Minicourse: Partial Differential Equations for Data Peeling

Jeff Calder*
School of Mathematics
University of Minnesota
jwcalder@umn.edu

Contents

*Much of this research was supported by NSF grant DMS-1500829.

1 Introduction

Many problems in science and engineering involve the sorting, or ordering, of large amounts of data. Being discrete in nature, these problems can be challenging in the context of big data. We consider here continuum limits for a class of algorithms for sorting, or arranging, a finite set $X \subset \mathbb{R}^d$ into layers by repeatedly peeling away extremal points. Different notions of extremality lead to different notions of sorting.

For example, if we say that coordinatewise minimal points are extremal, then we get *nondominated sorting* (Section 1.1), which is a fundamental algorithm in multi-objective optimization. If we choose vertices of the convex hull to be extremal, then we get *convex hull peeling* (Section 1.2), which is used to define multi-variate order statistics, such as the median. We can generalize the preceding two examples by repeatedly peeling away the so-called *Pareto envelope* of the set of points—we discuss this in Section 1.3. Finally, if our data is endowed with a graph structure, where pairs of points are connected by edges of varying weights, extremal points are often defined as points that have small degree. Peeling away points of small degree is called *graph peeling* (Section 1.4), and is used to discover significant parts of a graph in data science.

Since the size of typical datasets in data science problems is rapidly increasing, it is natural to ask what happens in the limit as the number of samples tends to infinity. This is a problem mathematicians can address, but we first need a model for our data. Probably the most common model is a sequence of independent and identically distributed random variables. For mathematical convenience, we will instead model our data as a spatial Poisson point process on $\mathbb{R}^d$ with intensity function $nf(x)$ where $n \in \mathbb{N}$. We denote the Poisson point process by X_{nf}. For the reader who is not familiar with Poisson point processes, we have summarized some important properties in the appendix in Section A. All of the results in the minicourse are equally true for a sequence of n independent and identically distributed random variables with probability density function f.

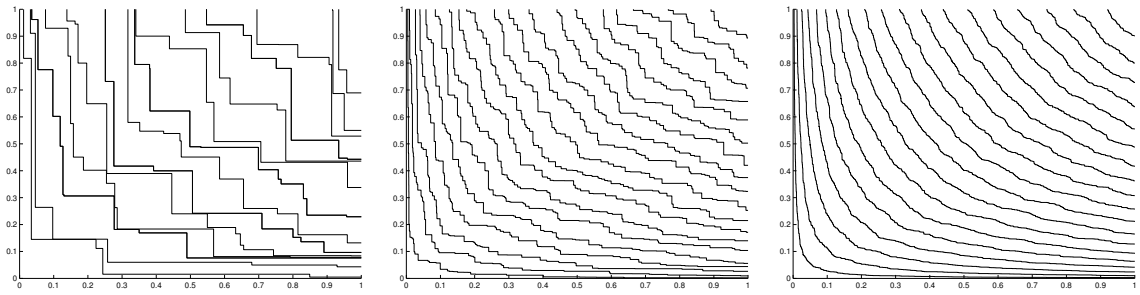
For all of the algorithms mentioned above, we can consider the problem of sorting the random Poisson cloud X_{nf} as $n \to \infty$. For nondominated sorting and convex hull peeling, we can prove that the sorting has a continuum limit that corresponds to solving a partial differential equation (PDE) in the viscosity sense. Similar results are expected to be true for peeling the Pareto envelope and graph peeling, and these are interesting problems for future research. Continuum limits for these peeling algorithms give us a better understanding of what the algorithms are doing, and sometimes can lead to fast approximate sorting algorithms based on solving the PDE instead of sorting.

In Sections 1.1–1.4, we describe each of the sorting algorithms above in more detail, and in Section 1.5 we give an overview of the content to be covered in the minicourse. The bulk of the minicourse is the content of Sections 3, 4, and 5.

We gratefully acknowledge the support of the National Science Foundation under grants NSF DMS-1500829, 1656030, which has supported much of the research described in this minicourse.

1.1 Nondominated sorting

Nondominated sorting is an algorithm for sorting points in Euclidean space into layers by repeatedly removing the coordinatewise minimal points. The algorithm is fundamental in multi-



(a) $n = 10^2$ points (b) $n = 10^4$ points (c) $n = 10^6$ points

Figure 1.1: Examples of Pareto fronts n random points on $[0,1]^2$.

objective optimization, where it is a key component of the so-called genetic and evolutionary algorithms for continuous multi-objective optimization [24, 25, 32, 33, 62]. Multi-objective optimization is ubiquitous in science and engineering, arising in problems such as control theory and path planning [46,50,52], gene selection and ranking [29–31,36–38,61], data clustering [35], database systems [45, 55] and image processing and computer vision [16, 54].

Let X_{nf} be a Poisson point process with intensity nf. Define the partial order

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

A point $x \in X_{nf}$ is minimal with respect to the coordinatewise partial order $\leqq$ if no other point $y \in X_{nf}$ is smaller (i.e., $y \leqq x$). The subset of minimal points in X_{nf} are called *nondominated*, or *Pareto optimal*, and constitute the first *Pareto front*, denoted $\mathcal{F}_1$. The second Pareto front, $\mathcal{F}_2$ consists of the minimal points from $X_{nf} \setminus \mathcal{F}_1$, and in general

$$\mathcal{F}_k = \text{Minimal elements of } X_{nf} \setminus (\mathcal{F}_1 \cup \cdots \cup \mathcal{F}_{k-1}).$$

The process of arranging X_{nf} into the Pareto fronts $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \dots$ is called *nondominated sorting* [25]. Figure 1.1 shows the Pareto fronts sorting random points.

Nondominated sorting is equivalent to the longest chain problem, which has a long history in probability [9, 26, 34, 48, 66, 67]. Recall that a *chain* in a partially ordered set is a totally ordered subset. In our setting, a chain in X_{nf} with respect to $\leqq$ is an up/right path through X_{nf}, that is, a path that with each step always increases in every coordinate. Let $\ell(S)$ denote the length of a longest chain in the set $S \subset \mathbb{R}^d$, and define

$$U_n(x) = \ell([0, x] \cap X_{nf}) \tag{1.2}$$

where $[0, x] := [0, x_1] \times [0, x_2] \times \cdots \times [0, x_d]$. We claim that for every $x \in X_{nf}$

$$U_n(x) = k \quad \text{if and only if} \quad x \in \mathcal{F}_k. \tag{1.3}$$

To see this, note first that any given Pareto front $\mathcal{F}_k$ can only contain at most one point from any chain. Thus, if $x \in \mathcal{F}_k$ we must have $U_n(x) \leq k$. To show equality, we argue how to construct a chain of length k. The point $x \in \mathcal{F}_k$ will be the last point on the chain. Now, we know $x \in \mathcal{F}_k$ precisely because there exists a point on the previous front $\mathcal{F}_{k-1}$ that is less than x with respect to $\leqq$. This point will be the second to last point on the chain. See Figure 1.2

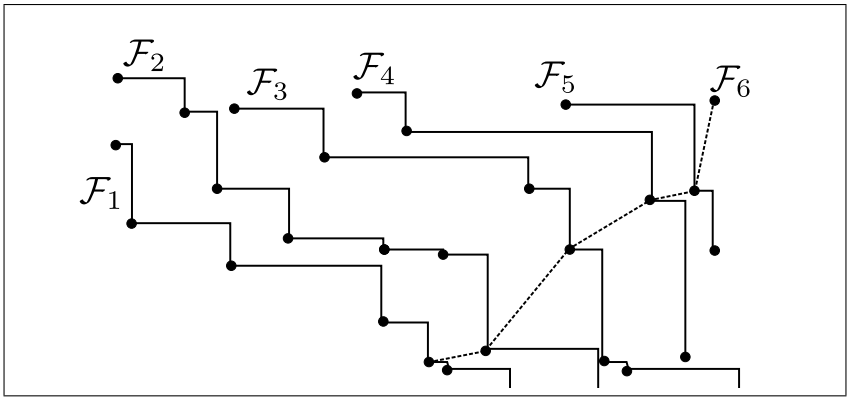


Figure 1.2: Illustration of the equivalence between the longest chain problem and nondominated sorting.

for reference—the chain is the dotted line. This point is on $\mathcal{F}_{k-1}$ because there exists a point on $\mathcal{F}_{k-2}$ less in all coordinates. This is the third to last point on the chain. Continuing by induction we can construct a chain of length exactly k starting on $\mathcal{F}_1$ and ending at $x \in \mathcal{F}_k$. This establishes (1.3).

The equivalence (1.3) says that the Pareto fronts $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \dots$ are encoded into the level sets (or jump sets) of the function U_n. In other words, U_n is a piecewise constant function that counts the nondominated layers. It turns out that the longest chain interpretation of nondominated sorting is very useful in the mathematical analysis of the algorithm. We also mention there are also other interesting connections between nondominated sorting and combinatorics [28, 49], molecular biology [3, 56], graph theory [49], Young Tableaux [28, 68] and even in physical layout problems in the design of integrated circuits [3].

Nondominated sorting has a continuum limit that corresponds to solving a Hamilton-Jacobi equation [11, 12]. One version of the theorem is as follows.

Theorem 1.1. *Assume $f \in C([0,\infty)^d)$. Then there exists a universal constant $C_d > 0$ such that with probability one*

$$n^{-1/d} U_n \longrightarrow C_d\, u \quad \text{locally uniformly on } [0,\infty)^d \text{ as } n \to \infty, \tag{1.4}$$

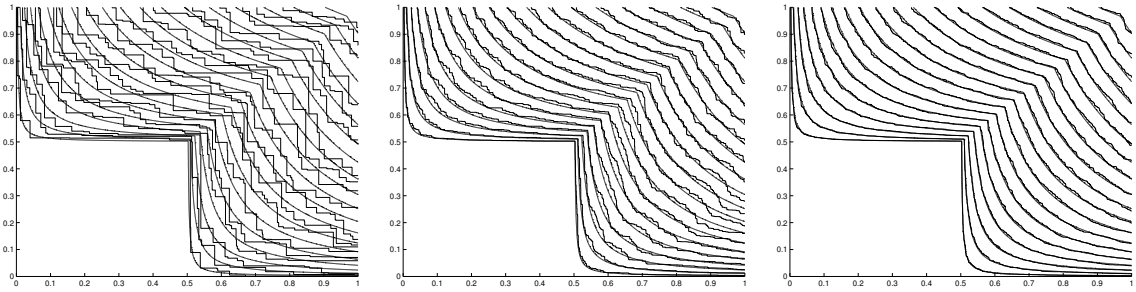
where $u \in C([0,\infty)^d)$ is the unique increasing[1] *viscosity solution of*

$$\begin{cases} u_{x_1} \cdots u_{x_d} = f, & \text{in } \mathbb{R}^d_+ \\ u = 0, & \text{on } \partial\mathbb{R}^d_+. \end{cases} \tag{1.5}$$

Theorem 1.1 states that the Pareto fronts $\mathcal{F}_1, \mathcal{F}_2, \dots$ converge almost surely to the level sets of the function u that satisfies the Hamilton-Jacobi equation (1.5) in the viscosity sense. See Figure 1.3 for an illustration of this continuum limit.

The continuum limit (1.5) opens the door to fast approximate sorting algorithms based on estimating the distribution f of the data and solving the PDE (1.5) numerically [13].

[1] Increasing means that $u(x) \le u(y)$ whenever $x_i \le y_i$ for all i.



(a) $n = 10^3$ (b) $n = 10^5$ (c) $n = 10^6$

Figure 1.3: Example of the continuum limit of nondominated sorting (Theorem 1.1) for $f(x) = 1$ for $x \in [0,1]^2 \setminus [0,0.5]^2$, and $f(x) = 0$ otherwise.

The resulting algorithm is called *PDE-based ranking*, and can be orders of magnitude faster than the discrete sorting problem, depending on the application. The PDE-based ranking algorithm has recently been applied in machine learning to the problem of anomaly detection and classification of streaming data [1] using the finite difference schemes for (1.5) developed recently [10].

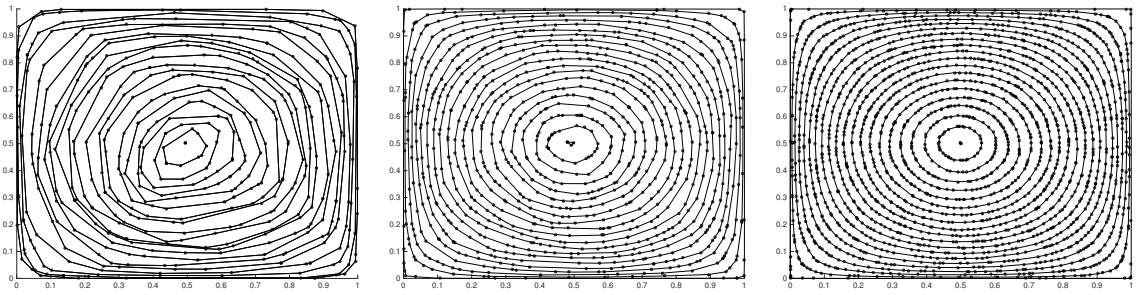
1.2 Convex hull peeling

The ordering of multivariate data is an important and challenging problem in statistics. One dimensional data can be ordered linearly from least to greatest, and the study of the distributional properties of this ordering is the subject of order statistics. An important order statistic is the median, or middle, of the dataset. In statistics, the median is generally preferred over the mean due to its robustness with respect to noise. In dimensions $d \geq 2$, there is no obvious generalization of the one dimensional order statistics, and no obvious candidate for the median. As such, many different types of orderings, and definitions of median, have been proposed for multivariate data. One of the first surveys on the ordering of multivariate data was given by Barnett [7], and more recent surveys are given by Small [60] and Liu [47].

