided DLPP model with discontinuous exponential or geometric weights. We proved convergence of a fast numerical scheme and showed how to use dynamic programming to find the asymptotic shapes of optimal DLPP paths. We also showed formally that this Hamilton-Jacobi equation is equivalent to the conservation law for the hydrodynamic limit of TASEP. This Hamilton-Jacobi equation is new in the context of DLPP and TASEP, and we believe it will prove to be a useful tool for studying properties of the hydrodynamic limits of these important stochastic processes.

This dissertation opens up several avenues for future research, so let us say a few words about some of these possibilities. First, we can view the PDE-based approach to non-dominated sorting presented in Chapter IV as merely one particular technique for estimating the viscosity solution U of the Hamilton-Jacobi equation (P) for non-dominated sorting from *i.i.d.* samples $X_1, \ldots, X_n$. There are certainly many other approaches one could take to perform this estimation, *e.g.,* one could use a different density estimator, or one could consider invoking the variational problem associated to the Hamilton-Jacobi equation. Thus, an important question is the following:

Among all estimators of U from $X_1, \ldots, X_n$, what is the best rate of convergence in n? A related problem is to construct an estimator that achieves the optimal rate. The question of optimal convergence rates is critical in practice, because it tells us in a very precise sense how large we need to make n to achieve a desired degree of ranking accuracy. This is particularly useful in the context of big data streaming problems, as it would tell us how much of the incoming data we can discard without sacrificing sorting accuracy.

A second direction for future work concerns the numerical scheme for solving the Hamilton-Jacobi equation (P). Since the complexity of our scheme grows exponentially fast in dimension, we found the algorithm useful in only relatively small dimensions, i.e., $d = 2, 3, 4$. Thus, one important questions is: How do we numerically solve (P) efficiently in higher dimensions? This is a crucial question if one is to use the Hamilton-Jacobi equation (P) for non-dominated sorting in even moderately high dimensions. Some ideas one could pursue include using the method of characteristics, and performing gradient ascent to find maximizing curves in the variational interpretation of (P). We provide evidence in Chapter IV indicating that our scheme has an accuracy on the order of $O(h^{\frac{1}{d}})$, which scales poorly in dimension. Thus another important question is the following: Is there a more accurate scheme for solving (P) even in low dimensions? One would hope to obtain an $O(h)$ scheme for a particular class of density functions, and we would like to discover such a scheme and prove the increased accuracy rigorously. Such a scheme would allow us to use a coarser grid, and may possibly allows us to use the scheme for non-dominated sorting in a higher dimension then we are currently able to.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] G. Adhar. Parallel algorithms for chains and anti-chains of points on a plane. In *International Conference on Parallel and Distributed Systems (ICPADS)*, volume 2, pages 1–7, Dec. 2007.

[2] D. Aldous and P. Diaconis. Hammersley's interacting particle process and longest increasing subsequences. *Probability Theory and Related Fields*, 103(2):199–213, 1995.

[3] D. Aldous and P. Diaconis. Longest increasing subsequences: from patience sorting to the baik-deift-johansson theorem. *Bulletin of the American Mathematical Society*, 36(4):413–432, 1999.

[4] T. Athan and P. Papalambros. A note on weighted criteria methods for compromise solutions in multi-objective optimization. *Engineering Optimization*, 27(2):155–176, 1996.

[5] F. Baccelli, A. Borovkov, and J. Mairesse. Asymptotic results on infinite tandem queueing networks. *Probability Theory and Related Fields*, 118(3):365–405, 2000.

[6] Z.-D. Bai, L. Devroye, H.-K. Hwang, and T.-H. Tsai. Maxima in hypercubes. *Random Structures & Algorithms*, 27(3):290–309, 2005.

[7] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, 33(5):1643–1697, 2005.

[8] J. Baik, P. L. Ferrari, and S. Péché. Convergence of the two-point function of the stationary TASEP. In *Singular Phenomena and Scaling in Mathematical Models*, pages 91–110. Springer, 2014.

[9] J. Baik and E. M. Rains. Limiting distributions for a polynuclear growth model with external sources. *Journal of Statistical Physics*, 100(3-4):523–541, 2000.

[10] M. Bardi and I. Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Springer, 1997.

[11] G. Barles, A. Briani, and E. Chasseigne. A bellman approach for two-domains optimal control problems in rn. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(03):710–739, 2013.

[12] G. Barles and P. E. Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic Analysis*, 4(3):271–283, 1991.

[13] O. Barndorff-Nielsen and M. Sobel. On the distribution of the number of admissible points in a vector random sample. *Theory of Probability and its Applications*, 11(2):249–269, 1966.

