A PDE-BASED APPROACH TO NON-DOMINATED SORTING*

JEFF CALDER†, SELIM ESEDOḠLU†, AND ALFRED O. HERO‡

Abstract. Non-dominated sorting is a fundamental combinatorial problem in multiobjective optimization, and is equivalent to the longest chain problem in combinatorics and random growth models for crystals in materials science. In a previous work [5], we showed that non-dominated sorting has a continuum limit that corresponds to solving a Hamilton–Jacobi equation. In this work we present and analyze a fast numerical scheme for this Hamilton–Jacobi equation, and show how it can be used to design a fast algorithm for approximate non-dominated sorting.

Key words. Hamilton–Jacobi equations, numerical schemes, algorithms, viscosity solutions, longest chain in Euclidean space, non-dominated sorting, Pareto-optimality

AMS subject classifications. 35F21, 35D40, 65N12, 65N06, 35Q68, 06A07

1. Introduction.

Non-dominated sorting is a combinatorial problem that is fundamental in multiobjective optimization, which is ubiquitous is scientific and engineering contexts [12, 8, 9]. The sorting can be viewed as arranging a finite set of points in Euclidean space into layers according to the componentwise partial order. The layers are obtained by repeated removal of the set of minimal elements. More formally, given a set $\mathcal{X}_n \subset \mathbb{R}^d$ of n points equipped with the componentwise partial order $\leqq$[1], the first layer, often called the first Pareto front and denoted $\mathcal{F}_1$, is the set of minimal elements in $\mathcal{X}_n$. The second Pareto front $\mathcal{F}_2$ is the set of minimal elements in $\mathcal{X}_n \setminus \mathcal{F}_1$, and in general the k^{th} Pareto front $\mathcal{F}_k$ is given by

$$\mathcal{F}_k = \text{mimimal elements of } \mathcal{X}_n \setminus \bigcup_{i \leq k-1} \mathcal{F}_i.$$

In the context of multiobjective optimization, the d coordinates of each point in $\mathcal{X}_n$ are the values of the d objective functions evaluated on a given feasible solution. In this way, each point in $\mathcal{X}_n$ corresponds to a feasible solution and the layers provide an effective ranking of all feasible solutions with respect to the given optimization problem. Rankings obtained in this way are at the heart of genetic and evolutionary algorithms for multiobjective optimization, which have proven to be valuable tools for finding solutions numerically [8, 9, 14, 15, 27]. Figure 1.1 gives a visual illustration of Pareto fronts for randomly generated points.

It is important to note that non-dominated sorting is equivalent to the longest chain problem in combinatorics, which has a long history beginning with Ulam's famous problem of finding the length of a longest increasing subsequence in a sequence of numbers (see [30, 17, 3, 11, 5] and the references therein). The longest chain problem is then intimately related to several other problems in combinatorics and graph theory [13, 23, 32], materials science [26], and molecular biology [25]. To see this connection, let $u_n(x)$ denote the length of a longest chain[2] in $\mathcal{X}_n$ consisting of points less than or equal to x with respect to $\leqq$. If all points in $\mathcal{X}_n$ are distinct, then

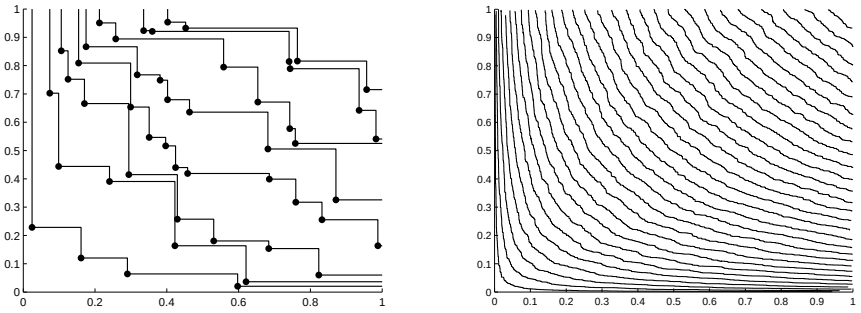
*The research in this paper was partially supported by NSF grants DMS-0914567 and CCF-1217880, and by ARO grant W911NF-09-1-0310.

†Department of Mathematics, University of Michigan. (`{jcalder,esedoglu}@umich.edu`)

‡Department of Electrical Engineering and Computer Science, University of Michigan. (`hero@eecs.umich.edu`)

[1] $x \leqq y \iff x_i \leq y_i$ for $i = 1, \ldots, d$.

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



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

FIG. 1.1. *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 out of the 1938 total fronts.*

a point $x \in \mathcal{X}_n$ is a member of $\mathcal{F}_1$ if and only if $u_n(x) = 1$. By peeling off $\mathcal{F}_1$ and making the same argument, we see that $x \in \mathcal{X}_n$ is a member of $\mathcal{F}_2$ if and only if $u_n(x) = 2$. In general, for any $x \in \mathcal{X}_n$ we have

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

This is a fundamental observation. It says that studying the shapes of the Pareto fronts is equivalent to studying the longest chain function u_n.

The longest chain problem has well-understood asymptotics as $n \to \infty$. In this context, we assume that $\mathcal{X}_n = \{X_1, \dots, X_n\}$ where $X_1, \dots, X_n$ are *i.i.d.* random variables in $\mathbb{R}^n$ and let $\ell(n)$ denote the length of a longest chain in $\mathcal{X}_n$. The seminal work on the problem was done by Hammersley [17], who studied the problem for $X_1, \dots, X_n$ *i.i.d.* uniform on $[0,1]^2$. He utilized subadditive ergodic theory to show that $n^{-\frac{1}{2}}\ell(n) \to c$ in probability, where $c > 0$. He conjectured that $c = 2$, and this was later proven by Vershik and Kerov [31] and Logan and Shepp [22]. Hammersley's results were generalized to higher dimensions by Bollobás and Winkler [3], who showed that $n^{-\frac{1}{d}}\ell(n) \to c_d$ almost surely, where $0 < c_d < e$ are constants tending to e as $d \to \infty$. The only known values of c_d are $c_1 = 1$ and $c_2 = 2$. Deuschel and Zeitouni [11] provided another generalization of Hammersley's results; for $X_1, \dots, X_n$ *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.

In [5], we studied the longest chain problem for $X_1, \dots, X_n$ *i.i.d.* on $\mathbb{R}^d$ with density function $f : \mathbb{R}^d \to \mathbb{R}$ that satisfies the following:

(H) There exists an open and bounded set $\Omega \subset (0,1)^d$ with Lipschitz boundary such that $f|_{\overline{\Omega}}$ is continuous and $\text{supp}(f) \subset \overline{\Omega}$.

This hypothesis simply states that $X_1, \dots, X_n$ live in the domain Ω and have a continuous density function f on that domain. Assuming (H) holds, we showed that

$n^{-\frac{1}{d}} u_n \longrightarrow c_d d^{-1} U$ in $L^\infty(\mathbb{R}^d)$ almost surely, where U is the viscosity solution of the Hamilton–Jacobi equation

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

Here $\mathbb{R}_+ = (0, \infty)$ and $\mathbb{R}^d_+ = (\mathbb{R}_+)^d$. The boundary condition $U = 0$ on $\partial\mathbb{R}^d_+$ can be interpreted as stipulating that the "zeroth" Pareto front lies on $\partial\mathbb{R}^d_+$. We also showed that when f satisfies (H) there exists a unique Pareto-monotone[3] viscosity solution of (P) satisfying the additional boundary condition at infinity given by (3.2). In a another more recent work [4] we showed that this additional boundary condition is actually redundant.