In Barnett's seminal paper [7], he introduced the idea of convex hull ordering. The idea is to sort a finite set $X \subset \mathbb{R}^d$ into convex layers by repeatedly removing the vertices of the convex hull. The process of sorting a set of points into convex layers is called convex hull peeling, convex hull ordering, and sometimes onion-peeling [22]. The index of the convex layer that a sample belongs to is called its *convex hull peeling depth*. This peeling procedure will eventually exhaust the entire dataset, and the *convex hull median* is defined as the centroid of the points on the final convex layer. Convex hull ordering is now extensively used in the field of robust statistics [27, 58], and is particularly useful in outlier detection [39].

Since affine transformations preserve the convexity of sets, the convex layers of a set of points are invariant under affine transformations. Due to this important property, Suk and Fusser [64] use convex hull peeling to match projectively deformed datasets. This is important, for example, in computer vision, where a common task is the recognition of objects viewed from different angles. There are also some interesting applications of convex hull peeling in fingerprint identification [57], and algorithmic drawing [23].

Convex hull peeling has a continuum limit that corresponds to solving a second order



(a) $n = 10^3$ points (b) $n = 10^4$ points (c) $n = 10^5$ points

Figure 1.4: Examples of convex layers for n random points on $[0,1]^2$.

degenerate elliptic PDE in the viscosity sense. We briefly summarize the result here, and it will appear in a forthcoming paper [14]. Let $V \subset \mathbb{R}^d$ be convex, open, and bounded, and let $f \in C(V)$ be bounded and positive on V. Let $V_n : \overline{V} \to \mathbb{R}$ be the *convex depth function* obtained by applying convex hull peeling to the Poisson points X_{nf}. The function V_n satisfies $V_n(x) = k$ if and only if x is in the k^{th} convex layer, and is thus a piecewise constant function that encodes the convex layers into its level sets. We have the following continuum limit.

Theorem 1.2 (J. Calder & C.K. Smart, 2017)**.** *There exists a universal constant α_d such that with probability one*

$$n^{-2/(d+1)} V_n \longrightarrow \alpha_d v \quad \textit{uniformly on } V \textit{ as } n \to \infty,$$

where $v \in C(\overline{V})$ is the unique viscosity solution of

$$\begin{cases} \nabla v \cdot \operatorname{cof}(-\nabla^2 v) \nabla v = f^2, & \text{in } V \\ v = 0, & \text{on } \partial V. \end{cases} \tag{1.6}$$

The solution of the nonlinear PDE (1.6) has the property that its level sets evolve with a normal velocity given by the $(d+1)^{\text{th}}$-root of Gaussian curvature multiplied $f(x)^{2/(d+1)}$. When $f \equiv 1$ this PDE is known as *affine invariant curvature motion* [15], *the affine flow* [4], and in two dimensions *affine curve shortening* [5, 53, 59].

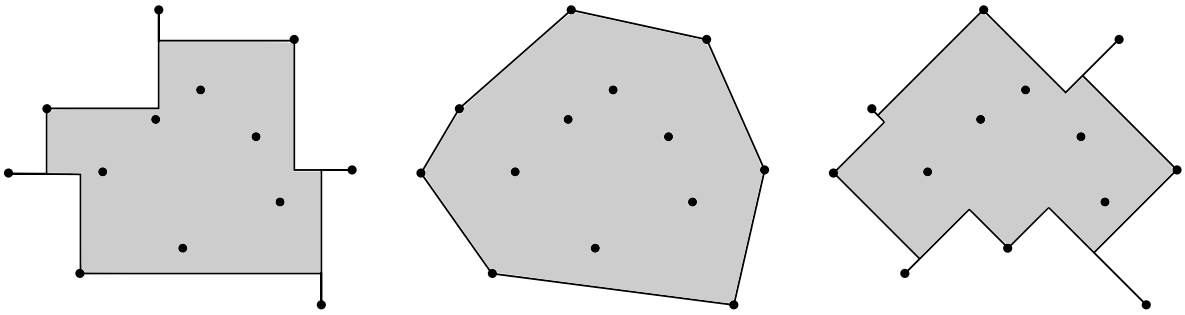
1.3 Pareto envelope peeling

Both non-dominated sorting and convex hull peeling can be unified under a more general peeling framework. Let $\|\cdot\|$ be any norm on $\mathbb{R}^d$ and let $X \subset \mathbb{R}^d$ be a set of points. We say $x \in \mathbb{R}^d$ is *dominated* by $y \in \mathbb{R}^d$ if

$$\forall z \in X,\ \|y - z\| \leq \|x - z\| \text{ and } \exists z \in X,\ \|y - z\| < \|y - x\|.$$

Basically, we are saying y dominates x if y is closer than x to every point in X with respect to the norm $\|\cdot\|$. The set of nondominated points is called the *Pareto envelope* [17, 65] of X, which we denote by

$$\operatorname{Hull}(X) = \text{Set of non-dominated points in } \mathbb{R}^d. \tag{1.7}$$



(a) $\|\cdot\|_1$ (b) $\|\cdot\|_2$ (c) $\|\cdot\|_\infty$

Figure 1.5: Examples of Pareto envelopes for different norms.

The Pareto envelope is sometimes also called the set of *efficient* points, and arises in planar location problems. It turns out that for any *round norm*[2] the Pareto envelope of a set of points coincides with its convex hull [65]. Figure 1.5 shows the Pareto envelope for a set of points in the ℓ_1, ℓ_2 and ℓ_∞ norms.

We can peel a point cloud by repeatedly removing the boundary of the Pareto envelopes. We call this *Pareto envelope peeling* and it can be defined by

$$K_1(X) = \text{Hull}(X), \quad K_{n+1}(X) = \text{Hull}(X \cap \text{Interior}(K_n(X))).$$

If $\|\cdot\|$ is any round norm, then Pareto envelope peeling is equivalent to convex hull peeling. If $\|\cdot\| = \|\cdot\|_1$ is the ℓ_1 norm, then Pareto envelope peeling is (almost) equivalent to non-dominated sorting. The difference is that Pareto envelope peeling allows the peeling to come from any of 2^d different directions, while nondominated sorting only allows peeling from the lower left side of the point cloud. If we sort a Poisson point process on $\mathbb{R}^d_+$ then Pareto envelope peeling and nondominated sorting are equivalent since the structure of the point cloud prohibits peeling from all but the lower left direction. Figure 1.6 shows an example of Pareto envelope peeling of random points with respect to the ℓ_1 norm. Notice the flow transforms the square into the ℓ_1 ball before vanishing.

Based on the continuum limit in Theorem 1.1, it is reasonable to conjecture that the continuum limit of ℓ_1-peeling is basically the same PDE

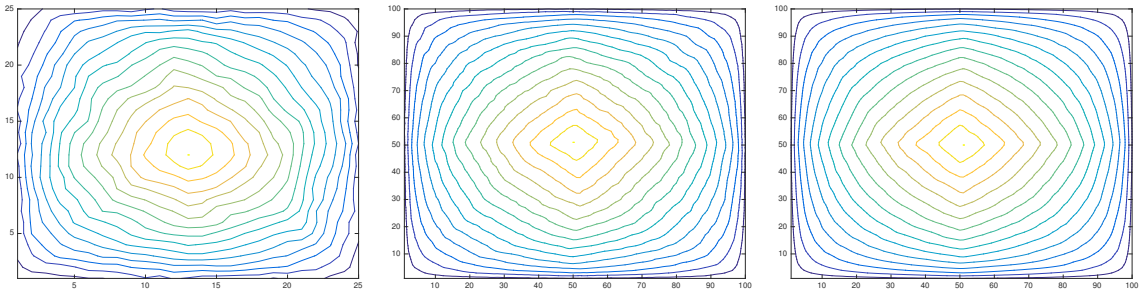
$$u_{x_1}^2 u_{x_2}^2 \cdots u_{x_d}^2 = f^2,$$

except now we do not require u to be increasing. It is an interesting question to determine the continuum PDEs for other non-round norms. For example, what are the continuum PDEs for block norms whose unit balls are convex polytopes?

1.4 Graph peeling

Our final example of a data peeling algorithm is *graph peeling*. Let (X, W) be a weighted graph. The set $X \subset \mathbb{R}^d$ are the vertices, and $W = (w_{xy})_{x,y\in X}$ are nonnegative edge weights. A weight of $w_{xy} = 0$ indicates that x and y are not connected. The weights usually represent a

[2] By round, we mean the unit ball is smooth and uniformly convex.



(a) $n = 10^4$ points (b) $n = 10^6$ points (c) $n = 10^7$ points

Figure 1.6: Example of ℓ_1 peeling for random points.

similarity between data samples. When $w_{xy} \approx 1$, x and y are very similar, and when $w_{xy} \approx 0$, x and y are very dissimilar. The weights can take discrete or continuous values. One common way to select the weights for machine learning or data science applications is

$$w_{xy} = \Phi\left(\frac{|x-y|}{h}\right)$$

where $h > 0$ is a length scale and Φ is a nonnegative decreasing function.

The *degree* of a vertex $x \in X$ is defined as

$$\deg(x) = \sum_{y \in X} w_{xy}.$$

The *graph k-peeling* algorithm arranges X into layers by repeatedly removing vertices with $\deg(x) < k$, and their connecting edges. When the edges are removed, the degrees of neighboring vertices are reduced and additional vertices may then have degree less than k and be removed in a later stage of the peeling. The peeling stops when all remaining vertices have degree at least k, or when the entire graph has been peeled away. The remaining vertices, if there are any, are called the *k-core* of the graph, and represent a significant component of the graph. The *peeling depth* of a vertex $x \in X$ is the smallest k such that x is removed by k-peeling. For references on graph peeling we refer the reader to [2, 41, 42].

Figure 1.7 shows colormaps of graph peeling for three different realizations. We used $n = 10^4$ random points on the box $[0,1]^d$, and set the weights to be $w_{xy} = 1$ if $|x-y| \leq h$ and $w_{xy} = 0$ otherwise, where $h = 1/4$. We set k to be exactly half of the expected number of points in the ball $B(x,h)$, that is $k = n(\pi h^2/2)$. The colors from red to blue indicate when the points were removed by the peeling algorithm (earlier vs later). The first and second simulations peeled away the entire graph, while the third simulation (on the right) reached a non-trivial k-core shown in blue. We note that each realization had distinct random features from the others, indicating that the stochastic elements may not wash away in any continuum limit. Figure 1.8 shows a single simulation of graph peeling.

It is interesting to ask what the continuum limit may be for graph peeling. In some special cases one would expect to get (stochastic) mean curvature motion, since the algorithm is similar to the diffusion generated curvature motion, also called threshold dynamics, of Merriman, Bence and Osher [51]. However, depending on the value of k and the local density of points, the macroscopic velocity of the peeling may often be dominated by first order terms and one may get a Hamilton-Jacobi equation.

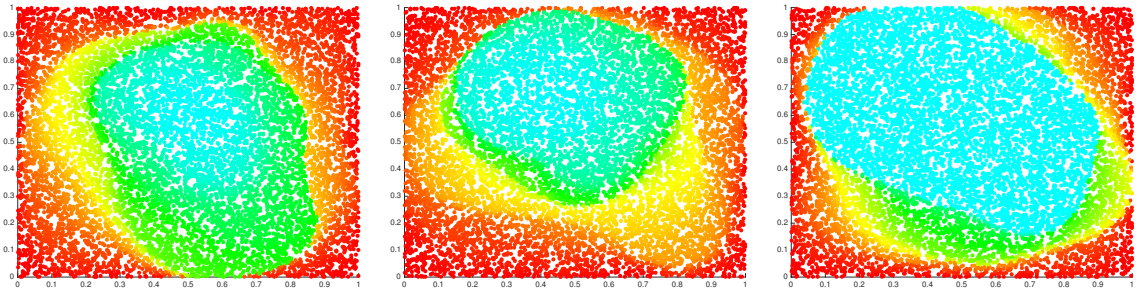


Figure 1.7: Colormaps for three different realizations of graph peeling. The colors from red to blue indicate the stage at which a point was peeled away.

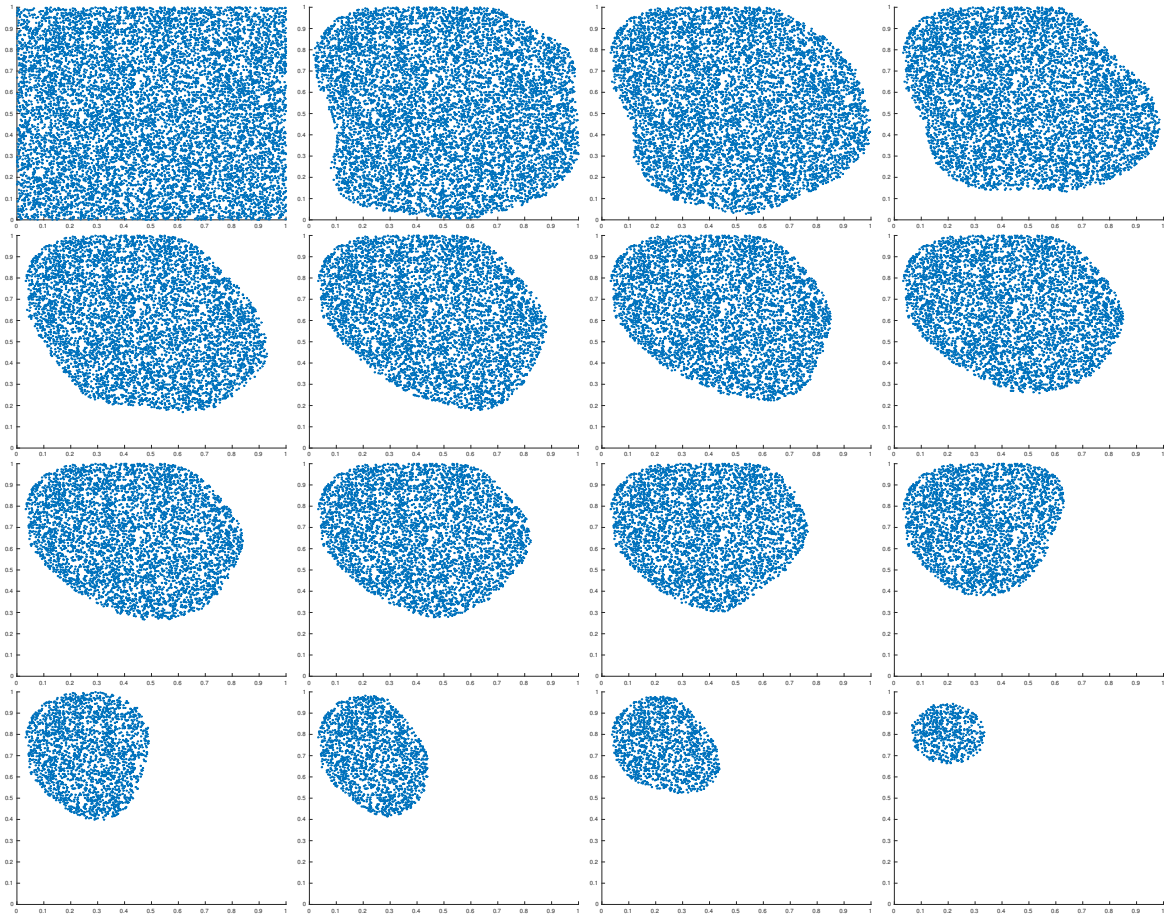


Figure 1.8: An illustration of graph peeling for a random graph.