[14] Y. Baryshnikov and J. E. Yukich. Maximal points and gaussian fields. Preprint, 2005.

[15] G. Ben Arous and I. Corwin. Current fluctuations for TASEP: A proof of the prähofer–spohn conjecture. *The Annals of Probability*, 39(1):104–138, 2011.

[16] B. Bollobás and G. Brightwell. The height of a random partial order: concentration of measure. *The Annals of Applied Probability*, 2(4):1009–1018, 1992.

[17] B. Bollobás and P. Winkler. The longest chain among random points in Euclidean space. *Proceedings of the American Mathematical Society*, 103(2):347–353, June 1988.

[18] E. Bolthausen. A note on the diffusion of directed polymers in a random environment. *Communications in Mathematical Physics*, 123(4):529–534, 1989.

[19] A. Borodin and S. Péché. Airy kernel with two sets of parameters in directed percolation and random matrix theory. *Journal of Statistical Physics*, 132(2):275–290, 2008.

[20] J. Calder, S. Esedoglu, and A. O. Hero. A Hamilton-Jacobi equation for the continuum limit of non-dominated sorting. *SIAM Journal on Mathematical Analysis*, 46(1):603–638, 2014.

[21] F. Camilli and E. Prados. Shape-from-shading with discontinuous image brightness. *Applied Numerical Mathematics*, 56(9):1225–1237, 2006.

[22] F. Camilli and A. Siconolfi. Hamilton-Jacobi equations with measurable dependence on the state variable. *Advances in Differential Equations*, 8(6):733–768, 2003.

[23] T. Cecil, J. Qian, and S. Osher. Numerical methods for high dimensional Hamilton–Jacobi equations using radial basis functions. *Journal of Computational Physics*, 196(1):327–347, 2004.

[24] T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, Feb. 2001.

[25] F. Comets, T. Shiga, and N. Yoshida. Probabilistic analysis of directed polymers in a random environment: a review. *Advanced Studies in Pure Mathematics*, 39:115–142, 2004.

[26] I. Corwin, P. L. Ferrari, and S. Péché. Limit processes for TASEP with shocks and rarefaction fans. *Journal of Statistical Physics*, 140(2):232–267, 2010.

[27] M. Crandall, H. Ishii, and P. Lions. User's guide to viscosity solutions of second order partial differential equations. *Bulletin of the American Mathematical Society*, 27(1):1–67, July 1992.

[28] K. Deb. *Multi-objective optimization using evolutionary algorithms*. Wiley, Chichester, UK, 2001.

[29] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

[30] K. Deckelnick and C. M. Elliott. Uniqueness and error analysis for Hamilton-Jacobi equations with discontinuities. *Interfaces and Free Boundaries*, 6(3):329–349, 2004.

[31] J.-D. Deuschel and O. Zeitouni. Limiting curves for i.i.d. records. *The Annals of Probability*, 23(2):852–878, 1995.

[32] M. Ehrgott. *Multicriteria Optimization (2. ed.)*. Springer, 2005.

[33] L. C. Evans. *Partial differential equations, 2nd edition*. Graduate studies in mathematics. American Mathematical Society, 1998.

[34] S. Felsner and L. Wernisch. Maximum k-chains in planar point sets: Combinatorial structure and algorithms. *SIAM Journal on Computing*, 28(1):192–209, 1999.

[35] P. L. Ferrari and H. Spohn. Scaling limit for the space-time covariance of the stationary totally asymmetric simple exclusion process. *Communications in Mathematical Physics*, 265(1):1–44, 2006.

[36] G. Fleury, A. Hero, S. Yoshida, T. Carter, C. Barlow, and A. Swaroop. Pareto analysis for gene filtering in microarray experiments. In *European Signal Processing Conference (EUSIPCO)*, 2002.

[37] G. Fleury, A. Hero, S. Zareparsi, and A. Swaroop. Gene discovery using Pareto depth sampling distributions. *Journal of the Franklin Institute*, 341(1–2):55–75, 2004.

[38] G. Fleury, A. Hero, S. Zareparsi, and A. Swaroop. Pareto depth sampling distributions for gene ranking. In *IEEE International Symposium on Biomedical Imaging: Nano to Macro* , volume 1, pages 65–68, Apr. 2004.

[39] C. Fonseca and P. Fleming. Genetic algorithms for multiobjective optimization : formulation, discussion and generalization. *Proceedings of the Fifth International Conference on Genetic Algorithms*, 1:416–423, July 1993.

[40] C. Fonseca and P. Fleming. An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary Computation*, 3(1):1–16, 1995.