In this paper we study a fast numerical scheme for (P), first proposed in [5], and prove convergence of this scheme. We then show how the scheme can be used to design a fast approximate non-dominated sorting algorithm, which we call *PDE-based ranking*, and we evaluate the sorting accuracy of PDE-based ranking on both real and synthetic 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 [9]. There are also potential applications in polynuclear growth of crystals in materials science [26]. Here, the scheme for (P) could be used to simulate polynuclear growth in the presence of a macroscopically inhomogeneous growth rate.

The proof of convergence of our scheme is modeled on the well-known framework of Barles and Souganidis [1]. The proof does not follow directly from [1] since (P) does not satisfy the strong uniqueness property, which is a comparison principle for semicontinuous viscosity solutions. We instead show that one can exploit a type of approximate Hölder-regularity of the numerical solutions to bypass strong uniqueness and substitute the ordinary uniqueness result that is known for (P).

The advantages of PDE-based ranking over existing algorithms are twofold. First, it is computationally more efficient in relatively low dimensions, and, in applications where only a subset of the data needs to be ranked, it can be sublinear in n. To the best of our knowledge, there are no existing algorithms that compare computationally for low dimensional problems. Second, although it is an approximate sorting algorithm, it is theoretically justified, and is guaranteed to give the exact sorting in the continuum limit. In Section 5.3 we propose another fast algorithm for non-dominated sorting based on sorting a small random subset of the data and interpolating. We call this algorithm *subset ranking*, and it compares well computationally to PDE-based ranking. However, it is not as accurate, and is based on heuristics for which there is no theoretical justification.

This paper is organized as follows. In Section 3 we prove that the numerical solutions converge to the viscosity solution of (P). We also prove a regularity result for the numerical solutions (see Lemma 3.3) and other important properties. In Section 4 we demonstrate the numerical scheme on several density functions, and in Section 5 we propose a fast algorithm for approximate non-dominated sorting that is based on numerical solving (P).

2. Numerical scheme. Let us first fix some notation. Given $x, y \in \mathbb{R}^d$ we write $x \leq y$ if $x \leqq y$ and $x \neq y$. We write $x < y$ when $x_i < y_i$ for all i. For $s, t \in \mathbb{R}$, $\leq$ and

[3] We say a function $u : \mathcal{O} \subset \mathbb{R}^d \to \mathbb{R}$ is Pareto-monotone if $x \leqq y \implies u(x) \leq u(y)$ for all $x, y \in \mathcal{O}$.

$<$ will retain their usual definitions. For $x \leqq y$ we define

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

and make similar definitions for $[x,y)$ and (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 $\mathbf{0} = (0,\dots,0) \in \mathbb{R}^d$ and $\mathbf{1} = (1,\dots,1) \in \mathbb{R}^d$. For $z \in [\mathbf{0},\infty)$, we denote by $\pi_z : \mathbb{R}^d \to [0,z]$ the projection mapping $\mathbb{R}^d$ onto $[\mathbf{0},z]$. For $x \in [\mathbf{0},\infty)$ this mapping is given explicitly by

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

We now recall the numerical scheme from [5]. Let $h > 0$ be the grid spacing. 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

$$\prod_{i=1}^{d} (U_h(x) - U_h(x - he_i)) = h^d f(x), \tag{2.1}$$

where $U_h : h\mathbb{N}_0^d \to \mathbb{R}$ is the numerical solution of (P) and $e_1,\dots,e_d$ are the standard basis vectors in $\mathbb{R}^d$. Given $U_h(x-he_1),\dots,U_h(x-he_d)$ and $f(x)$, there are in general d values of $U_h(x)$ that solve (2.1). However, since we are interested in the Pareto-monotone viscosity solution of (P), we should impose the constraint

$$U_h(x) \geq \max(U_h(x - he_1),\dots,U_h(x - he_d)), \tag{2.2}$$

Since the left hand side of (2.1) is increasing in $U_h(x)$ when (2.2) holds, we see that for $f(x) \geq 0$ there is *exactly one* value for $U_h(x)$ that satisfies (2.1) and (2.2). We define our numerical scheme by taking this distinguished value. At each grid point $x \in h\mathbb{N}^d$, $U_h(x)$ satisfying (2.1)–(2.2) can be computed numerically by either a binary search and/or Newton's method restricted to the interval

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

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

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

The numerical solution U_h is computed by visiting each grid point exactly once via any sweeping pattern that respects the partial order $\leqq$, and imposing the boundary condition $U_h(x) = 0$ for $x \in \partial\mathbb{R}^d_+$. The scheme therefore has linear complexity in the number of gridpoints.

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

$$S(h, x, u) = \prod_{i=1}^{d} \frac{u(x) - u(x - he_i)}{h}. \tag{2.3}$$

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).

3. 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). As in [5], we place the following assumption on $f : \mathbb{R}^d \to [0, \infty)$:

(H) There exists an open and bounded set $\Omega \subset (0, 1)^d$ with Lipschitz boundary such that $f|_{\overline{\Omega}}$ is continuous and $\operatorname{supp}(f) \subset \overline{\Omega}$.

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

$$\min\{F(x) \,:\, x \in \mathcal{K}\}, \tag{3.1}$$

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 (3.1), 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 $\mathcal{K}$ can induce a discontinuity in f along $\partial\Omega$.

In [5] we showed that, under hypothesis (H), there exists a unique Pareto-monotone viscosity solution U of (P) satisfying the additional boundary condition

$$U(x) = U(\pi_{\mathbf{1}}(x)) \quad \text{for all } x \in [\mathbf{0}, \infty). \tag{3.2}$$

The boundary condition (3.2) is natural for this problem. Indeed, since $\operatorname{supp}(f) \subset (0, 1)^d$, there are almost surely no random variables drawn outside of $(0, 1)^d$. Hence, for any $x \in [\mathbf{0}, \infty)$ we can write

$$u_n(x) = \max_{y \in [0,1]^d \,:\, y \leqq x} u_n(y).$$

Since u_n is Pareto-monotone, the maximum above is attained at $y = \pi_{\mathbf{1}}(x)$, and hence $u_n(x) = u_n(\pi_{\mathbf{1}}(x))$.

For completeness, let us now give a brief outline of the proof of uniqueness for (P). For more details, we refer the reader to [5]. The proof is based on the auxiliary function technique, now standard in the theory of viscosity solutions [7]. However, the technique must be modified to account for the fact that f is possibly discontinuous on $\partial\Omega$, and hence does not possess the required uniform continuity. A commonly employed technique is to modify the auxiliary function so that only a type of one-sided uniform continuity is required of f [28, 10]. This allows f to, for example, have a discontinuity along a Lipschitz curve, provided the jump in f is locally in the same direction (see [10] for more details). We cannot directly use these results because they require coercivity or uniform continuity of the Hamiltonian and/or Lipschitzness of

solutions—none of which hold for (P). Our technique for proving uniqueness for (P) employs instead an important property of viscosity solutions of (P)—namely that for any $z \in \mathbb{R}^d_+$, $U^z := U \circ \pi_z$ is a viscosity subsolution of (P). This property, called *truncatability* in [5], follows immediately from the variational principle [5]

$$U(x) = \sup_{\gamma' \geqq 0 \,:\, \gamma(1)=x} \int_0^1 f(\gamma(t))^{\frac{1}{d}} (\gamma_1'(t) \cdots \gamma_d'(t))^{\frac{1}{d}} \, dt.$$

This allows us to prove a comparison principle with no additional assumptions on the Hamiltonian.