1.5 Overview of minicourse

In this minicourse, we plan to give an entirely self-contained proof of the continuum limit of nondominated sorting stated in Theorem 1.1. The proof is based on [11] with some simplifications noticed recently by the author. At a high level the proof is split into three parts. First, in Section 3 we study a local problem of the longest chain among random points in the box $[0,1]^d$. The results in Section 3 are classical in probability theory and originally appeared in [8, 9, 34]. In Section 4 we introduce viscosity solutions and prove uniqueness of viscosity solutions of (1.5). Finally, in Section 5 we give the main portion of the proof of Theorem 1.1. The proof is based on establishing Hölder estimates on the sequence U_n, which enables us to prove uniform convergence of subsequences to Hölder continuous functions. The linchpin in the proof is showing that these uniform limits are in fact viscosity solutions of (1.5), and then uniqueness of viscosity solutions ensures that the limits of all subsequences are the same, and so the entire sequence converges.

Let us pause to give a few words of advice to the graduate student beginning their career in mathematics. There are at a broad level two important components of mathematical research. The first is to master the technical tools of a specific field so that you are ready to apply them to new research problems. The second, and arguably more important, is the "how" of mathematical research; that is, how does one discover good research problems and come up with new theorems to prove. This second part of mathematical research is deeply involved with arguing nonrigorously at an advanced level. To give an idea of how nonrigorous arguments are essential in research, we give in Section 2 two nonrigorous derivations of the PDE (1.5) in Theorem 1.1. The reader will quickly notice the gaps in our logic, but this is besides the point. The nonrigorous argument tells you what might (or should) be true, and produces a problem for us to study rigorously. It may sometimes, but not always, also suggest an avenue for a rigorous proof.

2 Two nonrigorous heuristics

We give in this section two nonrigorous arguments deriving the PDE (1.5) as the continuum limit of nondominated sorting. While the arguments are not rigorous, they give the essence of the result and provide a guideline for the construction of a rigorous proof.

2.1 PDE argument

We first give an argument using what we call PDE techniques, to distinguish them from the variational techniques used in the second heuristic argument in Section 2.2. We suppose we do not even know the growth rate of $n^{1/d}$. Thus, we will assume that for some $\alpha \in (0,1]$

$$n^{-\alpha} U_n \longrightarrow u \quad \text{as } n \to \infty$$

uniformly on $\mathbb{R}^d$, where $u \in C^1(\mathbb{R}^d)$. Implicit in this assumption is that the number of Pareto fronts among n random variables grows on the order of n^α, and we will use this throughout the argument. The end goal is to identify a PDE that u satisfies and the growth rate α.

Let us fix a point $x \in \mathbb{R}^d$ and zoom in to the level set $\{u = u(x)\}$ near x, which looks locally like a plane. Since u is the limit of $n^{-\alpha} U_n$, u must be increasing in each coordinate and so we also assume $u_{x_i} > 0$ for all i. Fix a vector $v \in \mathbb{R}^n$ with $\nabla u(x) \cdot v > 0$ and consider

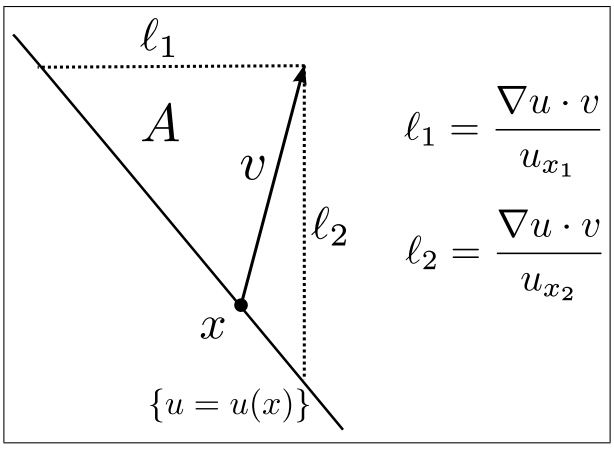


Figure 2.1: Illustration of some quantities from the nonrigorous derivation of (1.5).

the set A of points $y \in \mathbb{R}^d$ such that $u(y) > u(x)$ and $y_i \le x_i + v_i$ for all i (see Figure 2.1). The argument reduces to counting how many Pareto fronts pass through A. We note that in dimensions $d \ge 3$ the set A is a simplex with orthogonal corner at $x+v$. Let $\ell_1, \dots, \ell_d$ denote the side lengths of A, which satisfy

$$(v - \ell_i e_i) \cdot \nabla u(x) \approx 0,$$

for $|v|$ small. Therefore

$$\ell_i \approx \frac{\nabla u(x) \cdot v}{u_{x_i}(x)},$$

and up to a constant, the measure of A is

$$|A| \approx C \frac{(\nabla u(x) \cdot v)^d}{u_{x_1}(x) \cdots u_{x_d}(x)}.$$

When $|v|$ is small, the number of points $N(A)$ falling in A is (in expectation)

$$N(A) \approx n \int_A f \, dx \approx n|A| f(x) \approx C \left(\frac{f(x) n}{u_{x_1}(x) \cdots u_{x_d}(x)} \right) (\nabla u(x) \cdot v)^d.$$

Now, we claim we can count the number of Pareto fronts passing through A by considering only the Poisson points falling in A. The reason for this is that the level set $\{u = u(x)\}$ should be approximately close to a Pareto front, and so we can discard all points y with $u(y) < u(x)$. Then any point y with $u(y) > u(x)$ and $y_i > x_i + v_i$ for some i cannot affect the partial ordering of points within A, so any such points can also be discarded. Furthermore, when $|v|$ is small, the $N(A)$ points in A are roughly uniformly distributed (if f is smooth, so roughly constant on A), and we can scale A into a simplex with unit side lengths without changing the ordering of points in A by dilating along each coordinate axis. This all suggests that the number of Pareto fronts passing through A should be asymptotic to

$$C N(A)^\alpha \approx C \left(\frac{f(x) n}{u_{x_1}(x) \cdots u_{x_d}(x)} \right)^\alpha (\nabla u(x) \cdot v)^{d\alpha}.$$

The rest of the argument is calculus:

$$\begin{aligned}\nabla u(x)\cdot v &\approx u(x+v)-u(x)\\ &\approx n^{-\alpha}(U_n(x+v)-U_n(x))\\ &= n^{-\alpha}(\#\text{ of Pareto fronts passing through } A)\\ &\approx Cn^{-\alpha}\left(\frac{f(x)n}{u_{x_1}(x)\cdots u_{x_d}(x)}\right)^{\alpha}(\nabla u(x)\cdot v)^{d\alpha}\\ &= C\left(\frac{f(x)}{u_{x_1}(x)\cdots u_{x_d}(x)}\right)^{\alpha}(\nabla u(x)\cdot v)^{d\alpha}. \qquad (2.1)\end{aligned}$$

Since $v\in\mathbb{R}^d$ is arbitrary, we must have $d\alpha=1$, or $\alpha=1/d$. This tells us the growth rate should be $n^{1/d}$. We can now cancel $\nabla u\cdot v$ from both sides above to find that u must satisfy

$$u_{x_1}\cdots u_{x_d}=Cf.$$

Many of the steps in the argument above were purposefully not rigorous. It is important, however, to have an intuition about which steps are easy to make rigorous, and which are more limiting. The assumption that $u^{-\alpha}U_n$ converges uniformly to a function u is not terribly limiting, since we may replace it with convergence of a subsequence provided we establish some kind of compactness of the sequence $n^{-\alpha}U_n$. Since the limit u must be increasing (i.e., $u_{x_i}\geq 0$), the assumption that $u_{x_i}>0$ is also not limiting, since we can play a trick by bending the function u upwards so that it has strictly positive derivatives (see Section 5). The other steps, like the approximations of the area $|A|$ and the number of points $N(A)$ are all exact as $|v|\to 0$, and we can easily control the error, so these are not limiting.

The key gap in the argument is actually the assumption that the limit u is C^1, since this does not hold in general. The best we can hope for is Lipschitz continuity of u, and this allows the level sets $\{u=u(x)\}$ to have corners. A key part of our argument was that the shape of A is the same everywhere, so this is a serious problem. Overcoming this problem requires the viscosity solution framework that is introduced in Section 4. The idea, roughly, is that we perform the argument above on smooth test functions instead of on u directly. This is subtle and requires some work to make rigorous. We note the argument given above is the basis for the rigorous proof of Theorem 1.1 given in Sections 3, 4 and 5 of these notes, and in [11].

2.2 Variational argument

We give in this section another nonrigorous heuristic for deriving the PDE (1.5) in Theorem 1.1. Let $x\in\mathbb{R}^d_+$ and let $\mathbf{x}(t)=(x_1(t),\ldots,x_d(t))$ be an increasing (i.e., $x_i'(t)>0$ for all i) curve with $\mathbf{x}(0)=0$ and $\mathbf{x}(1)=x$. We aim to count the length of a longest chain near the curve $\mathbf{x}(t)$.

We again presume we do not know the order of growth $n^{1/d}$. Suppose all we know is that for n independent and uniformly distributed random variables in a box, the length of a longest chain is approximately c_dn^{α} for some $\alpha\in(0,1]$. Choose a time step $\Delta t=1/K$, set $t_j=j\Delta t$ and define the rectangles

$$R_j=\{x\in[0,\infty)^d\,:\,x_i(t_{j-1})\leq x_i<x_i(t_j)\quad\text{for all } i\}.$$

See Figure 2.2 for a depiction of the curve $\mathbf{x}(t)$ and the rectangles.

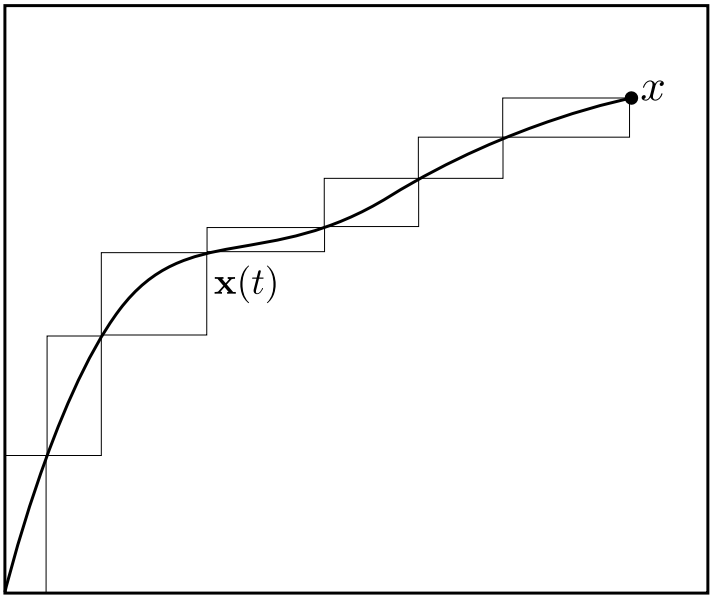


Figure 2.2: Depiction of some quantities from the variational derivation of the PDE (1.5).

If Δt is small, then the Poisson points within each rectangle R_j are approximately uniformly distributed. Let N_j denote the number of Poisson points in $|R_j|$. Then there should be a chain within each R_j of length approximately $c_d N_j^\alpha$. We can concatenate the chains from all rectangles to find that

$$U_n(x) \geq c_d \sum_{j=1}^{K} N_j^\alpha.$$

Now, $N_j \approx f(\mathbf{x}(t_j))|R_j|n$, and the measure of the rectangle R_j can be approximated by

$$|R_j| = x_1'(t_j) \cdots x_d'(t_j) \Delta t^d.$$

Thus we have

$$U_n(x) \geq c_d \sum_{j=1}^{K} f(\mathbf{x}(t_j))^\alpha (x_1'(t_j) \cdots x_d'(t_j))^\alpha \Delta t^{\alpha d}. \tag{2.2}$$

If $\alpha = 1/d$, then this is exactly a Riemann sum for the integral

$$J(\mathbf{x}) := c_d \int_0^1 f(\mathbf{x}(t))^{1/d} (x_1'(t) \cdots x_d'(t))^{1/d} \, dt. \tag{2.3}$$

If $\alpha < 1/d$, then the right hand side of (2.2) goes to ∞ as $\Delta t \to 0$, while if $\alpha > 1/d$, then the right hand side vanishes as $\Delta t \to 0$. This suggests that $\alpha = 1/d$ (at least this is the only interesting value).

This also suggests that in the limit as $n \to \infty$, finding the length of the longest chain should be equivalent to maximizing $J(\mathbf{x})$ over all monotone curves $\mathbf{x}$. More precisely, we expect that as $n \to \infty$, $n^{-1/d} U_n \to u$ where

$$u(x) = \sup \big\{ J(\mathbf{x}) \, : \, \mathbf{x}(0) = 0, \, \mathbf{x}(1) = x, \text{ and } x_i'(t) > 0 \text{ for all } i \big\}. \tag{2.4}$$

We note this variational problem originally appeared in [26] for the two-dimensional longest chain problem.