[41] M. Garavello and P. Soravia. Optimality principles and uniqueness for bellman equations of unbounded control problems with discontinuous running cost. *Nonlinear Differential Equations and Applications (NoDEA)*, 11(3):271–298, 2004.

[42] M. Garavello and P. Soravia. Representation formulas for solutions of the HJI equations with discontinuous coefficients and existence of value in differential games. *Journal of Optimization Theory and Applications*, 130(2):209–229, 2006.

[43] N. Georgiu, R. Kumar, and T. Seppäläinen. TASEP with discontinuous jump rates. *Latin American Journal of Probability and Mathematical Statistics*, 7:293–318, 2010.

[44] A. Gilbert and M. Strauss. Analysis of data streams: Computational and algorithmic challenges. *Technometrics*, 49(3):346–356, 2007.

[45] P. W. Glynn and W. Whitt. Departures from many queues in series. *The Annals of Applied Probability*, 1(4):546–572, 1991.

[46] J. Hammersley. A few seedlings of research. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 345–394, 1972.

[47] J. Handl and J. Knowles. Exploiting the trade-off – The benefits of multiple objectives in data clustering. In *Evolutionary Multi-Criterion Optimization*, pages 547–560. Springer, 2005.

[48] A. Hero. Gene selection and ranking with microarray data. In *IEEE International Symposium on Signal Processing and its Applications*, volume 1, pages 457–464, July 2003.

[49] A. Hero and G. Fleury. Posterior Pareto front analysis for gene filtering. In *Proceedings of the Workshop on Genomic Signal Processing and Statistics (GENSIPS)*, 2002.

[50] A. Hero and G. Fleury. Pareto-optimal methods for gene ranking. *VLSI Signal Processing*, 38(3):259–275, 2004.

[51] K.-J. Hsiao, K. Xu, J. Calder, and A. Hero. Multi-criteria anomaly detection using Pareto Depth Analysis. In *Advances in Neural Information Processing Systems 25*, pages 854–862. NIPS, 2012.

[52] D. A. Huse and C. L. Henley. Pinning and roughening of domain walls in ising systems due to random impurities. *Physical Review Letters*, 54(25):2708, 1985.

[53] H.-K. Hwang and T.-H. Tsai. Multivariate records based on dominance. *Electronic Journal of Probability*, 15(60):1863–1892, 2010.

[54] J. Z. Imbrie and T. Spencer. Diffusion of directed polymers in a random environment. *Journal of Statistical Physics*, 52(3-4):609–626, 1988.

[55] V. M. Ivanin. An asymptotic estimate of the mathematical expectation of the number of elements of the Pareto set. *Kibernetika (Kiev)*, 11(1):97–101, 1975.

[56] S. Janowsky and J. Lebowitz. Exact results for the asymmetric simple exclusion process with a blockage. *Journal of Statistical Physics*, 77(1-2):35–51, 1994.

[57] M. Jensen. Reducing the run-time complexity of multiobjective EAs: The NSGA-II and other algorithms. *IEEE Transactions on Evolutionary Computation*, 7(5):503–515, 2003.

[58] K. Johansson. Shape fluctuations and random matrices. *Communications in Mathematical Physics*, 209(2):437–476, 2000.

[59] A. P. Korostelev and A. B. Tsybakov. *Minimax theory of image reconstruction*. Springer-Verlag, New York, 1993.

[60] D. Kossmann, F. Ramsak, S. Rost, et al. Shooting stars in the sky: an online algorithm for skyline queries. In *Proceedings of the 28th International Conference on Very Large Data Bases*, pages 275–286, 2002.

[61] A. Krishnan. Variational formula for the time-constant of first-passage percolation I: Homogenization. *arXiv preprint:1311.0316*, 2013.

[62] H. Krug and H. Spohn. Kinetic roughening of growing surfaces. In C. Godrèche, editor, *Solids Far from Equilibrium*. Cambridge University Press, 1991.

[63] A. Kumar and A. Vladimirsky. An efficient method for multiobjective optimal control and optimal control subject to integral constraints. *Journal of Computational Mathematics*, 28(4):517–551, 2010.

[64] D. Loftsgaarden and C. Quesenberry. A nonparametric estimate of a multivariate density function. *The Annals of Mathematical Statistics*, 36(3):1049–1051, 1965.

[65] B. F. Logan and L. A. Shepp. A variational problem for random Young tableaux. *Advances in Mathematics*, 26(2):206–222, 1977.

[66] R. Lou and M. Sarrafzadeh. An optimal algorithm for the maximum three-chain problem. *SIAM Journal on Computing*, 22(5):976–993, 1993.