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 [1]. Their framework requires that the scheme be stable, monotone, consistent, and that the PDE satisfy a *strong uniqueness property* [1]. 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 [1] 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 3.1) and is proved in Lemma 3.3. The monotonicity property requires the following:

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

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 (H). Since f may be discontinuous along $\partial\Omega$, we can only establish a comparison principle for continuous viscosity sub- and supersolutions (see [5, Theorem 4]).

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 (H) 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 3.3 we establish approximate Hölder regularity of U_h of the form

$$|U_h(x) - U_h(y)| \leq C(|x - y|^{\frac{1}{d}} + h^{\frac{1}{d}}). \tag{3.3}$$

As we verify in Appendix A, the approximate Hölder estimate (3.3) along with the stability of (S) allows us to apply the Arzelà-Ascoli Theorem to the sequence U_h, and extract a subsequence that converges uniformly to a Hölder-continuous function u. Since (S) is consistent and monotone, it is a standard result that u is a viscosity solution of (P). The convergence proof is then completed by invoking uniqueness of Pareto-monotone viscosity solutions of (P) [5, Theorem 5].

3.1. 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 3.3. For the remainder of this section, we fix $h > 0$.

LEMMA 3.1 (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*

$$S(h, x, u) \leq S(h, x, v) \quad \text{for all } x \in (h\mathbf{1}, z]. \tag{3.4}$$

Then $u \leq v$ on $\Gamma_h \cap [\mathbf{0}, z]$ implies that $u \leq v$ on $[\mathbf{0}, z]$.

Proof. Suppose that $\sup_{[\mathbf{0},z]}(u - v) > 0$ and set

$$T_r = \{x \in [\mathbf{0}, \infty) : (x_1 + \cdots + x_d) \leq rd\} \text{ and } R = \sup\{r > 0 : u \leq v \text{ on } T_r \cap [\mathbf{0}, z]\}.$$

Since $u \leq v$ 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 $u \leq v$ on $T_s \cap [\mathbf{0}, z]$ and hence

$$u(x - he_i) \leq v(x - he_i) \leq v(x) \quad \text{for } i = 1, \ldots, d. \tag{3.5}$$

The second inequality above follows from Pareto-monotonicity of v. Since u and v are Pareto-monotone and $u(x) > v(x)$ we have

$$\prod_{i=1}^{d}(u(x) - u(x - he_i)) > \prod_{i=1}^{d}(v(x) - u(x - he_i)) \overset{(3.5)}{\geq} \prod_{i=1}^{d}(v(x) - v(x - he_i)).$$

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 the boundary condition at infinity (3.2).

PROPOSITION 3.2. *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*

$$\operatorname{supp}\{x \mapsto S(h, x, u)\} \subset [\mathbf{0}, z]. \tag{3.6}$$

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 3.1 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 3.3. *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*

$$|u(x) - u(y)| \le 2d^2 R^{\frac{d-1}{d}} \|S(h,\cdot,u)\|^{\frac{1}{d}}_{L^\infty((h,R]^d)} (|x-y|^{\frac{1}{d}} + h^{\frac{1}{d}}) \tag{3.7}$$

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

$$\psi(x) = \begin{cases} d(x_1 \cdots x_d)^{\frac{1}{d}} & \text{if } x \in (\mathbf{0},\infty), \\ 0 & \text{otherwise.} \end{cases} \tag{3.8}$$

By the concavity of $t \mapsto t^{\frac{1}{d}}$ we have

$$\psi(x) - \psi(x - he_i) = d(x_1 \cdots x_d)^{\frac{1}{d}} x_i^{-\frac{1}{d}} (x_i^{\frac{1}{d}} - (x_i - h)^{\frac{1}{d}}) \ge x_i^{-1} (x_1 \cdots x_d)^{\frac{1}{d}} h,$$

for any $x \in (h\mathbf{1},\infty)$ and hence

$$S(h,x,\psi) \ge 1 \quad \text{for all} \quad x \in (h\mathbf{1},\infty). \tag{3.9}$$

By the translation invariance of S and (3.9) we have

$$S(h,x,\psi(\cdot - b)) \ge 1 \quad \text{for all } b \in [\mathbf{0},\infty),\ x \in (b + h\mathbf{1},\infty). \tag{3.10}$$

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

$$\begin{aligned} S(h,x,w) &\ge \frac{1}{h^d} \prod_{i=1}^d \Big(\widehat{u}(x) - \widehat{u}(x - he_i) \\ &\qquad + \|S(h,\cdot,u)\|^{\frac{1}{d}}_{L^\infty((h,R]^d)} (\psi(x - b^k) - \psi(x - b^k - he_i)) \Big) \\ &\ge S(h,x,\widehat{u}) + \|S(h,\cdot,u)\|_{L^\infty((h,R]^d)} S(h,x,\psi(\cdot - b^k)) \\ &\overset{(3.10)}{\ge} S(h,x,\widehat{u}) + \|S(h,\cdot,u)\|_{L^\infty((h,R]^d)} \\ &\ge S(h,x,u). \end{aligned}$$

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 $S(h,x,w) \ge S(h,x,u)$. Since $w \ge u = 0$ on $\Gamma_h \cap [0,R]^d$, we can apply Lemma 3.1 to obtain $w \ge u$ on $[0,R]^d$, which yields

$$\begin{aligned} u(y_0) - \widehat{u}(y_0) &\le \|S(h,\cdot,u)\|^{\frac{1}{d}}_{L^\infty((h,R]^d)} \sum_{i=1}^d \psi(y_0 - b^i) \\ &\le dR^{\frac{d-1}{d}} \|S(h,\cdot,u)\|^{\frac{1}{d}}_{L^\infty((h,R]^d)} \sum_{i=1}^d (y_{0,i} - x_{0,i} + h)^{\frac{1}{d}} \\ &\le d^2 R^{\frac{d-1}{d}} \|S(h,\cdot,u)\|^{\frac{1}{d}}_{L^\infty((h,R]^d)} (|x_0 - y_0|^{\frac{1}{d}} + h^{\frac{1}{d}}). \end{aligned} \tag{3.11}$$

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. □

3.2. Convergence theorem. Our main result is the following convergence statement for the scheme (S).

THEOREM 3.4. *Let f be nonnegative and satisfy (H). Let U be the unique Pareto-monotone viscosity solution of (P) satisfying* (3.2). *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 (H) we have that $f(x) = 0$ for $x \notin (0, 1)^d$, and hence $\operatorname{supp}(f(\lfloor \cdot \rfloor_h)) \subset [0, 1]^d$. Therefore, by Proposition 3.2, we have that U_h satisfies (3.2). Combining this with Lemma 3.3 we have

$$\|U_h\|_{L^\infty([\mathbf{0},\infty))} \leq C \|f\|^{\frac{1}{d}}_{L^\infty([\mathbf{0},\infty))}, \tag{3.12}$$

for all $h > 0$. Similarly, combining (3.2) with Lemma 3.3 we have

$$|U_h(x) - U_h(y)| \leq 2d^2 \|f\|^{\frac{1}{d}}_{L^\infty([\mathbf{0},\infty))} (|x - y|^{\frac{1}{d}} + h^{\frac{1}{d}}) \text{ for all } x, y \in [\mathbf{0}, \infty), \tag{3.13}$$

for every $h > 0$. The estimates in (3.12) and (3.13) 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 (3.2), 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) [1]. Note that U_h is Pareto-monotone, $U_h = 0$ on Γ_h, and U_h satisfies (3.2). Since $U_{h_k} \to u$ uniformly, it follows that u is Pareto-monotone, $u = 0$ on $\partial\mathbb{R}^d_+$, and u satisfies (3.2). By uniqueness for (P) [5, Theorem 5] 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 $[\mathbf{0}, \infty)$. □