We claim that u defined by (2.4) solves the Hamilton-Jacobi equation (1.5). To see this, note by Hölder's inequality that

$$\int_0^1 (x_1'(t)\cdots x_d'(t))^{1/d}\,dt \le \left(\prod_{i=1}^d \int_0^1 x_i'(t)\,dt\right)^{1/d} = \prod_{i=1}^d (x_i(1)-x_i(0))^{1/d},$$

with equality if $\mathbf{x}(t)$ is a straight line. Therefore, at least locally, straight lines maximize $J(\mathbf{x})$. Therefore, for $r>0$ small we have the dynamic programming principle

$$u(x) \approx \sup\left\{u(y) + c_d f(x)^{1/d} \,:\, \prod_{i=1}^d (x_i-y_i)^{1/d} = r \text{ and } y_i \le x_i \text{ for all } i\right\}.$$

Expanding u in a Taylor expansion we have

$$\inf\left\{\nabla u(x)\cdot(x-y) \,:\, \prod_{i=1}^d (x_i-y_i)^{1/d} = r \text{ and } y_i \le x_i \text{ for all } i\right\} \approx c_d f(x)^{1/d}.$$

Dividing by $r>0$ and setting $a=(x-y)/r$ yields

$$\inf\left\{\nabla u(x)\cdot a \,:\, a_1\cdots a_d = 1 \text{ and } a_i > 0 \text{ for all } i\right\} \approx c_d f(x)^{1/d}. \tag{2.5}$$

By the Lagrange multiplier method, the optimal a satisfies

$$a_i u_{x_i} = \lambda \quad \text{for all } i,$$

where $\lambda>0$ is the Lagrange multiplier. Therefore $\lambda^d = u_{x_1}\cdots u_{x_d}$ and

$$a_i = \frac{(u_{x_1}\cdots u_{x_d})^{1/d}}{u_{x_i}}.$$

Plugging this back into (2.5) yields

$$d(u_{x_1}\cdots u_{x_d})^{1/d} = c_d f(x)^{1/d},$$

which is the Hamilton-Jacobi equation (1.5), up to a constant. We note that the connection between the variational problem (2.4) and the Hamilton-Jacobi equation (1.5) is a special case of optimal control theory [6]

It is possible to construct a rigorous proof of Theorem 1.1 using the ideas above, and this was in fact the original proof that appeared in [12]. The majority of the proof centers around the variational problem (2.4), and the PDE (1.5) makes a surprise appearance at the end.

3 The local problem

We first study the local problem of the longest chain in a box. Let X_n be a Poisson point process with intensity n on $\mathbb{R}^d$ and let

$$L_n = \ell(X_n \cap [0,1]^d). \tag{3.1}$$

Establishing a law of large numbers for L_n is a classical application of subadditive ergodic theory. The original ideas were developed by Hammersley [34]. In this section we prove the following Theorem.

Theorem 3.1 (Hammersley, 1972). *There exists a constant $c_d > 0$ such that with probability one*

$$\lim_{n\to\infty} \frac{L_n}{n^{1/d}} = c_d. \tag{3.2}$$

Before giving the proof of Theorem 3.1, we note a useful corollary.

Corollary 1. *For any rectangle $R \subset \mathbb{R}^d$ we have*

$$\lim_{n\to\infty} \frac{\ell(X_n \cap R)}{n^{1/d}} = c_d |R|^{1/d} \quad \text{almost surely,} \tag{3.3}$$

where $|R|$ denotes the Lebesgue measure of R.

Proof. By scaling the rectangle R into the unit rectangle, we find that $\ell(X_n \cap R)$ has the same distribution as $L_{n|R|}$. Theorem 3.1 gives

$$\lim_{n\to\infty} \frac{\ell(X_n \cap R)}{n^{1/d}|R|^{1/d}} = c_d. \qquad \square$$

The proof of Theorem 3.1 is split into two parts.

3.1 Convergence in mean

We first prove convergence of the mean value $\mathbb{E}[L_n]$. A key tool in the proof is the following sub- (or rather super-) additivity lemma.

Lemma 3.1. *Suppose $f : [0,\infty) \to [0,\infty)$ satisfies*

$$f(s) + f(t) \le f(s+t) \quad \text{for all } s,t \in [0,\infty). \tag{3.4}$$

Then there exists $c \in [0,\infty]$ such that $f(t) \le ct$ for all $t > 0$ and

$$\lim_{t\to\infty} \frac{f(t)}{t} = c. \tag{3.5}$$

Proof. Let $c = \limsup_{t\to\infty} \frac{f(t)}{t}$ and assume first that $c < \infty$. We first prove that

$$f(t) \le ct \quad \text{for all } t > 0. \tag{3.6}$$

To see this, let $\varepsilon > 0$ and choose $s > 0$ such that $\frac{f(\tau)}{\tau} \le c + \varepsilon$ for all $\tau \ge s$. By iterating the superadditivity property (3.4) we have

$$mf(t) \le f(mt) \quad \text{for all } t \in \mathbb{R},\ m \in \mathbb{N}. \tag{3.7}$$

Therefore

$$\frac{f(t)}{t} \le \frac{f(mt)}{mt} \le c + \varepsilon,$$

for a sufficiently large choice of $m \in \mathbb{N}$. This establishes the claim.

We now prove the limit (3.5) holds. Let $\varepsilon > 0$ and choose $s > 0$ such that $\frac{f(s)}{s} \geq c - \varepsilon$. Let $t > s$ and choose $m \in \mathbb{N}$ such that $t = ms + \tau$, where $0 \leq \tau \leq s$. By (3.4) and (3.7) we have

$$f(t) \geq f(ms) + f(\tau) \geq mf(s).$$

Therefore

$$\frac{f(t)}{t} \geq \frac{ms}{t}\frac{f(s)}{s} \geq \frac{ms}{t}(c - \varepsilon).$$

Sending $t \to \infty$ we have $ms/t \to 1$ and we obtain

$$\liminf_{t\to\infty} \frac{f(t)}{t} \geq c - \varepsilon.$$

This completes the proof for $c < \infty$. The proof for $c = \infty$ is similar. □

Remark 3.1. When using Lemma 3.1, one must show that the constant c is not trivial, that is, one needs to establish that $c > 0$ and $c < \infty$. Since the lemma shows that $f(t) \leq ct$ for all $t > 0$, we know that $c > 0$ whenever f is not identically zero. This is presumably easy to check. Showing that $c < \infty$ is usually the more challenging direction, and is often proved by establishing that $f(t) \leq Ct$ for some larger constant $C > c$.

Since the growth rate of $n^{1/d}$ is not linear in n, we cannot expect L_t to be superadditive. Hence, the key to using Lemma 3.1 is to restate the problem in terms of a new parameter $t = n^{1/d}$ (or $n = t^d$). In particular, we claim that $f(t) := \mathbb{E}[L_{t^d}]$ is superadditive. To see this, let $s, t > 0$ and set $n = (s + t)^d$ and $\lambda = s/(s + t)$. Since we can concatenate maximal chains in $[0, \lambda)^d \cap X_n$ and $[\lambda, 1]^d \cap X_n$ to obtain a chain in $[0, 1]^d \cap X_n$ we have

$$\ell([0, \lambda)^d \cap X_n) + \ell([\lambda, 1]^d \cap X_n) \leq \ell([0, 1]^d \cap X_n) = L_n. \tag{3.8}$$

See Figure 3.1 for an illustration. The inequality can be strict if the longest chain in $[0, 1]^d \cap X_n$ contains points not in either sub-rectangle—this is illustrated by the dotted line in the figure. The number of Poisson points in $[0, \lambda)^d$ is a Poisson random variable with mean $\lambda^d n = s^d$. Therefore A has the same distribution as L_{s^d}. Similarly, B has the same distribution as L_{t^d} and taking expectations in (3.8) yields

$$f(s) + f(t) \leq f(s + t).$$

By Lemma 3.1 there exists $c_d > 0$ (possibly $c_d = \infty$) such that $\lim_{t\to\infty} \frac{f(t)}{t} = c_d$. Since $t = n^{1/d}$ this gives

$$\lim_{n\to\infty} \frac{\mathbb{E}[L_n]}{n^{1/d}} = c_d. \tag{3.9}$$

We can also use Lemma 3.1 to show that $\mathbb{E}[L_n] \leq c_d n^{1/d}$ for all n.

The last step is to prove that $c_d < \infty$. For this, we condition on the number of points N in the Poisson point process $X_n \cap [0, 1]^d$. Recall N is a Poisson random variable with mean n.

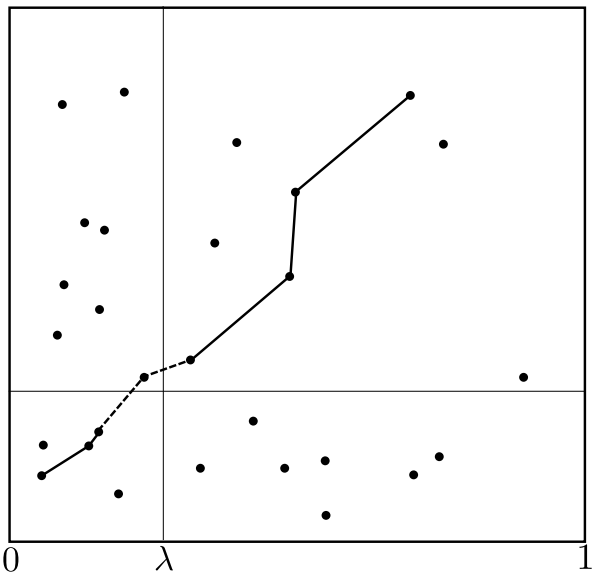


Figure 3.1: Illustration of the superadditivity property (3.8) for the longest chain problem.

For $\varepsilon > 0$ small we have

$$\begin{aligned}
\mathbb{E}[L_n] &\leq \mathbb{E}(L_n \,|\, N \leq (1+\varepsilon)n)\, \mathbb{P}(N \leq (1+\varepsilon)n) + \sum_{k>(1+\varepsilon)n} k\, \mathbb{P}(N = k) \\
&\leq \sum_{j \leq (1+\varepsilon)n} \mathbb{P}(L_n \geq j \,|\, N \leq (1+\varepsilon)n) + \left(\sum_{k=1}^{\infty} k^2\, \mathbb{P}(N = k) \right)^{1/2} \left(\sum_{k>(1+\varepsilon)n} \mathbb{P}(N = k) \right)^{1/2} \\
&\leq k + 2n\, \mathbb{P}(L_n \geq k \,|\, N \leq (1+\varepsilon)n) + \mathrm{Var}(N)^{1/2}\, \mathbb{P}(N > (1+\varepsilon)n) \\
&\leq k + 2n\, \mathbb{P}(L_n \geq k \,|\, N \leq (1+\varepsilon)n) + n^{1/2} e^{-n\varepsilon^2/2},
\end{aligned}$$

for any $k \geq 1$, where we used standard Poisson tail bounds in the last line. Now, any subset of X_n of size k is a chain with probability $1/k!^{d-1}$. Conditioned on $N \leq (1+\varepsilon)n$ there are at most $\binom{m}{k}$ subsets of size k, where m is the least integer greater than $(1+\varepsilon)n$. By Stirling's approximation $k! \geq (k/e)^k$ and the union bound we have

$$\mathbb{P}(L_n \geq k \,|\, N \leq (1+\varepsilon)n) \leq \binom{m}{k} \frac{1}{k!^{d-1}} \leq \frac{m^k}{k!^d} \leq \left(\frac{me^d}{k^d} \right)^k.$$

Choosing $k^d = m(1+\varepsilon)e^d$ we have

$$\mathbb{E}[L_n] \leq e(1+\varepsilon)^{1/d} m^{1/d} + 2n(1+\varepsilon)^{-k} + n^{1/2} e^{-n\varepsilon^2/2}.$$

Sending $n \to \infty$ we find that

$$\limsup_{n\to\infty} \frac{\mathbb{E}[L_n]}{n^{1/d}} \leq (1+\varepsilon)^{1/d} e$$

for every $\varepsilon > 0$. Therefore c_d is finite, and in fact $c_d \le e$. With more work [9] it is possible to show that

$$\frac{d^2}{d!^{1/d}\Gamma(\frac{1}{d})} \le c_d < e$$

for all d, and that $c_2 = 2$ [48, 67]. Another application of Stirling's approximation shows that $c_d \to e$ as $d \to \infty$.

3.2 Almost sure convergence

We can immediately upgrade (3.9) to almost sure convergence by invoking Kingman's subadditive ergodic theory [43]. However, we take another path here, in order to introduce other important tools from probability.

A well-traveled path from convergence in mean to almost sure convergence is controlling the fluctuations between L_n and $\mathbb{E}[L_n]$. If we can show that for any $\varepsilon > 0$

$$\sum_{n=1}^{\infty} \mathbb{P}\left(|L_n - \mathbb{E}[L_n]| > \varepsilon n^{1/d}\right) < \infty,$$

then Theorem 3.1 follows from (3.9) and the Borel-Cantelli Lemma. Sharp estimates on the probability $P(|Y - \mathbb{E}[Y]| > \varepsilon)$ are referred to as concentration of measure.

A useful tool for proving concentration of measure is Azuma's inequality, also known as the *method of bounded differences* [63], which is a generalization of the Chernoff-Hoeffding bounds [18, 40]. To describe the setup, let $X_1, \dots, X_n$ be a sequence of independent random variables and set

$$Y_n = f(X_1, \dots, X_n), \tag{3.10}$$

where f satisfies the *bounded differences* condition

$$|f(x_1, \dots, x_i, \dots, x_n) - f(x_1, \dots, \widetilde{x}_i, \dots, x_n)| \le r, \tag{3.11}$$

for some $r > 0$ and all $i = 1, \dots, n$. The bounded differences condition says that f does not depend too much on any one of its entries.

Example 1 (Coin toss)**.** Suppose that $X_1, \dots, X_n$ are the results of a sequence of n coin tosses, so each X_i takes values 1 (heads) and 0 (tails) with equal probability. Suppose $Y_n = X_1 + \cdots + X_n$ is the number of heads. Then $f(x_1, \dots, x_n) = x_1 + \cdots + x_n$ satisfies the bounded differences condition with $r = 1$.