[67] N. Madavan. Multiobjective optimization using a Pareto differential evolution approach. In *Proceedings of the Congress on Evolutionary Computation (CEC)*, 2002.

[68] B. Majecka. Statistical models of pedestrian behaviour in the forum. Master's thesis, School of Informatics, University of Edinburgh, 2009.

[69] J. J. Manfredi, M. Parviainen, and J. D. Rossi. On the definition and properties of p-harmonious functions. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze-Serie V*, 11(2):215, 2012.

[70] J. B. Martin. Linear growth for greedy lattice animals. *Stochastic Processes and their Applications*, 98(1):43–66, 2002.

[71] J. B. Martin. Limiting shape for directed percolation models. *The Annals of Probability*, 32(4):2908–2937, 2004.

[72] I. Mitchell and S. Sastry. Continuous path planning with multiple constraints. In *IEEE Conference on Decision and Control*, volume 5, pages 5502–5507, Dec. 2003.

[73] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[74] R. Newcomb and J. Su. Eikonal equations with discontinuities. *Differential and Integral Equations-Athens*, 8:1947–1947, 1995.

[75] D. Papadias, Y. Tao, G. Fu, and B. Seeger. Progressive skyline computation in database systems. *ACM Transactions on Database Systems (TODS)*, 30(1):41–82, 2005.

[76] P. Pevzner. *Computational Molecular Biology*. The MIT Press, 2000.

[77] M. Prähofer and H. Spohn. Universal distributions for growth processes in 1+1 dimensions and random matrices. *Physical Review Letters*, 84(21):4882–4885, 2000.

[78] M. Prähofer and H. Spohn. Current fluctuations for the totally asymmetric simple exclusion process. In *In and Out of Equilibrium*, pages 185–204. Springer, 2002.

[79] A. Prekopa. On logarithmic concave measures and functions. *Acta Sci. Math.(Szeged)*, 34:335–343, 1973.

[80] F. Rezakhanlou. Continuum limit for some growth models. *Stochastic Processes and their Applications*, 101(1):1–41, 2002.

[81] L. T. Rolla and A. Q. Teixeira. Last passage percolation in macroscopically inhomogeneous media. *Electronic Communications in Probability*, 13:131–139, 2008.

[82] H. Rost. Non-equilibrium behaviour of a many particle process: Density profile and local equilibria. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 58(1):41–53, 1981.

[83] T. Seppäläinen. Hydrodynamic scaling, convex duality, and asymptotic shapes of growth models. Unpublished, 1996.

[84] T. Seppäläinen. Hydrodynamic profiles for the totally asymmetric exclusion process with a slow bond. *Journal of Statistical Physics*, 102(1-2):69–96, 2001.

[85] T. Seppäläinen. A growth model in multiple dimensions and the height of a random partial order. In *Asymptotics: particles, processes and inverse problems*, volume 55 of *IMS Lecture Notes Monograph Series*, pages 204–233. Institute for Mathematical Statistics, Beachwood, OH, 2007.

[86] P. Soravia. Boundary value problems for Hamilton-Jacobi equations with discontinuous Lagrangian. *Indiana University Mathematics Journal*, 51(2):451–477, 2002.

[87] P. Souganidis. Stochastic homogenization of Hamilton–Jacobi equations and some applications. *Asymptotic Analysis*, 20(1):1–11, 1999.

[88] T. Speed. *Statistical analysis of gene expression microarray data*, volume 11. Chapman & Hall/CRC, 2003.

[89] N. Srinivas and K. Deb. Muiltiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, 2(3):221–248, 1994.

[90] A. Tourin. A comparison theorem for a piecewise Lipschitz continuous Hamiltonian and application to shape-from-shading problems. *Numerische Mathematik*, 62(1):75–85, 1992.

[91] A. Tsybakov. *Introduction to nonparametric estimation*. Springer, 2009.

[92] S. Ulam. Monte carlo calculations in problems of mathematical physics. In E. F. Beckenbach, editor, *Modern Mathematics for Engineers*, pages 261–281. McGraw–Hill, 1961.

[93] A. Vershik and S. Kerov. Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tables. *Soviet Doklady Mathematics*, 18(527-531):38, 1977.

[94] A. M. Vershik. Asymptotic combinatorics and algebraic analysis. In *Proceedings of the International Congress of Mathematicians*, pages 1384–1394. Springer, 1995.

[95] G. Viennot. Chain and antichain families grids and Young tableaux. *North-Holland Mathematics Studies*, 99:409–463, 1984.