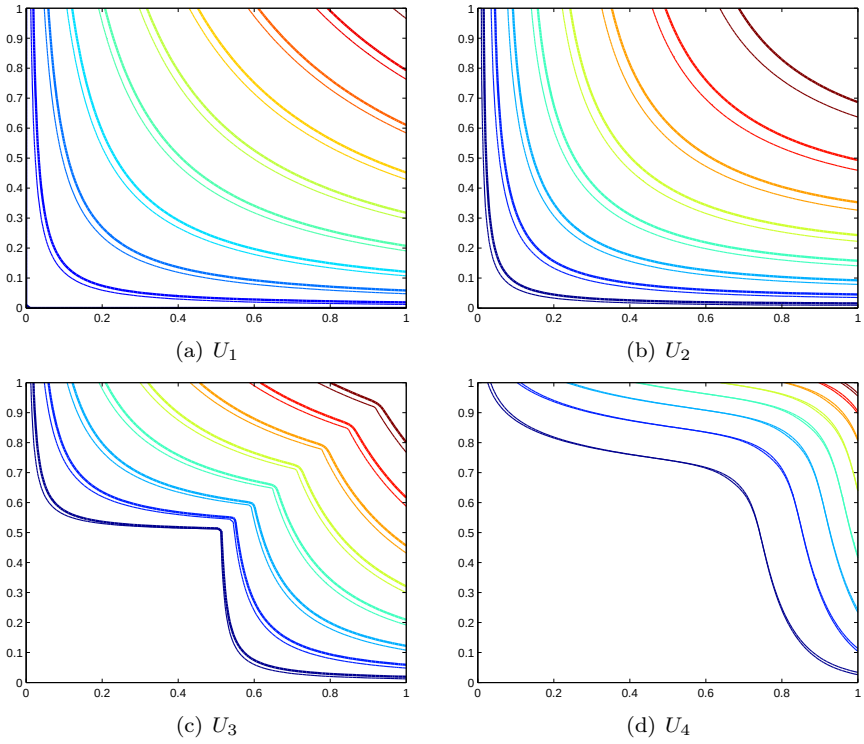
In Section 4, we observe that the numerical scheme provides a fairly consistent underestimate of the exact solution of (P). The following lemma shows that this is indeed the case whenever the solution U of (P) is concave.

LEMMA 3.5. *Let f be nonnegative and satisfy (H). Let U be the unique Pareto-monotone viscosity solution of (P) satisfying* (3.2). *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.[4] 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

$$U(x) - U(x - he_i) \geq hU_{x_i}(x) \text{ for all } i.$$

[4]The fact that U is Pareto-monotone also implies differentiability almost everywhere.



(a) U_1 (b) U_2

(c) U_3 (d) U_4

FIG. 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.*

Since U is a viscosity solution of (P) and f is continuous at x we have

$$S(h, x, U) \geq U_{x_1}(x) \cdots U_{x_d}(x) = f(x).$$

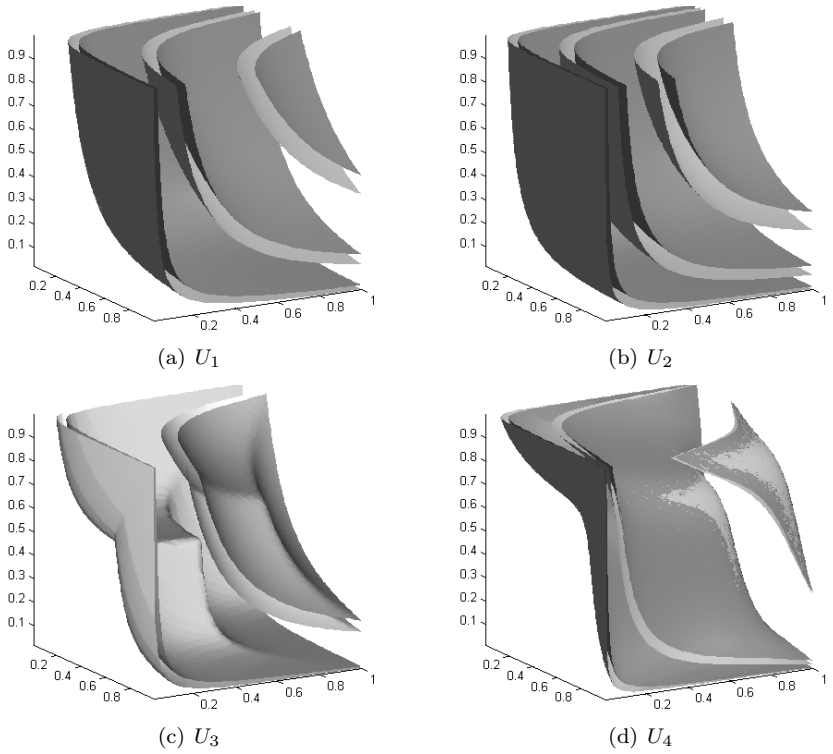
Since $x \mapsto S(h, x, U)$ is continuous, we see that $S(h, x, U) \geq f_*(x) = f(x)$ for all $x \in (h\mathbf{1}, \infty]$. Now define $W_h(x) = U(\lfloor x \rfloor_h)$. Then we have

$$S(h, x, W_h) \geq f(\lfloor x \rfloor_h) \quad \text{for all } x \in (h\mathbf{1}, x],$$

and $W_h = 0$ on Γ_h. It follows from Lemma 3.1 that $U_h \leq W_h$. Since U is Pareto-monotone, we have $W_h \leq U$, which completes the proof. □

4. 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 $f_1(x) = 1$, $f_2(x) = \frac{2^d}{\pi^{d/2}} e^{-|x|^2}$,

$$f_3(x) = 1 - \chi_{[0,1/2]^d}(x) \quad \text{and} \quad f_4(x) = \left(\sum_{i=1}^d x_i^9 \right)^{1-d} \prod_{i=1}^d \left(9x_i^9 + \sum_{i=1}^d x_i^9 \right).$$