The bounded differences condition gives us tight concentration of Y_n about its mean.

Theorem 3.2 (Azuma's inequality)**.** *For any $\lambda > 0$ we have*

$$\mathbb{P}(|Y_n - \mathbb{E}[Y_n]| > \lambda) \le 2\exp\left(\frac{-\lambda^2}{2nr^2}\right). \tag{3.12}$$

Remark 3.2. If we choose $\lambda = t\sqrt{nr^2}$ in Azuma's inequality we get

$$\mathbb{P}(|Y_n - \mathbb{E}[Y_n]| > tr\sqrt{n}) \le 2\exp\left(-t^2/2\right).$$

Hence, Azuma's inequality says that the fluctuations between Y_n and its mean $\mathbb{E}[Y_n]$ are with high probability smaller than $O(r\sqrt{n})$. This only gives useful information when $\mathbb{E}[Y_n] \gg \sqrt{n}$.

Proof. Let $\mathcal{F}_i = \sigma(X_1, \dots, X_i)$ be the σ-algebra generated by $X_1, \dots, X_i$ and define the differences

$$Z_i := \mathbb{E}[Y_n \,|\, \mathcal{F}_i] - \mathbb{E}[Y_n \,|\, \mathcal{F}_{i-1}],$$

where $\mathcal{F}_0 = \{\varnothing, \Omega\}$ is the trivial sigma algebra. The idea is that we are incrementally adding information about each X_i and observing how much the conditional expectation can change by. Since $\mathbb{E}[Y_n \,|\, \mathcal{F}_n] = Y_n$ and $\mathbb{E}[Y_n \,|\, \mathcal{F}_0] = \mathbb{E}[Y_n]$ we have a telescoping sum

$$Y_n - \mathbb{E}[Y_n] = \sum_{i=1}^{n} Z_i. \tag{3.13}$$

By the law of conditional expectation

$$\mathbb{E}[Z_i \,|\, \mathcal{F}_{i-1}] = \mathbb{E}[\mathbb{E}[Y_n \,|\, \mathcal{F}_i] - \mathbb{E}[Y_n \,|\, \mathcal{F}_{i-1}] \,|\, F_{i-1}] = \mathbb{E}[Y_n \,|\, \mathcal{F}_{i-1}] - \mathbb{E}[Y_n \,|\, \mathcal{F}_{i-1}] = 0. \tag{3.14}$$

It follows that

$$\mathbb{E}[Z_{i_1} Z_{i_2} \cdots Z_{i_k}] = 0 \tag{3.15}$$

for any $1 \le i_1 < i_2 < \cdots < i_k \le n$. Indeed, we simply condition on $\mathcal{F}_{i_k - 1}$ and apply (3.14) to find

$$\mathbb{E}[Z_{i_1} Z_{i_2} \cdots Z_{i_k}] = \mathbb{E}[\mathbb{E}[Z_{i_1} Z_{i_2} \cdots Z_{i_k} \,|\, \mathcal{F}_{i_k - 1}]] = \mathbb{E}[Z_{i_1} Z_{i_2} \cdots Z_{i_{k-1}} \mathbb{E}[Z_{i_k} \,|\, \mathcal{F}_{i_k - 1}]] = 0.$$

By Markov's inequality

$$\mathbb{P}(Y_n - \mathbb{E}[Y_n] > \lambda) = \mathbb{P}\left(e^{\sum_{i=1}^n Z_i t} > e^{\lambda t}\right) \le e^{-\lambda t} \mathbb{E}\left[e^{\sum_{i=1}^n Z_i t}\right] = e^{-\lambda t} \mathbb{E}\left[\prod_{i=1}^{n} e^{Z_i t}\right], \tag{3.16}$$

for any $t > 0$. We now use the swapping trick and the bounded differences condition to bound the Z_i. Let $\widetilde{X}_i$ be an independent copy of X_i. Then

$$\begin{aligned} |Z_i| &= |\mathbb{E}[f(X_1, \dots, X_i, \dots, X_n) \,|\, \mathcal{F}_i] - \mathbb{E}[f(X_1, \dots, X_i, \dots, X_n) \,|\, \mathcal{F}_{i-1}]| \\ &= |\mathbb{E}[f(X_1, \dots, X_i, \dots, X_n) \,|\, \mathcal{F}_i] - \mathbb{E}[f(X_1, \dots, \widetilde{X}_i, \dots, X_n) \,|\, \mathcal{F}_i]| \\ &\le \mathbb{E}[|f(X_1, \dots, X_i, \dots, X_n) - f(X_1, \dots, \widetilde{X}_i, \dots, X_n)| \,|\, \mathcal{F}_i] \le r. \end{aligned}$$

It follows from convexity of the exponential function e^{ax} that

$$e^{ax} \le (x+1)\left(\frac{e^a - e^{-a}}{2}\right) + e^{-a} = x \sinh a + \cosh a,$$

for any $a > 0$ and $x \in [-1, 1]$. Using $x = Z_i / r$ and $a = tr$ we have

$$e^{Z_i t} \le \frac{Z_i}{r} \sinh(tr) + \cosh(tr).$$

It follows from (3.16) and the multiplicative property (3.15) that

$$\mathbb{P}(Y_n - \mathbb{E}[Y_n] > \lambda) \le e^{-\lambda t} \mathbb{E}\left[\prod_{i=1}^{n} \left(\frac{Z_i}{r} \sinh(tr) + \cosh(tr)\right)\right] = e^{-\lambda t} \cosh^n(tr) \le e^{-\lambda t + n t^2 r^2 / 2},$$

where we used the inequality $\cosh x \le e^{x^2/2}$ above. Optimizing over t yields $t = \lambda / (nr^2)$ and

$$\mathbb{P}(Y_n - \mathbb{E}[Y_n] > \lambda) \le \exp\left(\frac{-\lambda^2}{2nr^2}\right). \qquad \square$$

The addition or removal of a single point X_i changes the length of a longest chain by at most 1. Hence, the longest chain among a sequence of *i.i.d.* random variables satisfies the bounded differences condition (3.11) with $r = 1$. However, by Remark 3.2 a direct application of Azuma's inequality is only useful when $\mathbb{E}[L_n] \gg \sqrt{n}$, and we know from Section 3.1 that $\mathbb{E}[L_n] \sim c_d n^{1/d} \leq c_d\sqrt{n}$ for all $d \geq 2$. Thus, a naive direct application of Azuma's inequality is not useful here.

Nevertheless, the ideas behind Azuma's inequality can be cleverly applied to obtain the following result [8].

Theorem 3.3 (Bollobás and Brightwell, 1992)**.** *For each $d \geq 2$ there exists $K_d > 0$ such that*

$$\mathbb{P}\left(|L_n - \mathbb{E}[L_n]| > \frac{\lambda K_d n^{1/2d} \log n}{\log\log n}\right) \leq 4\lambda^2 \exp(-\lambda^2) \tag{3.17}$$

for all λ with $2 < \lambda < n^{1/2d}/\log\log n$.

Let us sketch the proof of Theorem 3.3, which is based on Azuma's inequality. The idea is to sub-divide the points in the Poisson point process X_n in such a way that L_n does not depend too much on how the random process behaves within any of the pieces. A natural way to do this is to break up the points in X_n into "slices" that should all be roughly orthogonal to the longest chain. Thus, each slice will contribute only a few points to the longest chain.

More specifically, let m be the smallest integer larger than $n^{1/d}$ and define the slices

$$S_j = \left\{x \in [0,1]^d \,:\, j - 1 \leq \frac{m}{d}\sum_{i=1}^{d} x_i < j\right\}. \tag{3.18}$$

Let $Z_j = S_j \cap X_n$ be the Poisson points falling in each slice, and note the Z_j are mutually independent. The idea is to apply Azuma's inequality to the representation

$$L_n = f(Z_1, \ldots, Z_m).$$

However, the function f does not have bounded differences (i.e., (3.11) does not hold), since any individual slice could contain an arbitrarily large number of points and make a huge contribution to L_n. While this is possible, it is very unlikely, and the bulk of the argument amounts to controlling this very unlikely event.

To handle this, we define another version of L_n that does satisfy the bounded differences condition. For a fixed $k \in \mathbb{N}$, let L'_n be the length of a longest chain in X_n that contains at most k points from each slice Z_j. The bounded differences condition amounts to swapping out the points in a chosen slice Z_j and counting how much L'_n can change by. Thus, L'_n satisfies the bounded differences condition with $r = k$. Azuma's inequality gives

$$\mathbb{P}(|L'_n - \mathbb{E}[L'_n]| > \lambda) \leq 2\exp\left(\frac{-\lambda^2}{2mk^2}\right). \tag{3.19}$$

The idea is to choose k as small as possible to get the best estimate, while ensuring $L_n - L'_n$ is small. To illustrate the idea behind the proof, we set $k = n^{1/8d}$ and $\lambda = n^{3/4d}$ in (3.19) to get

$$\mathbb{P}(|L'_n - \mathbb{E}[L'_n]| > n^{3/4d}) \leq 2\exp\left(-Cn^{1/4d}\right). \tag{3.20}$$

To show that $L_n - L'_n$ is small, we divide the cube $[0,1]^d$ into (approximately) n identical subcubes of measure $1/n$ and let N_j denote the number of Poisson points in the j^{th} subcube. We claim there exists $c > 0$ such that if $N_j \leq ck$ for all j then $L'_n = L_n$. Briefly, any chain in a slice S_j can only pass through a constant $M = M(d)$ number of subcubes, so each slice can contribute at most cMk points to the longest chain. If we choose $c = 1/M$, then each slice contributes at most k points to the longest chain and hence $L_n = L'_n$.

Each random variable N_j is an independent Poisson random variable with mean 1. By tail bounds on Poisson random variables

$$\mathbb{P}(N_j \geq k) \leq \exp(-k) = \exp\left(-n^{1/8d}\right), \tag{3.21}$$

for n large enough so that $k \geq e^2$. By the union bound we have

$$\mathbb{P}(L_n - L'_n \geq 1) \leq n \exp\left(-n^{1/8d}\right).$$

By arguments similar to those at the end of Section 3.1

$$\mathbb{E}[L_n - L'_n] \leq 1 + Cn^2 \exp\left(-n^{1/8d}\right) = 1 + C \exp\left(-n^{1/8d} + 2\log n\right).$$

Thus, for n large enough we have $\mathbb{E}[L_n - L'_n] \leq 2$. Combining this with (3.20) yields

$$\mathbb{P}(|L_n - \mathbb{E}[L_n]| > 2n^{3/4d}) \leq n \exp\left(-n^{1/8d}\right) \tag{3.22}$$

for sufficiently large n. Almost sure convergence follows now from the Borel-Cantelli Lemma, as described at the start of the section.

We note that our concentration of measure result (3.22) is not as tight as Theorem 3.3. The reason is that our choice of $k = n^{1/8d}$, while simplifying the proof, is far too large. The proof of Theorem 3.3 in [8] uses $k = C \log n / \log\log n$ and slightly more subtle arguments for bounding $L_n - L'_n$. We leave the details to the reader.

4 Viscosity solutions

Hamilton-Jacobi equations like (1.5) do not in general admit classical C^1 solutions that satisfy the PDE everywhere, due to the possibility of crossing characteristics. On the other hand, there are in general infinitely many Lipschitz continuous functions that satisfy the PDE almost everywhere. The notion of viscosity solution, developed in [20,21], selects the *physically* correct Lipschitz almost everywhere solution for a *very wide* range of applications. Viscosity solutions enjoy very strong stability and uniqueness properties that make them extremely useful tools for a wide range of problems, especially when passing to a limit in a sequence of PDEs (or approximations thereof).

4.1 Definitions

We now give the definition of viscosity solution.

Definition 1 (Viscosity solutions)**.** Let $U \subset \mathbb{R}^n$ be open and $u \in C(\overline{U})$.

(i) We say u is a *viscosity subsolution* of $H(x, u, \nabla u) = 0$ in U (written $H(x, u, \nabla u) \leq 0$ in U) if for every $x \in U$ and $\varphi \in C^\infty(U)$ such that $u - \varphi$ has a local maximum at x

$$H(x, u(x), \nabla\varphi(x)) \leq 0.$$

(ii) We say u is a *viscosity supersolution* of $H(x, u, \nabla u) = 0$ in U (written $H(x, u, \nabla u) \geq 0$ in U) if for every $x \in U$ and $\varphi \in C^\infty(U)$ such that $u - \varphi$ has a local minimum at x

$$H(x, u(x), \nabla\varphi(x)) \geq 0.$$

(iii) We say u is a *viscosity solution* of $H(x, u, \nabla u) = 0$ in U if u is both a viscosity sub- and supersolution.

Remark 4.1. The definition of viscosity solution can ostensibly be applied to any PDE of any order, not just first order Hamilton-Jacobi type equations. However, uniqueness of viscosity solutions is based on the maximum principle, which applies only to PDEs of degree at most two (since the necessary conditions for a maximum only involve first and second derivatives). We focus mostly on first order Hamilton-Jacobi equations here; for the theory of viscosity solutions for second order degenerate elliptic equations we refer the reader to the user's guide [19].

The original motivation for viscosity solutions is the method of vanishing viscosity, whereby a *viscosity term* $-\varepsilon\Delta u$ is added to the Hamilton-Jacobi equation to regularize the equation, resulting in a unique smooth solution u_ε. As $\varepsilon \to 0^+$ the solutions u_ε can be shown to converge uniformly to a continuous function u that satisfies the definition of viscosity solution above. While this is appealing intuitively, it is not generally the most useful way to think about viscosity solutions.

