A continuum limit for non-dominated sorting

Jeff Calder
Department of Mathematics
University of Michigan
Ann Arbor, MI 48109, USA
Email: jcalder@umich.edu

Selim Esedoḡlu
Department of Mathematics
University of Michigan
Ann Arbor, MI 48109, USA
Email: esedoglu@umich.edu

Alfred O. Hero III
Department of Electrical Engineering and
Computer Science, University of Michigan
Ann Arbor, MI 48109, USA
Email: hero@eecs.umich.edu

***Abstract*— Non-dominated sorting is an important combinatorial problem in multi-objective optimization, which is ubiquitous in many fields of science and engineering. In this paper, we overview the results of some recent work by the authors on a continuum limit for non-dominated sorting. In particular, we have discovered that in the (random) large sample size limit, the non-dominated fronts converge almost surely to the level sets of a function that satisfies a Hamilton-Jacobi partial differential equation (PDE). We show how this PDE can be used to design a fast, potentially sublinear, approximate non-dominated sorting algorithm, and we show the results of applying the algorithm to real data from an anomaly detection problem.**

Keywords: Non-dominated sorting, Pareto-optimality, multi-objective optimization, longest chain problem, antichain partition, partial differential equations, Hamilton-Jacobi equations, numerical schemes

I. Introduction

Non-dominated sorting is an important combinatorial problem in multi-objective optimization [1], [2], [3], [4]. The sorting can be viewed as arranging points in Euclidean space into layers, or fronts, by repeated removal of the set of minimal elements with respect to a partial order. It is an essential step in the so-called genetic and evolutionary algorithms for continuous multi-objective optimization, where it is employed to determine fitness levels of feasible solutions in the current population [3], [5], [6], [2], [7]. Of course, multi-objective optimization is ubiquitous in many fields of science and engineering, including control theory and path planning [8], [9], [10], gene selection and ranking [11], [12], [13], [14], [15], [16], [17], data clustering [18], database systems [19], [20] and image processing and computer vision [21], [22].

In probability theory, non-dominated sorting goes by the pseudonym 'The longest chain problem', and has a long history beginning with Ulam's famous problem [23] of finding the length of a longest increasing subsequence of a random permutation. Some of the major breakthroughs on probabilistic aspects of the problem were obtained by Hammersley [24], Vershik and Kerov [25], Logan and Shepp [26], Bollobás and Winkler [27], and Deuschel and Zeitouni [28]. In combinatorics, non-dominated sorting is called the 'canonical antichain partition' [29], and there are further striking applications in molecular biology [30], graph theory [31], Young Tableaux [32], [29], materials science [33], and even in physical layout problems in the design of integrated circuits [34].

In this paper, we overview the results of some recent work by the authors on a continuum limit for non-dominated sorting [35], [36]. In particular, we prove that in the (random) large sample size limit, the non-dominated fronts converge almost surely to the level sets of a function that satisfies a Hamilton-Jacobi partial differential equation (PDE). We also give a fast numerical scheme for solving the PDE and use it to develop a fast, potentially sublinear, approximate non-dominated sorting algorithm. We apply this algorithm to real data from an anomaly detection problem [37], and show that the algorithm achieves excellent accuracy while significantly reducing the computational complexity of non-dominated sorting. We believe this work has the potential to be particularly useful in the context of big data streaming problems [38], which involve constant re-sorting of massive datasets upon the arrival of new samples.

The rest of the paper is organized as follows. In Section II we describe non-dominated sorting and present our main result—a PDE for the limiting shapes of the fronts. In Section III, we present a fast numerical scheme for the PDE and a fast approximate non-dominated sorting algorithm based on this scheme. Finally in Section III-C we apply our fast non-dominated sorting algorithm to real data, and show some experimental results.