(a) U_1 (b) U_2

(c) U_3 (d) U_4

FIG. 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.*

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

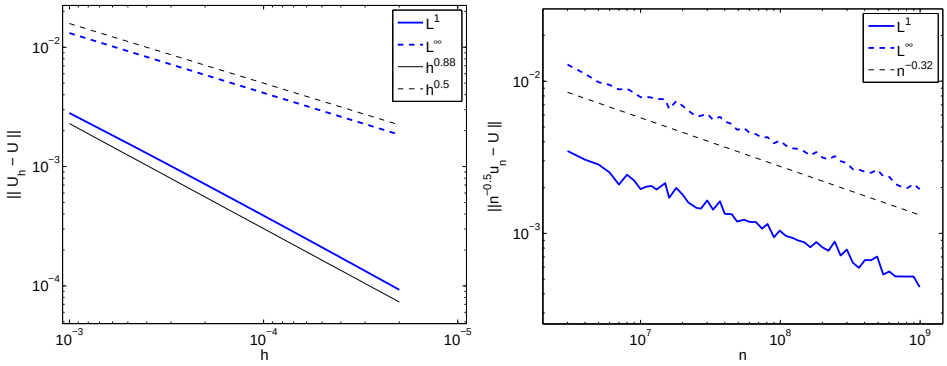
$$U_3(x) = d \max_{i\in\{1,\dots,d\}} \left\{ \left(x_i - \frac{1}{2}\right)_+ \prod_{j\neq i} x_j \right\}^{\frac{1}{d}}, \quad U_4(x) = d\left(\prod_{i=1}^d x_i \cdot \sum_{i=1}^d x_i^9\right)^{\frac{1}{d}},$$

where $\operatorname{erf}(x)$ is the error function defined by $\operatorname{erf}(x) = 2/\sqrt{\pi}\int_0^x e^{-t^2}\,dt$, and $x_+ := \max(0, x)$. The solutions U_1 and U_2 are special cases of the formula

$$U(x) = d\left(\int_{[\mathbf{0},x]} f(y)\,dy\right)^{\frac{1}{d}}, \tag{4.1}$$

which holds when f is separable, i.e., $f(x) = f_1(x_1)\cdots f_d(x_d)$ [5]. 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 [12, 5].

We computed the numerical solutions for $d = 2$ and $d = 3$. For $d = 2$ we used a 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



(a) Convergence rate of scheme (S) (b) Stochastic convergence rates

FIG. 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.*

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 3.5 that $U_h \leq U$ 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 3.5 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.

4.1. 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. 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

$$\|U_h - U\|_{L^\infty(\mathbb{R}^d_+)} \leq C h^{\frac{1}{d}}, \tag{4.2}$$

for some constant $C > 0$. We intend to investigate this in a future work. 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

$$|n^{-\frac{1}{d}} u_n - c_d d^{-1} U|_{L^\infty} := \max_{1 \leq i \leq n} |n^{-\frac{1}{d}} u_n(X_i) - c_d d^{-1} U(X_i)| \tag{4.3}$$

and

$$|n^{-\frac{1}{d}}u_n - c_d d^{-1}U|_{L^1} := \frac{1}{n}\sum_{i=1}^{n} |n^{-\frac{1}{d}}u_n(X_i) - c_d d^{-1}U(X_i)| \tag{4.4}$$

Figure 4.3(b) shows (4.3) and (4.4) versus n for the same density f_3. For each n the values of (4.3) and (4.4) 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 (4.4) and $\alpha = 0.3144$ for the L^∞ norm (4.3). These results are in line with the known convergence rates for the longest chain problem with a uniform distribution on $[0,1]^d$ [2].

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).

5. 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 5.4. We assume here that we are given data $X_1, \ldots, X_n$ that *i.i.d.* samples 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.

In practice, one is given only the data $X_1, \ldots, X_n$, and does not know the underlying distribution f, which is required for solving (P). Therefore, the first step in our algorithm will be to form an estimate $\widehat{f}$ of f from 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 [29] or a k-nearest neighbor estimator [21]. 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). Let us denote the grid spacing by h. 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 $\widehat{f}_h$ 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

$$\prod_{i=1}^{d} (\widehat{U}_h(y) - \widehat{U}_h(y - h_i e_i)) = h_1 \cdots h_d \widehat{f}_h(x), \tag{5.1}$$

where $h_i = y_i - (x_i - h)$. In (5.1) we compute $\widehat{U}_h(y - h_i e_i)$ by linear interpolation using adjacent grid points. Figure 5.1 illustrates the grid used for computing $\widehat{U}_h(y)$.

We call this algorithm *PDE-based ranking*, and the algorithm is summarized below.

Algorithm 1 (PDE-based ranking). *Fast approximate non-dominated sorting*

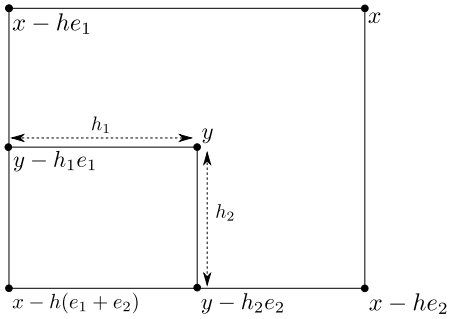


FIG. 5.1. *Depiction of the grid used for computing $\widehat{U}_h(y)$ according to (5.1). The values of $\widehat{U}_h(y - h_1e_1)$ and $\widehat{U}_h(y - h_2e_2)$ are computed by linear interpolation using adjacent grid points, i.e., $\widehat{U}_h(y - h_1e_1)$ is computed via linearly interpolating between $\widehat{U}_h(x - he_1)$ and $\widehat{U}_h(x - h(e_1 + e_2))$.*

1. *Select k points from $X_1, \dots, X_n$ at random. Call them $Y_1, \dots, Y_k$.*
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.,*

$$\widehat{f}_h(x) = \frac{1}{kh^d} \cdot \#\left\{ Y_i \,:\, x \leqq Y_i \leqq x + h\mathbf{1} \right\} \text{ for } x \in h\mathbb{N}_0^d. \tag{5.2}$$

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 $i = 1, \dots, n$ via interpolation.*

For simplicity of discussion, we have assumed that $X_1, \dots, X_n$ 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 1.

5.1. Convergence rates for PDE-based ranking. It is important to understand how the parameters k and h in PDE-based ranking (Algorithm 1) affect the accuracy of the estimate $\widehat{U}_h$. We first consider the estimate $\widehat{f}_h$. By (5.2), 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

$$p = \int_{[x, x+h\mathbf{1}]} f(y)\, dy. \tag{5.3}$$

By the central limit theorem, the fluctuations of $\widehat{f}_h(x)$ about its mean satisfy

$$\left| \widehat{f}_h(x) - \frac{p}{h^d} \right| \leq C \frac{1}{\sqrt{k} h^d}, \tag{5.4}$$

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 (5.3) we have

$$\left| f(x) - \frac{p}{h^d} \right| \leq C\sqrt{d} h.$$

Combining this with (5.4) we have

$$\|\widehat{f}_h - f\|_{L^\infty([0,1]^d \cap h\mathbb{N}^d)} \leq C\left(\frac{1}{\sqrt{k}h^d} + h\right), \tag{5.5}$$

with high probability. By the discrete comparison principle (Lemma 3.1) and (5.5) we have that

$$\|\widehat{U}_h - U_h\|_{L^\infty([0,1]^d)} \leq d\|\widehat{f}_h - f\|^{\frac{1}{d}}_{L^\infty([0,1]^d \cap h\mathbb{N}^d)} \leq C\left(k^{-\frac{1}{2d}}h^{-1} + h^{\frac{1}{d}}\right), \tag{5.6}$$

with high probability. Based on the numerical evidence presented in Section 4.1, it is reasonable to suspect that $\|U - U_h\|_{L^\infty([0,1]^d)} \leq Ch^{\frac{1}{d}}$. If this is indeed the case, then in light of (5.6) we have

$$\|\widehat{U}_h - U\|_{L^\infty([0,1]^d)} \leq C\left(k^{-\frac{1}{2d}}h^{-1} + h^{\frac{1}{d}}\right), \tag{5.7}$$

with high probability.