Being the limit a sequence of semilinear elliptic equations, it turns out that the viscosity solution inherits useful properties from elliptic equations. The most important property is the *maximum principle*, and a useful way to think about viscosity solutions is that they are a notion of weak solution that obeys the maximum principle. Indeed, we can immediately see from the definitions that viscosity solutions enjoy the comparison property against smooth strict super/subsolutions. That is, suppose that $v \in C^\infty(\overline{U})$ satisfies $H(x, \nabla v(x)) > 0$ in U, u is a viscosity subsolution of $H(x, \nabla u) = 0$ in U, and U is bounded. Then $u - v$ attains its maximum value at some $x \in \overline{U}$. By the definition of viscosity subsolution we must have $x \in \partial U$ (why?) and thus

$$\max_{\overline{U}}(u - v) = \max_{\partial U}(u - v). \tag{4.1}$$

A similar result holds for smooth strict subsolutions. We note that (4.1) is referred to as the maximum principle (or sometimes the comparison principle). We prove in Theorem 4.1 a comparison principle for viscosity solutions when neither u nor v is smooth.

One may naturally be inclined to ask why viscosity solutions are the right notion of solution for the continuum limits of the data peeling problems we are considering. The reason is actually quite simple; all of our data peeling algorithms satisfy a type of *monotonicity*. That is, when we add new points, the peeling depth function U_n can only increase. So if $f \leq g$ and U_n and V_n are the Pareto depth functions for Poisson point processes X_{nf} and X_{ng} with intensity nf and ng, respectively, then $U_n \leq V_n$ (provided we construct the point processes so that $X_{nf} \subset X_{ng}$). This is simply a discrete version of the maximum principle (or comparison principle) given in (4.1) and Theorem 4.1 below.

4.2 The maximum principle

We now give the proof of the comparison principle for viscosity solutions of Hamilton-Jacobi equations. To simplify the proof as much as possible, we present the proof for strict sub/supersolution pairs on a bounded domain $U \subset \mathbb{R}^d$ for the simple Hamilton-Jacobi equation

$$H(\nabla u) = f \quad \text{in } U,$$

where $U \subset \mathbb{R}^d$ is *bounded*. This version of the proof contains the essence of the result, without all the bells and whistles attached to a more complicated setup.

Theorem 4.1 (Strict comparison)**.** *Let $f \in C(\overline{U})$ and $U \subset \mathbb{R}^d$ be bounded. If $u, v \in C(\overline{U})$ satisfy*

$$H(\nabla u) \leq f \quad \text{and} \quad H(\nabla v) \geq f + \varepsilon \quad \text{in } U \tag{4.2}$$

in the viscosity sense, where $\varepsilon > 0$ is a constant, then

$$\max_{\overline{U}}(u - v) = \max_{\partial U}(u - v). \tag{4.3}$$

If $u, v \in C^1(\overline{U})$ then the proof of Theorem 4.1 follows a standard maximum principle argument. We simply examine the point $x \in \overline{U}$ at which $u - v$ attains its maximum value. The point x cannot be in the interior of the domain, since we would then have $\nabla u(x) = \nabla v(x)$, which contradicts (4.2) since $\varepsilon > 0$. Therefore $x \in \partial U$ and (4.3) follows. It is remarkable that this argument can be rescued when u and v are not differentiable, and are merely viscosity sub- and supersolution of (4.2). The proof is based on the clever idea of doubling the number of variables and looking instead at the maximum of $u(x) - v(y) - k|x - y|^2$, where k is large. The additional $-k|x-y|^2$ term serves as a smooth test function from the definition of viscosity solution that touches u from above and v from below.

Proof. Without loss of generality we may assume that $\max_{\partial U}(u - v) = 0$. Assume to the contrary now that

$$\delta := \max_{\overline{U}}(u - v) > 0 = \max_{\partial U}(u - v). \tag{4.4}$$

We define the auxiliary function

$$\Phi_k(x, y) = u(x) - v(y) - k|x - y|^2, \tag{4.5}$$

for $x, y \in \overline{U}$ and $k > 0$. Let $(x_k, y_k) \in \overline{U} \times \overline{U}$ such that

$$M_k := \max_{\overline{U}\times\overline{U}} \Phi_k = \Phi(x_k, y_k). \tag{4.6}$$

Since $M_k \geq 0$, $u(x_k) - v(y_k) - k|x_k - y_k|^2 \geq 0$, and so

$$|x_k - y_k| \leq \frac{C}{\sqrt{k}}, \tag{4.7}$$

since u and v are bounded. Therefore, there exists $x_0 \in \overline{U}$ such that, passing to a subsequence if necessary, we have $x_k \to x_0$ and $y_k \to x_0$ as $k \to \infty$. Noticing that $M_k \geq \delta$ and so

$$\delta \leq \liminf_{k\to\infty} M_k = \limsup_{k\to\infty} M_k \leq u(x_0) - v(x_0) \leq \delta,$$

we see that $u(x_0) - v(x_0) = \delta > 0$ and so $x_0 \in U$.

Since $x_0 \in U$, we know that $x_k, y_k \in U$ for k sufficiently large. Note that $u - \varphi$ attains its maximum over $\overline{U}$ at $x_k \in U$, where $\varphi(x) = k|x - y_k|^2$. It follows from the definition of viscosity subsolution that

$$H(p_k) = H(\nabla\varphi(x_k)) \leq f(x_k)$$

where $p_k = 2k(x_k - y_k)$. Similarly, $v - \psi$ has a minimum over $\overline{U}$ at y_k, where $\psi(y) = -k|x_k - y|^2$. Therefore

$$H(p_k) = H(\nabla\psi(y_k)) \geq f(y_k) + \varepsilon.$$

Combining the inequalities above yields

$$f(x_k) - f(y_k) \geq \varepsilon.$$

By the uniform continuity of f and (4.7), the left hand side above tends to zero as $k \to \infty$, which is a contradiction. □

Relaxing the strictness condition (i.e., taking $\varepsilon = 0$) in Theorem 4.1 requires some further structural conditions on H. We show how to relax this condition for the PDE (1.5) below.

Theorem 4.2. *If $f \in C([0,\infty)^d)$ then there is at most one* increasing *viscosity solution of* (1.5).

In the statement of Theorem 4.2, *increasing* means that $u(x) \leq u(y)$ whenever $x_i \leq y_i$ for all i. Thus, u is increasing with respect to the partial order for non-dominated sorting, which is a natural condition to place on solutions of (1.5). We also remark that the key points in the proof of Theorem 4.2 are (1) dealing with the unbounded domain $\mathbb{R}^d_+$ and (2) relaxing the strictness condition from Theorem 4.1. We sketch the proof below, leaving some details as exercises.

Proof. Let $u, v \in C([0,\infty)^d)$ be two increasing viscosity solutions of (1.5). For $\varepsilon > 0$ define

$$v^\varepsilon(x) = v(x) + \varepsilon^{1/d}(x_1 + \cdots + x_d).$$

Then it follows that v^ε satisfies

$$\begin{cases} v^\varepsilon_{x_1} \cdots v^\varepsilon_{x_d} \geq f + \varepsilon, & \text{in } \mathbb{R}^d_+ \\ v^\varepsilon \geq 0, & \text{on } \partial\mathbb{R}^d_+ \end{cases} \tag{4.8}$$

in the viscosity sense. Let $\Phi : \mathbb{R} \to \mathbb{R}$ be a smooth function satisfying $\Phi(t) = t$ for $t \leq 1$, and $0 \leq \Phi'(t) \leq 1$ and $\Phi(t) \leq 2$ for all t. Define

$$u^\varepsilon(x) = \varepsilon^{-1}\Phi(\varepsilon u(x)).$$

Then it follows that

$$\begin{cases} u^\varepsilon_{x_1} \cdots u^\varepsilon_{x_d} \le f, & \text{in } \mathbb{R}^d_+ \\ u^\varepsilon = 0, & \text{on } \partial\mathbb{R}^d_+ \end{cases} \tag{4.9}$$

in the viscosity sense. By Theorem 4.1 applied to $U = (0,R)^d$ we have

$$\max_{[0,R]^d}(u^\varepsilon - v^\varepsilon) = \max_{\partial(0,R)^d}(u^\varepsilon - v^\varepsilon)$$

for any $R > 0$. Since u^ε is bounded (by $2\varepsilon^{-1}$) and $v^\varepsilon(x) \ge \varepsilon^{1/d}(x_1 + \cdots + x_d)$ (since $v \ge 0$) we have that $\max_{\partial(0,R)^d}(u^\varepsilon - v^\varepsilon) \le 0$ for all $R > 0$ sufficiently large. Therefore $u^\varepsilon \le v^\varepsilon$ on $\mathbb{R}^d_+$. Sending $\varepsilon \to 0^+$ we obtain $u \le v$. Swapping the roles of u and v completes the proof. □

Exercise 1. Verify that v^ε satisfies (4.8) in the viscosity sense, and that u^ε satisfies (4.9) in the viscosity sense.

It is possible to prove uniqueness of viscosity solutions of (1.5) when f has certain kinds of discontinuities. For example, Theorem 4.2 holds provided there exists some open and bounded set $U \subset \mathbb{R}^d_+$ with Lipschitz boundary such that $f = 0$ for $x \notin U$ and f is nonnegative and uniformly continuous on U. The proof follows closely the proof of Theorem 4.1 with some adjustments to the auxiliary function Φ_k to account for the discontinuity of f along ∂U. The notion of viscosity solution must be adjusted slightly for discontinuous f; we refer the reader to [12] for more details. Recall that f is the probability density for the random points we are sorting with non-dominated sorting. The setup described above models the case where the points are sampled from the domain $U \subset \mathbb{R}^d_+$. For more information on viscosity solutions we refer the reader to [6, 19].

5 Convergence to continuum limit

We give in this section the proof of Theorem 1.1, which is split into three main parts.

5.1 Compactness

The first step is to establish a form of compactness for the sequence U_n. Without loss of generality we assume that $0 \le f \le 1$.

Lemma 5.1 (Compactness)**.** *Let $x, y \in [0,M]^d$. Then with probability one*

$$\limsup_{n\to\infty} n^{-1/d}(U_n(x) - U_n(y)) \le CM^{(d-1)/d}|x-y|^{1/d}. \tag{5.1}$$

Proof. It is enough to show that

$$\limsup_{n\to\infty} n^{-1/d}(U_n(x + he_1) - U_n(x)) \le CM^{(d-1)/d}h^{1/d}$$

holds with probability one for any $h > 0$. We may also assume $x_j > 0$ for all $j \ge 2$, otherwise $U_n(x + he_1) = U_n(x) = 0$ almost surely for all n.

Let R denote the rectangle spanning x_1e_1 and $x + he_1$; see Figure 5.1 for an illustration. Define

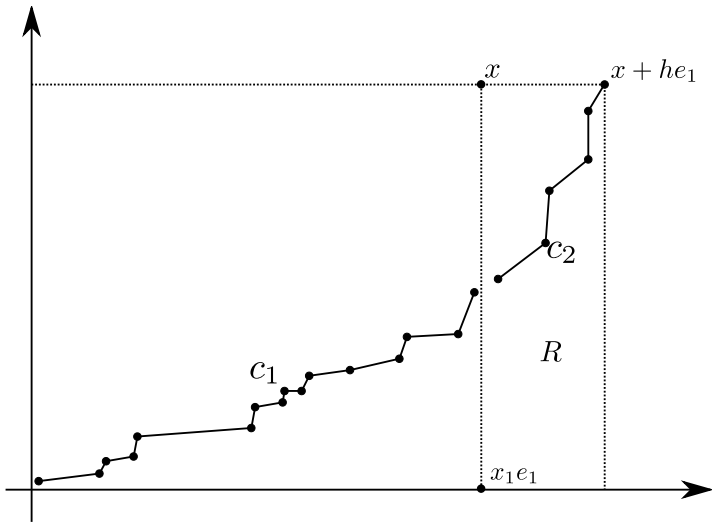


Figure 5.1: Illustration of quantities from the proof of Lemma 5.1

$$g(x) = \begin{cases} 1, & \text{if } x \in R \\ f(x), & \text{if } x \notin R, \end{cases}$$

and let X_{ng} be a Poisson point process with intensity ng constructed so that $X_{nf} \subset X_{ng}$. Any maximal chain in the rectangle $[0, x + he_1]$ can be subdivided into two pieces; one piece c_1 contained entirely in $[0, x]$ of length at most $U_n(x)$, and one piece c_2 in R of length at most $\ell(X_{nf} \cap R)$ (see, again, Figure 5.1). Therefore

$$U_n(x + he_1) \le U_n(x) + \ell(X_{nf} \cap R) \le U_n(x) + \ell(X_{ng} \cap R).$$

By Corollary 1

$$\limsup_{n\to\infty} n^{-1/d}(U_n(x + he_1) - U_n(x)) \le \lim_{n\to\infty} \frac{\ell(X_{ng} \cap R)}{n^{1/d}} = c_d|R|^{1/d} \le c_d M^{(d-1)/d} h^{1/d}$$

with probability one, which completes the proof. □

Lemma 5.1 is sufficient to apply an argument similar to the Arzelà-Ascoli Theorem to obtain uniform convergence of subsequences of $n^{-1/d}U_n$.

Lemma 5.2. *Suppose $u_n : [0,\infty)^d \to \mathbb{R}$ are increasing functions satisfying $u_n(0) = 0$ and*

$$\limsup_{n\to\infty} (u_n(x) - u_n(y)) \le C|x - y|^\alpha \quad \text{for all } x, y \in [0,\infty)^d \cap \mathbb{Q}^d, \tag{5.2}$$

where $\alpha > 0$ and C depends only on $\max\{x_1, y_1, \ldots, x_d, y_d\}$. Then there exists a subsequence $n_k \to \infty$ and an increasing function $u \in C([0,\infty)^d)$ such that

$$u_{n_k} \longrightarrow u \text{ locally uniformly on } [0,\infty)^d \text{ as } k \to \infty.$$

Proof. By (5.2) and the assumption that $u_n(0) = 0$, the sequence $\{u_n(x)\}_{n=1}^\infty$ is bounded for each $x \in [0,\infty)^d \cap \mathbb{Q}^d$. Therefore we can use the Bolzano-Weierstrass Theorem and a diagonal

argument to obtain a subsequence u_{n_k} such that the sequence $\{u_{n_k}(x)\}_{k=1}^{\infty}$ is convergent for all $x \in [0,\infty)^d \cap \mathbb{Q}^d$. Let us write $u(x) := \lim_{k\to\infty} u_{n_k}(x)$. For any $x, y \in [0,\infty)^d \cap \mathbb{Q}^d$

$$|u(x) - u(y)| = \lim_{k\to\infty} |u_{n_k}(x) - u_{n_k}(y)| \le C|x-y|^{\alpha}.$$

Therefore, u can be extended uniquely to a continuous increasing function $u : [0,\infty)^d \to \mathbb{R}$.

We now show that $u_{n_k} \to u$ locally uniformly on $[0,\infty)^d$. Fix $M > 0$ and let $\varepsilon > 0$. Let $h > 0$ be small enough so that

$$u(x + \mathbf{1}h) - u(x) < \varepsilon \quad \text{for all } x \in h\mathbb{Z}^d \cap [0,M]^d,$$

where $\mathbf{1} = (1,\dots,1) \in \mathbb{R}^d$. Let $y \in [0,M]^d$ and choose $x \in h\mathbb{Z}^d \cap [0,M]^d$ such that $x_i \le y_i < x_i + h$ for all i. Then since u_{n_k} and u are increasing we have

$$u_{n_k}(y) - u(y) \le u_{n_k}(x + \mathbf{1}h) - u(x + \mathbf{1}h) + \varepsilon,$$

and

$$u(y) - u_{n_k}(y) \le u(x) - u_{n_k}(x) + \varepsilon.$$

Therefore

$$\|u_{n_k} - u\|_{L^\infty([0,M]^d)} \le \max_{x \in h\mathbb{Z}^d \cap [0,M+h]^d} |u_{n_k}(x) - u(x)| + \varepsilon$$

and so

$$\limsup_{k\to\infty} \|u_{n_k} - u\|_{L^\infty([0,M]^d)} \le \varepsilon.$$

It follows that $u_{n_k} \to u$ uniformly on $[0,M]^d$, which completes the proof. □

5.2 Longest chain in a simplex

As we did in the nonrigorous argument in Section 2, the next step is to count the length of a longest chain in simplices of the form

$$S_v = \left\{ x \in (-\infty, 0]^d \, : 1 + x \cdot v \ge 0 \right\}, \tag{5.3}$$

where $v \in (0,\infty)^d$. The simplex S_v is a simplex with orthogonal corner at the origin and side lengths $\ell_1 = v_1^{-1}, \ell_2 = v_2^{-1}, \dots, \ell_d = v_d^{-1}$ (i.e., the set A from Figure 2.1).

Lemma 5.3. *For any $v \in (0,\infty)^d$ and $x \in \mathbb{R}^d$ we have*

$$\limsup_{n\to\infty} \frac{\ell((x + S_v) \cap X_{nf})}{n^{1/d}} \le \frac{c_d}{d} \left(\frac{\sup_{x+S_v} f}{v_1 \cdots v_d} \right)^{1/d}, \tag{5.4}$$

and

$$\liminf_{n\to\infty} \frac{\ell((x + S_v) \cap X_{nf})}{n^{1/d}} \ge \frac{c_d}{d} \left(\frac{\inf_{x+S_v} f}{v_1 \cdots v_d} \right)^{1/d}, \tag{5.5}$$

with probability one.

Proof. We sketch the proof of (5.4); the proof of (5.5) is similar. Without loss of generality we can take $x = 0$.

We first prove the result for $f = 1$ and $v = \mathbf{1}$. In this case we claim that

$$\lim_{n\to\infty} \frac{\ell(X_n \cap S_{\mathbf{1}})}{n^{1/d}} = \frac{c_d}{d} \quad \text{almost surely.} \tag{5.6}$$

To see this, note that the rectangle of largest area contained in $S_{\mathbf{1}}$ is $R = [-1/d, 0]^d$, which has measure $|R| = 1/d^d$. Let $\varepsilon > 0$ and let $R_1, \ldots, R_m$ be m rectangles such that $|R_j| \le 1/d^d + \varepsilon$ for all j, $S_{\mathbf{1}} \subset \cup_{j=1}^m R_j$, and $0 \in R_j$ for all j. Then any chain in $X_n \cap S_{\mathbf{1}}$ is contained in at least one of the rectangles R_j and hence

$$\ell(X_n \cap S_{\mathbf{1}}) \le \max_{1\le j\le m} \ell(X_n \cap R_j).$$

By Corollary 1

$$\limsup_{n\to\infty} \frac{\ell(X_n \cap S_{\mathbf{1}})}{n^{1/d}} \le c_d \left(\frac{1}{d^d} + \varepsilon\right)^{1/d} \quad \text{almost surely.}$$

Sending $\varepsilon \to 0^+$ we get

$$\limsup_{n\to\infty} \frac{\ell(X_n \cap S_{\mathbf{1}})}{n^{1/d}} \le \frac{c_d}{d} \quad \text{almost surely.}$$

The lim inf inequlity is immediate, since $S_{\mathbf{1}} \supset R$. This establishes (5.6).

We now prove (5.4). We define

$$g(x) = \begin{cases} \sup_{S_v} f, & \text{if } x \in S_v, \\ f, & \text{otherwise,} \end{cases}$$

and let X_{ng} be Poisson point process with intensity ng constructed so that $X_{nf} \subset X_{ng}$. Let $\Phi(x) = (v_1 x_1, \ldots, v_d x_d)$. Then $X := \Phi(X_{ng})$ is a Poisson point process with intensity

$$\lambda = |\det \Phi|^{-1} n \sup_{S_v} f = \frac{n \sup_{S_v} f}{v_1 \cdots v_d}$$

on the simplex $S_{\mathbf{1}} = \Phi(S_v)$. Since Φ sends chains to chains we have

$$\ell(S_v \cap X_n) \le \ell(S_v \cap X_{ng}) = \ell(\Phi(S_v) \cap \Phi(X_{ng})) = \ell(S_{\mathbf{1}} \cap X). \tag{5.7}$$

By (5.6) we have

$$\limsup_{n\to\infty} \frac{\ell(S_v \cap X_n)}{n^{1/d}} \le \lim_{n\to\infty} \left(\frac{\lambda^{1/d}}{n^{1/d}}\right) \frac{\ell(S_{\mathbf{1}} \cap X)}{\lambda^{1/d}} = \frac{c_d}{d} \left(\frac{\sup_{S_v} f}{v_1 \cdots v_d}\right)^{1/d} \quad \text{almost surely.} \qquad \square$$

5.3 Proof of main result

We now give the proof of Theorem 1.1.

Proof of Theorem 1.1. Let $\Omega_{x,y}$ denote the event that (5.1) holds for $x, y \in [0,\infty)^d$, and let Ω denote the intersection of $\Omega_{x,y}$ over the countable collection of $x, y \in [0,\infty)^d \cap \mathbb{Q}^d$. By Lemma 5.1 we know that $\mathbb{P}(\Omega_{x,y}) = 1$, and so $\mathbb{P}(\Omega) = 1$. By the same argument, the event Ω' that (5.4) and (5.5) hold for all $x \in [0,\infty)^d \cap \mathbb{Q}^d$ and all $v \in (0,\infty)^d \cap \mathbb{Q}^d$ has probability one. For the remainder of the proof we fix a realization in the probability one event $\Omega \cap \Omega'$. The remainder of the proof is deterministic.

The sequence $u_n := n^{-1/d} U_n$ satisfies the hypotheses of Lemma 5.2. Therefore, passing to a subsequence, which we again denote by u_n for convenience, there exists $u \in C([0,\infty)^d)$ such that $u_n \to u$ locally uniformly on $[0,\infty)^d$.[3] We claim that u is the unique viscosity solution of

$$\begin{cases} u_{x_1} \cdots u_{x_d} = \dfrac{c_d^d}{d^d} f, & \text{in } \mathbb{R}^d_+ \\ u = 0, & \text{on } \partial\mathbb{R}^d_+. \end{cases} \tag{5.8}$$

This will complete the proof, since all subsequences contain subsequences converging to the solution of (5.8), so the entire sequence must converge to the same function. Note in Theorem 1.1 we absorb the constant c_d^d/d^d into the convergence statement (1.4) instead of the PDE.

To show that u is the viscosity solution of (5.8) we have to verify the sub- and supersolution conditions.

1. We first verify that u is a viscosity subsolution of (5.8). Let $x_0 \in \mathbb{R}^d_+$ and let $\varphi \in C^\infty(\mathbb{R}^d)$ such that $u - \varphi$ has a local maximum at x_0. Since u is increasing we must have $\varphi_{x_i}(x_0) \geq 0$ for all i (why?), and if $\varphi_{x_i}(x_0) = 0$ for any i then the subsolution property holds trivially, since $f(x_0) \geq 0$. Therefore we may assume that $\varphi_{x_i}(x_0) > 0$ for all i. By modifying φ away from x_0, we may also assume, without loss of generality, that $u(x_0) = \varphi(x_0)$, $u(x) \leq \varphi(x)$ for all $x \in [0, x_0]$, and φ is increasing for all $x \in [0, x_0]$.

Let $\varepsilon > 0$, $v \in (0,\infty)^d$, and set[4]

$$A_n = \{x \in [0, x_0] \,:\, u_n(x) \geq \varphi(x_0 - \varepsilon v)\}. \tag{5.9}$$

and

$$A = \{x \in [0, x_0] \,:\, \varphi(x) \geq \varphi(x_0 - \varepsilon v) - \varepsilon^2\}. \tag{5.10}$$

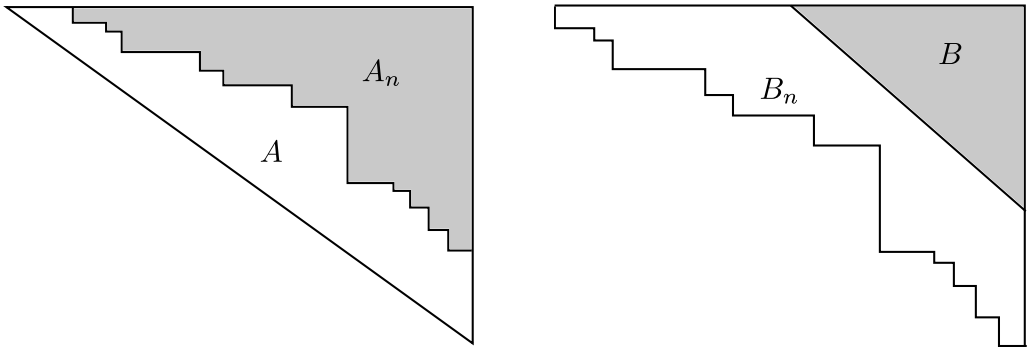
We claim that

$$U_n(x_0) \leq n^{1/d} \varphi(x_0 - \varepsilon v) + \ell(A_n \cap X_{nf}). \tag{5.11}$$

Briefly, the idea is to peel away all points on fronts with indices less than or equal to $n^{1/d}\varphi(x_0 - \varepsilon v)$, and then compute $U_n(x_0)$ by peeling (or counting the longest chain in) A_n. Furthermore, since $u_n \to u$ uniformly and $u \leq \varphi$ for $x \in [0, x_0]$, we have $A_n \subset A$ for n sufficiently large. See Figure 2(a) for an illustration. Therefore

[3] The reader should note that the particular subsequence we pass to, and the limit u both depend on the realization selected earlier in the proof.

[4] Notice we are "looking backwards", contrary to the nonrigorous argument in Section 2.



(a) $A_n \subset A$ (b) $B \subset B_n$

Figure 5.2: Illustration of the inclusions $A_n \subset A$ and $B \subset B_n$ from the proof of Theorem 1.1.

$$\begin{aligned}\varepsilon\nabla\varphi(x_0)\cdot v - C\varepsilon^2 &\leq \varphi(x_0) - \varphi(x_0 - \varepsilon v)\\ &= u(x_0) - \varphi(x_0 - \varepsilon v)\\ &= \lim_{n\to\infty} u_n(x_0) - \varphi(x_0 - \varepsilon v)\\ &\leq \lim_{n\to\infty} n^{-1/d}\ell(A_n \cap X_{nf})\\ &\leq \limsup_{n\to\infty} n^{-1/d}\ell(A \cap X_{nf}).\end{aligned}$$

The reader should contrast this with the first part of (2.1) from Section 2.

Since $A \subset B(x_0, C\varepsilon)$ for some $C > 0$ (why?), we can apply Taylor expansions to φ in (5.10) to obtain

$$\varepsilon\nabla\varphi(x_0)\cdot v + C\varepsilon^2 + \nabla\varphi(x_0)\cdot(x - x_0) \geq 0$$

whenever $x \in A$. Letting $y \in [0,\infty)^d \cap \mathbb{Q}^d$ with $y_i \geq x_{0,i}$ for all i and $|y - x_0| \leq \varepsilon^2$ we find that

$$1 + (x - y)\cdot p \geq 0$$

whenever $x \in A$, where

$$p = \frac{\nabla\varphi(x_0)}{\varepsilon\nabla\varphi(x_0)\cdot v + C\varepsilon^2}.$$

Let $q \in \mathbb{Q}^d$ with $0 < q_i < p_i$ for all i and $|q - p| \leq \varepsilon^2$. Then for any $x \in A$

$$1 + (x - y)\cdot q \geq 0,$$

and so $A \subset y + S_q$, where y and q have rational coordinates. We therefore have

$$\varepsilon\nabla\varphi(x_0)\cdot v - C\varepsilon^2 \leq \limsup_{n\to\infty} n^{-1/d}\ell((y + S_q)\cap X_{nf}) \leq \frac{c_d}{d}\left(\frac{\sup_{y+S_q} f}{q_1\cdots q_d}\right)^{1/d}.$$

Sending $\varepsilon \to 0^+$ we obtain

$$1 \leq \frac{c_d}{d}\left(\frac{f(x_0)}{\varphi_{x_1}(x_0)\cdots\varphi_{x_d}(x_0)}\right)^{1/d},$$

which is exactly the subsolution property.