The right side of the inequality (5.7) 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 [20]. Similarly to [20] we can use the bound in (5.7) 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 (5.7). This yields $h = k^{-\frac{1}{2(d+1)}}$ and (5.7) becomes

$$\|\widehat{U}_h - U\|_{L^\infty([0,1]^d)} \leq Ck^{-\frac{1}{2d(d+1)}} = Ch^{\frac{1}{d}}, \tag{5.8}$$

with high probability.

In PDE-based ranking, we rank the samples with $\widehat{U}_h$ in place of u_n. Consider the corresponding L^1 sorting error

$$\begin{aligned}\left|c_d d^{-1}\widehat{U}_h(X_i) - n^{-\frac{1}{d}}u_n(X_i)\right|_{L^1} &\leq \left|c_d d^{-1}\widehat{U}_h(X_i) - U(X_i)\right|_{L^1} \\ &\quad + \left|c_d d^{-1}U(X_i) - n^{-\frac{1}{d}}u_n(X_i)\right|_{L^1} \\ &\overset{(5.7)}{\leq} C\big(k^{-\frac{1}{2d}}h^{-1} + h^{\frac{1}{d}}\big) + \left|c_d d^{-1}U(X_i) - n^{-\frac{1}{d}}u_n(X_i)\right|_{L^1},\end{aligned} \tag{5.9}$$

which holds with high probability. The right hand side of (5.9) decomposes the sorting error into two terms. The first term captures the effect of errors in estimating f and solving (P) numerically, while the second term captures errors due to approximating non-dominated sorting by its continuum limit.

Notice that Steps 1-3 in PDE-based ranking, i.e., computing $\widehat{U}_h$, require $O(k + h^{-d})$ operations. In practice, it is often the case that one need not rank all n samples (e.g., in a streaming application [16]). So suppose we are required to sort, relative to the entire dataset, a subset size $N \ll n$, and suppose we have a sorting error tolerance of $\varepsilon > 0$. We can use (5.9) to choose k and h so that

$$\left|c_d d^{-1}\widehat{U}_h(X_i) - n^{-\frac{1}{d}}u_n(X_i)\right|_{L^1} \leq \frac{\varepsilon}{2} + \left|c_d d^{-1}U(X_i) - n^{-\frac{1}{d}}u_n(X_i)\right|_{L^1} \tag{5.10}$$

	$\lambda_{i,1}$	$\lambda_{i,2}$	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 5.1
Parameter values for mixture of Gaussians density

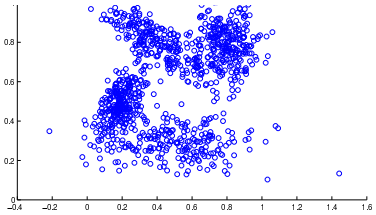


FIG. 5.2. *Depiction of random samples from the mixture of Gaussians density.*

holds with high probability. For example, we can choose $h = k^{-\frac{1}{2(d+1)}}$ to equate the two terms in (5.7) and then choose k large enough. Using PDE-based ranking to sort the subset of size N then has complexity $O(k + h^{-d} + N)$. When n is large enough, the L^1 sorting error will be less than ε by (5.10). If N is constant in n, or a sublinear function of n, then PDE-based ranking provides a *sublinear in* n algorithm for ranking all N points relative to the entire dataset with an error less than ε. Existing non-dominated sorting algorithms would need to sort all n points in order to rank a subset of size N, and hence have complexity at least $O(n \log n)$.

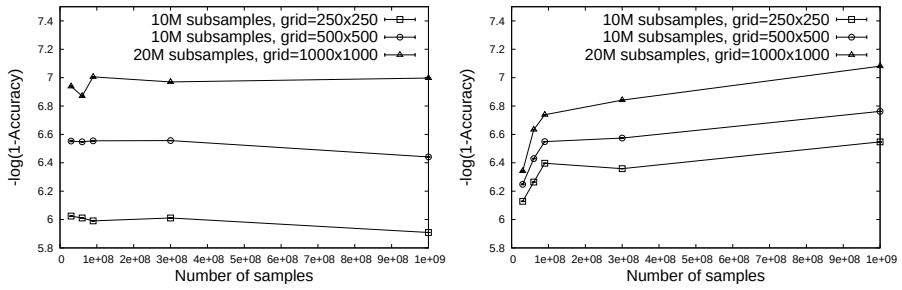
We emphasize that the sublinear nature of the algorithm lies in the computation of $\widehat{U}_h$, which allows one to approximately rank any point in $O(1)$ time. Ranking all samples, i.e., evaluating $\widehat{U}_h$ at each of $X_1, \ldots, X_n$, of course requires $O(n)$ operations. We show in Section 5.2 that even in this case, PDE-based ranking is 8-10 times faster than non-dominated sorting in two dimensions.

5.2. Evaluation of PDE-based ranking. In this section, we evaluate PDE-based ranking in dimension $d = 2$ for the discontinuous density $f_3(x) = 1 - \chi_{[0,1/2]^d}(x)$ and a mixture of Gaussians density 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 5.1, and the density is illustrated in Figure 5.2. In Section 5.4, we provide further evaluation of our proposed PDE-based ranking on real-world data from an anomaly detection problem.

In practical applications of non-dominated sorting, the numerical ranks assigned to each data point are important only insomuch as they provide a relative ranking among samples. Thus, rankings that differ only by composition with monotone increasing functions should be regarded as equivalent. As a consequence, the usual L^p norms are inadequate and irrelevant for measuring sorting accuracy. We propose an accuracy measure for comparing PDE-based ranking to exact non-dominated sorting that measures the fraction of pairs (X_i, X_j) that are ordered correctly. In this way, the accuracy score only takes into account the relative orderings between pairs of data points. Since the true Pareto rank of X_i is given by $u_n(X_i)$, this accuracy score can be expressed as

$$\text{Accuracy} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \psi(u_n(X_i) - u_n(X_j), \widehat{U}_h(X_i) - \widehat{U}_h(X_j)), \tag{5.11}$$

where $\psi(x, y) = 1$ if $xy > 0$ and $\psi(x, y) = 0$ otherwise. In other words, for each pair (X_i, X_j) we add 1 to the accuracy score if and only if $u_n(X_i) - u_n(X_j)$ and



(a) Mixture of Gaussians (b) f_3

FIG. 5.3. *Comparison of accuracy versus number of samples for various grid sizes and number of subsamples k used to estimate f.*

$\widehat{U}_h(X_i) - \widehat{U}_h(X_j)$ have the same sign, and are both non-zero, i.e., the relative ordering between X_i and X_j given by PDE-based ranking is correct. The accuracy score is then normalized by the number of pairs, which is $\binom{n}{2} = n(n-1)/2$. It turns out that the accuracy scores 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.