2. We now show that u is a viscosity supersolution of (5.8). Let $x_0 \in \mathbb{R}^d_+$ and let $\varphi \in C^\infty(\mathbb{R}^d)$ such that $u - \varphi$ has a local minimum at x_0. Since u is increasing we must have $\varphi_{x_i}(x_0) \geq 0$ for all i (why?). Let $\lambda > 0$ and define

$$\psi(x) = \varphi(x) + \lambda(x_1 + \cdots + x_d).$$

Then for $x \in B(x_0, r) \cap [0, x_0]$ we have

$$\begin{aligned} u(x) - \psi(x) &= u(x) - \varphi(x) - \lambda(x_1 + \cdots + x_d) \\ &\geq u(x_0) - \varphi(x_0) - \lambda(x_{0,1} + \cdots + x_{0,d}) \\ &= u(x_0) - \psi(x_0). \end{aligned}$$

Therefore $u - \psi$ has a local maximum at x_0 relative to $[0, x_0]$ and $\psi_{x_i}(x_0) \geq \lambda > 0$. We can again translate ψ and modify ψ away from x_0 so that $u(x_0) = \psi(x_0)$, $u \geq \psi$ on $[0, x_0]$, and ψ is increasing on $[0, x_0]$.

As before, let $\varepsilon > 0$, $v \in (0, \infty)^d$ and set

$$B_n = \{x \in [0, x_0] \,:\, u_n(x) \geq \psi(x_0 - \varepsilon v)\}. \tag{5.12}$$

and

$$B = \{x \in [0, x_0] \,:\, \psi(x) \geq \psi(x_0 - \varepsilon v) + \varepsilon^2\}. \tag{5.13}$$

Similar to (5.11) we have

$$U_n(x_0) \geq n^{1/d}\psi(x_0 - \varepsilon v) + \ell(B_n \cap X_{nf}) - 2.$$

Since $u_n \to u$ uniformly and $u \geq \psi$ on $[0, x_0]$, we have $B \subset B_n$ for n sufficiently large. See Figure 2(b) for an illustration. Therefore

$$\begin{aligned} \varepsilon \nabla\varphi(x_0) \cdot v + C\varepsilon^2 &\geq \varphi(x_0) - \varphi(x_0 - \varepsilon v) \\ &= u(x_0) - \varphi(x_0 - \varepsilon v) \\ &= \lim_{n\to\infty} u_n(x_0) - \varphi(x_0 - \varepsilon v) \\ &\geq \lim_{n\to\infty} n^{-1/d}\ell(B_n \cap X_{nf}) \\ &\geq \liminf_{n\to\infty} n^{-1/d}\ell(B \cap X_{nf}). \end{aligned}$$

Again, for $\varepsilon > 0$ small enough we have $B \subset B(x_0, C\varepsilon)$, and so we can apply Taylor expansions to ψ to find that if $x \in B(x_0, C\varepsilon)$ and

$$\varepsilon \nabla\psi(x_0) \cdot v - C\varepsilon^2 + \nabla\psi(x_0) \cdot (x - x_0) \geq 0$$

then $x \in B$. Let $y \in [0, \infty)^d \cap \mathbb{Q}^d$ such that $y_i \leq x_{0,i}$ for all i and $|y - x_0| \leq \varepsilon^2$. Then if $x \in B(x_0, C\varepsilon)$ and $1 + (x - y) \cdot p \geq 0$ then $x \in B$, where

$$p = \frac{\nabla\psi(x_0)}{\varepsilon \nabla\psi(x_0) \cdot v - C\varepsilon^2}.$$

It follows that $y + S_q \subset B$ for any $q \in \mathbb{Q}^d$ with $q_i > p_i$ and, say, $|q - p| \le \varepsilon^2$. Therefore

$$\varepsilon \nabla\varphi(x_0) \cdot v + C\varepsilon^2 \ge \liminf_{n\to\infty} n^{-1/d} \ell((y + S_q) \cap X_{nf}) \ge \frac{c_d}{d} \left(\frac{\inf_{y+S_q} f}{q_1 \cdots q_d} \right)^{1/d}.$$

Sending $\varepsilon \to 0^+$ and recalling $\psi_{x_i} = \varphi_{x_i} + \lambda$ we find that

$$(\varphi_{x_1}(x_0) + \lambda) \cdots (\varphi_{x_d}(x_0) + \lambda) \ge \frac{c_d^d}{d^d} f(x_0).$$

Sending $\lambda \to 0^+$ we obtain the supersolution property

$$\varphi_{x_1}(x_0) \cdots \varphi_{x_d}(x_0) \ge \frac{c_d^d}{d^d} f(x_0).$$

This completes the proof. □

Exercise 2. Fill in some of the missing steps marked by "(why?)" in the proof of Theorem 1.1.

A Poisson point processes

A Poisson point process X_f with intensity function f is a random at most countable collection of points in $\mathbb{R}^d$ with the following properties:

(i) For every Borel set $A \subset \mathbb{R}^d$, the number of points in $X_f \cap A$, denoted $N(A)$, is a Poisson random variable with mean $\lambda = \int_A f\,dx$.[5]

(ii) For disjoint Borel sets $A, B \subset \mathbb{R}^d$, the random variables $N(A)$ and $N(B)$ are independent.

Poisson point processes are not all that far removed from sequences of independent and identically distributed (*i.i.d.*) random variables. Indeed, if $\int_{\mathbb{R}^d} f\,dx = 1$ then one way to construct the Poisson point process X_{nf} for $n \in \mathbb{N}$ is

$$X_{nf} = \{Y_1, \dots, Y_N\},$$

where $Y_1, Y_2, Y_3, \dots$ is a sequence of *i.i.d.* random variables with probability density function f, and N is a Poisson random variable with mean n.

Poisson point processes have many properties that make them more mathematically convenient compared to *i.i.d.* sequences. First, note that the independence of $N(A)$ and $N(B)$ does not hold if X_f is replaced by a collection of n *i.i.d.* random variables $X_1, \dots, X_n$ (they are negatively correlated). Second, the restriction $Y = X_f \cap A$ to any Borel set A is again a Poisson point process on A with intensity f. Third, the union $Y = X_f \cup X_g$ is a Poisson point process with intensity $f + g$, i.e., $Y = X_{f+g}$. Fourth, given a Poisson point process X_f and another intensity function g with $f \le g$ (resp. $f \ge g$) it is possible to construct a Poisson point process X_g so that $X_f \subset X_g$ (resp. $X_f \supset X_g$). In the former case, we set $X_g = X_f \cup X_{g-f}$, while the latter case involves randomly deleting points from X_f (called "thinning") to recover $X_g \subset X_f$. Finally, if $A \in \mathbb{R}^{d\times d}$ is a nonsingular matrix, and $b \in \mathbb{R}^d$, then $Y := AX_f + b$ is again a Poisson point process with intensity

$$g(y) = |\det A|^{-1} f(A^{-1}(y - b)).$$

For more details on Poisson point processes we refer the reader to Kingman's book [44].

[5] N is Poisson with mean λ if $\mathbb{P}(N = k) = e^{-\lambda}\lambda^k/k!$.

References

[1] B. Abbasi, J. Calder, and A. M. Oberman. Anomaly detection and classification for streaming data using partial differential equations. *arXiv preprint:1608.04348*, 2016.

[2] J. Abello and F. Queyroi. Fixed points of graph peeling. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 256–263. ACM, 2013.

[3] 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.

[4] B. Andrews et al. Contraction of convex hypersurfaces by their affine normal. *Journal of Differential Geometry*, 43(2):207–230, 1996.

[5] S. Angenent, G. Sapiro, and A. Tannenbaum. On the affine heat equation for non-convex curves. *Journal of the American Mathematical Society*, 11(3):601–634, 1998.

[6] M. Bardi and I. Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Springer, 1997.

[7] V. Barnett. The ordering of multivariate data. *Journal of the Royal Statistical Society. Series A (General)*, pages 318–355, 1976.

[8] 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.

[9] 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.

[10] J. Calder. Numerical schemes and rates of convergence for the Hamilton-Jacobi equation continuum limit of nondominated sorting. *arXiv preprint:1508.01557*, 2015.

[11] J. Calder. A direct verification argument for the Hamilton–Jacobi equation continuum limit of nondominated sorting. *Nonlinear Analysis: Theory, Methods & Applications*, 141:88–108, 2016.

[12] 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.

[13] J. Calder, S. Esedo$\bar{g}$lu, and A. O. Hero. A PDE-based approach to nondominated sorting. *SIAM Journal on Numerical Analysis*, 53(1):82–104, 2015.

[14] J. Calder and C. K. Smart. The limit shape of convex peeling. *In preparation*, 2017.

[15] F. Cao. Morphological scale space and mathematical morphology. In *Scale-Space Theories in Computer Vision*, pages 164–174. Springer, 1999.

[16] T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, Feb. 2001.

[17] V. Chepoi, K. Nouioua, E. Thiel, and Y. Vaxes. Pareto envelopes in simple polygons. *International Journal of Computational Geometry & Applications*, 20(06):707–721, 2010.

[18] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, pages 493–507, 1952.

[19] 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.

[20] M. G. Crandall, L. C. Evans, and P.-L. Lions. Some properties of viscosity solutions of hamilton-jacobi equations. *Transactions of the American Mathematical Society*, 282(2):487–502, 1984.

[21] M. G. Crandall and P.-L. Lions. Viscosity solutions of hamilton-jacobi equations. *Transactions of the American Mathematical Society*, 277(1):1–42, 1983.

[22] K. Dalal. Counting the onion. *Random Structures & Algorithms*, 24(2):155–165, 2004.

[23] M. Damian, E. D. Demaine, M. L. Demaine, V. Dujmovic, D. El-Khechen, R. Flatland, J. Iacono, S. Langerman, H. Meijer, S. Ramaswami, D. L. Souvaine, P. Taslakian, and G. T. Toussaint. Curves in the sand: Algorithmic drawing. In *Canadian Conference on Computational Geometry*, 2006.

[24] K. Deb. *Multi-objective optimization using evolutionary algorithms*. Wiley, Chichester, UK, 2001.

[25] 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.

[26] J.-D. Deuschel and O. Zeitouni. Limiting curves for i.i.d. records. *The Annals of Probability*, 23(2):852–878, 1995.

[27] D. L. Donoho and M. Gasko. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, pages 1803–1827, 1992.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] C. Fonseca and P. Fleming. An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary Computation*, 3(1):1–16, 1995.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] A. Hero and G. Fleury. Pareto-optimal methods for gene ranking. *VLSI Signal Processing*, 38(3):259–275, 2004.

[39] V. J. Hodge and J. Austin. A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2):85–126, 2004.

[40] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[41] J. Jiang, M. Mitzenmacher, and J. Thaler. Parallel peeling algorithms. *ACM Transactions on Parallel Computing (TOPC)*, 3(1):7, 2016.

[42] R. M. Karp and M. Sipser. Maximum matching in sparse random graphs. In *Foundations of Computer Science, 1981. SFCS'81. 22nd Annual Symposium on*, pages 364–375. IEEE, 1981.

[43] J. F. Kingman. The ergodic theory of subadditive stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 499–510, 1968.

[44] J. F. C. Kingman. *Poisson processes*. Wiley Online Library, 1993.

[45] 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.

[46] 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.

[47] R. Y. Liu, J. M. Parelius, and K. Singh. Multivariate analysis by data depth: descriptive statistics, graphics and inference. *The Annals of Statistics*, 27(3):783–858, 1999.

[48] B. F. Logan and L. A. Shepp. A variational problem for random Young tableaux. *Advances in Mathematics*, 26(2):206–222, 1977.

[49] R. Lou and M. Sarrafzadeh. An optimal algorithm for the maximum three-chain problem. *SIAM Journal on Computing*, 22(5):976–993, 1993.

[50] N. Madavan. Multiobjective optimization using a Pareto differential evolution approach. In *Proceedings of the Congress on Evolutionary Computation (CEC)*, 2002.

[51] B. Merriman, J. K. Bence, and S. Osher. *Diffusion generated motion by mean curvature.* Department of Mathematics, University of California, Los Angeles, 1992.

[52] 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.

[53] L. Moisan. Affine plane curve evolution: a fully consistent scheme. *IEEE Transactions on Image Processing*, 7(3):411–420, 1998.

[54] 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.

[55] 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.

[56] P. Pevzner. *Computational Molecular Biology.* The MIT Press, 2000.

[57] M. Poulos, S. Papavlasopoulos, and V. Chrissikopoulos. An application of the onion peeling algorithm for fingerprint verification purposes. *Journal of Information and Optimization Sciences*, 26(3):665–681, 2005.

[58] P. Rousseeuw and A. Struyf. Computation of robust statistics: depth, median, and related measures. In J. Goodman and J. O'Rourke, editors, *Handbook of Discrete and Compututational Geometry*, Discrete Mathematics and its Applications, pages 1279–1292. Chapman & Hall-CRC Press, 2004.

[59] G. Sapiro and A. Tannenbaum. Affine invariant scale-space. *International Journal of Computer Vision*, 11(1):25–44, 1993.

[60] C. G. Small. A survey of multidimensional medians. *International Statistical Review*, pages 263–277, 1990.

[61] T. Speed. *Statistical analysis of gene expression microarray data*, volume 11. Chapman & Hall/CRC, 2003.

[62] N. Srinivas and K. Deb. Muiltiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, 2(3):221–248, 1994.

[63] J. M. Steele. *Probability theory and combinatorial optimization.* SIAM, 1997.

[64] T. Suk and J. Flusser. Convex layers: a new tool for recognition of projectively deformed point sets. In *Computer Analysis of Images and Patterns*, pages 454–461. Springer, 1999.

[65] J.-F. Thisse, J. E. Ward, and R. E. Wendell. Some properties of location problems with block and round norms. *Operations Research*, 32(6):1309–1327, 1984.

[66] S. Ulam. Monte carlo calculations in problems of mathematical physics. *Modern Mathematics for the Engineers*, pages 261–281, 1961.

[67] 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.

[68] G. Viennot. Chain and antichain families, grids and Young tableaux. In *Orders: description and roles*, volume 99 of *North-Holland Mathematics Studies*, pages 409–463. 1984.