The complexity of computing the accuracy score via (5.11) is $O(n^2)$, which is intractable for even moderate values of n. We note however that (5.11) is, at least formally, a Monte-Carlo approximation of

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \psi(U(x) - U(y), U_h(x) - U_h(y)) f(x) f(y)\, dxdy.$$

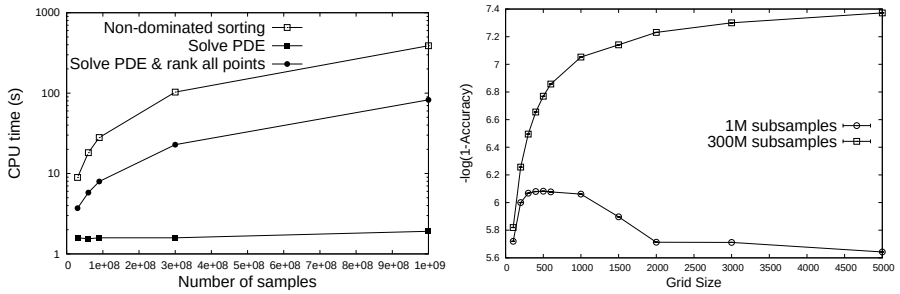
Hence it is natural to use a truncated Monte-Carlo approximation to estimate (5.11). This is done by selecting n pairs $(X_{i_1}, X_{j_1}), \ldots, (X_{i_n}, X_{j_n})$ at random and computing

$$\frac{1}{n} \sum_{k=1}^{n} \psi(u_n(X_{i_k}) - u_n(X_{j_k}), \widehat{U}_h(X_{i_k}) - \widehat{U}_h(X_{j_k})).$$

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 small enough so that they are contained within the data point itself.

Figure 5.3 shows the sorting accuracy versus the number n of points to sort for various grid sizes and number k of subsamples. We see in Figure 5.3 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 compared the performance of our algorithm against the fast two dimensional non-dominated sorting algorithm presented in [13], 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 5.4(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 PDE-based ranking) separately from the CPU time required to execute *all* of PDE-based ranking,



(a) CPU time (b) Accuracy vs grid size

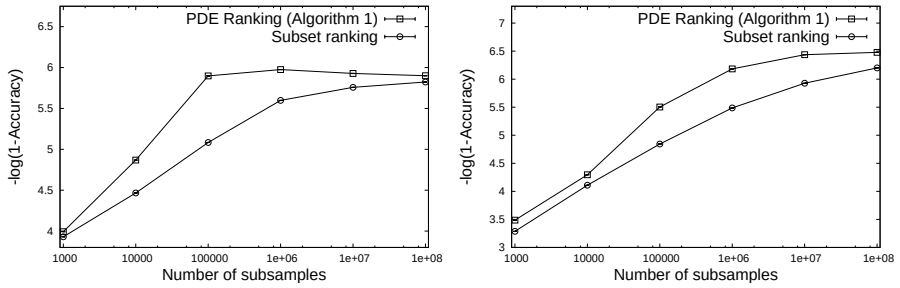
FIG. 5.4. *(a) Comparison of CPU time versus number of samples for a grid size of* 250×250 *and* $k = 10^7$ *subsamples for estimating the density. (b) Comparison of accuracy versus grid size for* $k = 10^6$ *and* $k = 3 \times 10^8$ *subsamples for non-dominated sorting of* $n = 3 \times 10^8$ *points. 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.*

since the former is sublinear in n. We see in Figure 5.4(a) that using PDE-based ranking to rank all n data points is approximately 8 to 10 times faster than non-dominated sorting. While this is a substantial improvement, the more important observation from Figure 5.4(a) is that estimating the density and accurately solving the PDE to compute the Pareto ranking function $\widehat{U}_h$ has roughly *constant* complexity in n. Many applications, such as the anomaly detection problem discussed in Section 5.4, only require estimating this ranking function, and in these cases PDE-based ranking provides a sublinear, in fact constant time, algorithm.

It is interesting to consider more closely the relationship between the grid size and the number of subsamples k. In Figure 5.4(b), we show 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 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 PDE-based ranking (5.7). Intuitively, when $k = 10^6$ and the grid size is greater 1000, there are more grid cells, or histogram bins, than the number k of points used for estimating the density. Hence it is very likely that each bin contains at most 1 sample, and many bins contain no samples. The error from the corresponding histogram estimation is therefore so large that it overwhelms any improvement one would expect to see from solving the PDE on a finer grid.

5.3. Comparison to a heuristic algorithm. 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.

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 1), we use grid interpolation, using the same grid



(a) Mixture of Gaussians (b) f_3

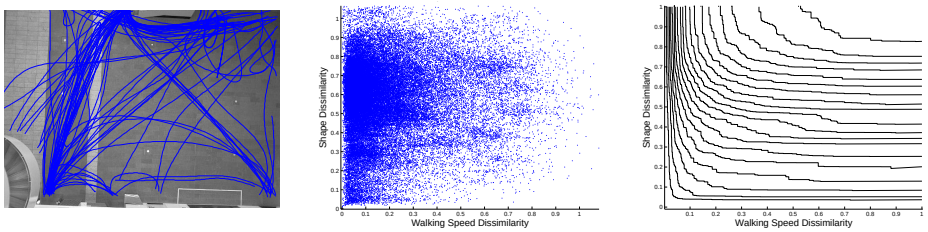
FIG. 5.5. *Comparison of PDE-based ranking (Algorithm 1) and subset interpolation ranking for sorting $n = 10^8$ samples. Accuracy scores are shown for various numbers of subsamples ranging from $k = 10^3$ to $k = 10^8$.*

size as used to solve the PDE. For grid cells $\alpha \in \mathbb{N}^d$ that contain at least one sample from $Y_1, \ldots, Y_k$, we assign a rank r_α by averaging all samples that fall in that cell. For grid cells that contain no samples from $Y_1, \ldots, Y_k$, we define the rank recursively by $r_\alpha = \max_i\{r_{\alpha - e_i}\}$, and we set $r_\alpha = 0$ for $\alpha \notin \mathbb{N}^d$. This can be computed by sweeping through the grid once in any direction that respects the partial order $\leqq$. 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 1).

Figure 5.5 shows the accuracy scores for PDE-based ranking (Algorithm 1) and subset ranking of $n = 10^8$ samples drawn from the mixture of Gaussians distribution and f_3. 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.

5.4. Application to anomaly detection. We now demonstrate PDE-based ranking on a large scale real data application of anomaly detection [18]. The data consists of thousands of pedestrian trajectories, captured from an overhead camera, and the goal is to differentiate nominal from 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 [24]. Figure 5.6(a) shows 100 of the over 100,000 trajectories captured from the overhead camera.

The approach to anomaly detection employed in [18] 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 [18] are a walking speed dissimilarity, and a trajectory shape dissimilarity. 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



(a) Example trajectories (b) 50000 Pareto points (c) Pareto fronts

FIG. 5.6. *(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).*

$X_{i,j} = (c_{speed}(\gamma_i, \gamma_j), c_{shape}(\gamma_i, \gamma_j))$ for every pair of trajectories (γ_i, γ_j), yielding $\binom{M}{2} \approx 6 \times 10^9$ Pareto points. Figure 5.6(b) shows an example of 50000 Pareto points and Figure 5.6(c) shows the respective Pareto fronts. In [18], 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 [18] performs non-dominated sorting on the Pareto points $\{X_{i,j}\}_{1 \leq i < j \leq M}$, 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 PDE-based ranking, we can approximate u_n using only a small fraction of the Pareto points $\{X_{i,j}\}_{1 \leq 1 < j \leq M}$, thus alleviating the computational burden of computing all pairwise dissimilarities. Figure 5.7 shows the accuracy scores for PDE-based ranking 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 PDE-based ranking, it is important to have a ground truth sorting to compare against. 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 5.3. 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.

5.5. Discussion. We have provided theory demonstrating that for $X_1, \ldots, X_n$ *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

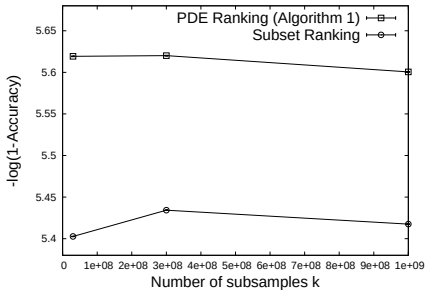


FIG. 5.7. *Accuracy scores for PDE-based ranking and subset ranking for sorting* 10^9 *Pareto points from the pedestrian anomaly detection problem versus the number of subsamples* k.

could explore other numerical techniques for solving (P) which do not utilize a fixed grid [6]. At present, there is also no good algorithm for non-dominated sorting in high dimensions. The fastest known algorithm is $O(n(\log n)^{d-1})$ [19], 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 [16], where it would be important to be able to construct an approximation of the Pareto depth function u_n without visiting all the datapoints $X_1, \ldots, X_n$. For example, the data may be arriving piece by piece, and it may be impossible to keep a history of all samples. In such a setting, one could slightly modify PDE-based ranking 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 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 5.4, 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. For instance, one would expect to obtain better results by using more sophisticated density 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.

Acknowledgments. We thank Ko-Jen Hsiao for providing code for manipulating the pedestrian trajectory database.

Appendix A. 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

$$\forall x, y \in X,\ |x-y| < \delta \implies |f_n(x) - f_n(y)| < \varepsilon + \frac{1}{n}, \tag{A.1}$$

for every $n \in \mathbb{N}$.

THEOREM A.1. *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

$$|f_n(x) - f_n(y)| < \frac{\varepsilon}{4} + \frac{1}{n} \text{ for all } x, y \in X \text{ with } |x-y| < \delta. \tag{A.2}$$

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 (A.2) 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. □

REFERENCES

[1] G. BARLES AND PANAGIOTIS E. SOUGANIDIS, *Convergence of approximation schemes for fully nonlinear second order equations*, Asymptotic Analysis, 4 (1991), pp. 271–283.

[2] B. BOLLOBÁS AND G. BRIGHTWELL, *The height of a random partial order: concentration of measure*, The Annals of Applied Probability, 2 (1992), pp. 1009–1018.

[3] B. BOLLOBÁS AND P. WINKLER, *The longest chain among random points in Euclidean space*, Proceedings of the American Mathematical Society, 103 (1988), pp. 347–353.

[4] JEFF CALDER, *Directed last passage percolation with discontinuous weights*, arXiv preprint:1402.3381, (2014).

[5] J. CALDER, S. ESEDOĞLU, AND A.O. HERO, *A Hamilton-Jacobi equation for the continuum limit of non-dominated sorting*, SIAM Journal on Mathematical Analysis, 46 (2014), pp. 603–638.

[6] T. CECIL, J. QIAN, AND S. OSHER, *Numerical methods for high dimensional Hamilton–Jacobi equations using radial basis functions*, Journal of Computational Physics, 196 (2004), pp. 327–347.

[7] M.G. Crandall, H. Ishii, and P.L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bulletin of the American Mathematical Society, 27 (1992), pp. 1–67.
[8] K. Deb, *Multi-objective optimization using evolutionary algorithms*, Wiley, Chichester, UK, 2001.
[9] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, *A fast and elitist multiobjective genetic algorithm: NSGA-II*, IEEE Transactions on Evolutionary Computation, 6 (2002), pp. 182–197.
[10] K. Deckelnick and C.M. Elliott, *Uniqueness and error analysis for Hamilton-Jacobi equations with discontinuities*, Interfaces and Free Boundaries, 6 (2004), pp. 329–349.
[11] J.-D. Deuschel and O. Zeitouni, *Limiting curves for i.i.d. records*, The Annals of Probability, 23 (1995), pp. 852–878.
[12] M. Ehrgott, *Multicriteria Optimization (2. ed.)*, Springer, 2005.
[13] S. Felsner and L. Wernisch, *Maximum k-chains in planar point sets: Combinatorial structure and algorithms*, SIAM Journal on Computing, 28 (1999), pp. 192–209.
[14] C.M. Fonseca and P.J. Fleming, *Genetic algorithms for multiobjective optimization : formulation, discussion and generalization*, Proceedings of the Fifth International Conference on Genetic Algorithms, 1 (1993), pp. 416–423.
[15] ———, *An overview of evolutionary algorithms in multiobjective optimization*, Evolutionary Computation, 3 (1995), pp. 1–16.
[16] A.C. Gilbert and M.J. Strauss, *Analysis of data streams: Computational and algorithmic challenges*, Technometrics, 49 (2007), pp. 346–356.
[17] J.M. Hammersley, *A few seedlings of research*, in Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, 1972, pp. 345–394.
[18] K.-J. Hsiao, K. Xu, J. Calder, and A.O. Hero, *Multi-criteria anomaly detection using Pareto Depth Analysis*, in Advances in Neural Information Processing Systems 25, NIPS, 2012, pp. 854–862.
[19] M.T. Jensen, *Reducing the run-time complexity of multiobjective EAs: The NSGA-II and other algorithms*, IEEE Transactions on Evolutionary Computation, 7 (2003), pp. 503–515.
[20] A. P. Korostelev and A. B. Tsybakov, *Minimax theory of image reconstruction*, Springer-Verlag, New York, 1993.
[21] D.O. Loftsgaarden and C.P. Quesenberry, *A nonparametric estimate of a multivariate density function*, The Annals of Mathematical Statistics, (1965), pp. 1049–1051.
[22] B. F. Logan and L. A. Shepp, *A variational problem for random Young tableaux*, Advances in Mathematics, 26 (1977), pp. 206–222.
[23] R.D. Lou and M. Sarrafzadeh, *An optimal algorithm for the maximum three-chain problem*, SIAM Journal on Computing, 22 (1993), pp. 976–993.
[24] B. Majecka, *Statistical models of pedestrian behaviour in the forum*, Master's thesis, School of Informatics, University of Edinburgh, (2009).
[25] P. Pevzner, *Computational Molecular Biology*, The MIT Press, 2000.
[26] M. Prähofer and H. Spohn, *Universal distributions for growth processes in 1+ 1 dimensions and random matrices*, Physical Review Letters, 84 (2000), pp. 4882–4885.
[27] N. Srinivas and K. Deb, *Muiltiobjective optimization using nondominated sorting in genetic algorithms*, Evolutionary Computation, 2 (1994), pp. 221–248.
[28] A. Tourin, *A comparison theorem for a piecewise Lipschitz continuous Hamiltonian and application to shape-from-shading problems*, Numerische Mathematik, 62 (1992), pp. 75–85.
[29] A.B. Tsybakov, *Introduction to nonparametric estimation*, Springer, 2009.
[30] S.M. Ulam, *Monte carlo calculations in problems of mathematical physics*, Modern Mathematics for the Engineers, (1961), pp. 261–281.
[31] A.M. Vershik and S.V. Kerov, *Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tables*, Soviet Doklady Mathematics, 18 (1977), p. 38.
[32] G. Viennot, *Chain and antichain families, grids and Young tableaux*, in Orders: Description and Roles, vol. 99 of North-Holland Mathematics Studies, 1984, pp. 409–463.