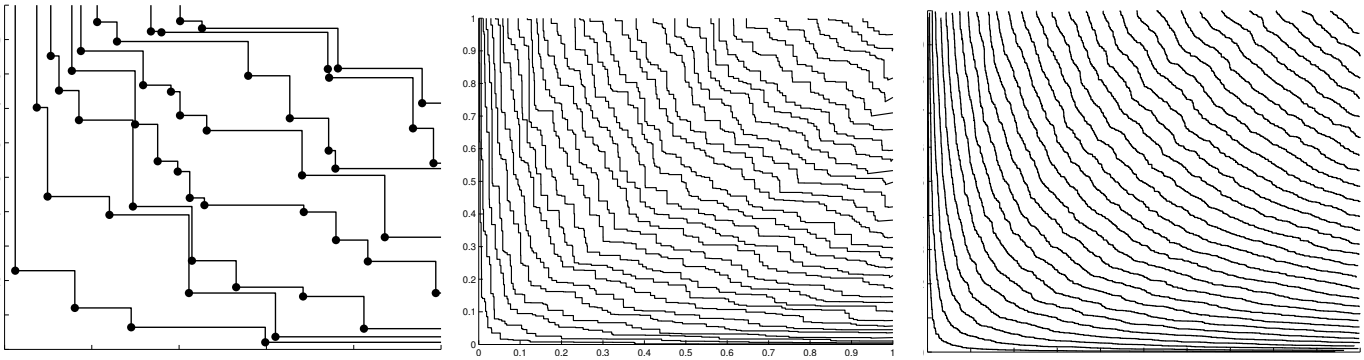
II. Main result

A. Non-dominated sorting

In a discrete multi-objective optimization problem, we are given a number of functions $g_i : S \to [0,\infty)$, where $i = 1,\dots,d$ and $S = \{x^1,\dots,x^n\}$, and the problem is to find an element $x \in S$ that minimizes *all* of the functions $g_1,\dots,g_d$ simultaneously. Since no such solution exists in general, one is instead interested in a family of solutions that are 'optimal' in certain sense. Formally, we say a feasible solution $x \in S$ is *Pareto-optimal* if

$$\forall y \in S, \left\{\exists i,\, g_i(y) > g_i(x) \text{ or } \forall i,\, g_i(y) = g_i(x)\right\}.$$

In other words, no other feasible solution is better in every objective. The set of Pareto-optimal solutions is denoted $\mathscr{F}_1$ and usually called the *first Pareto front*, or *first non-dominated front*. It is a natural notion of solution for a multi-objective optimization problem in which one has no *a priori* information



(a) $n = 50$ points (b) $n = 10^4$ points (c) $n = 10^6$ points

Fig. 1. Examples of exact finite sample Pareto fronts for $X_1, \dots X_n$ chosen from the uniform distribution on $[0,1]^2$. In (b) and (c), 29 equally spaced fronts are depicted.

concerning the relative importance of each objective. The second Pareto front, $\mathscr{F}_2$, consists of the Pareto-optimal elements of $S \setminus \mathscr{F}_1$, and in general

$$\mathscr{F}_k = \text{Pareto optimal elements of } S \setminus \bigcup_{j<k} \mathscr{F}_j. \tag{1}$$

Now set $X_i = (g_1(x^i), \dots, g_d(x^i)) \in \mathbb{R}^d$ for $i = 1, \dots, n$. In this paper, we view non-dominated sorting as acting on $X_1, \dots, X_n$, which are simply points in $\mathbb{R}^d$, so that the problem of non-dominated sorting is removed from the underlying multi-objective optimization problem. Figure 1 shows the Pareto fronts obtained by non-dominated sorting of $n = 50, 10^4$, and $n = 10^6$ points chosen from the uniform distribution on $[0,1]^2$.

Let us now describe the connection to the longest chain problem. Let $X_1, \dots, X_n$ be independent and identically distributed (*i.i.d.*) random variables. The points $\mathscr{X}_n = \{X_1, \dots, X_n\}$ form a (random) partially ordered set under the usually coordinate-wise partial order

$$x \leqq y \iff x_i \leq y_i \quad \text{for} \quad i = 1, \dots, d. \tag{2}$$

Let $\ell(n)$ denote the length of a longest chain[1] in $\mathscr{X}_n$, and for $x \in \mathbb{R}^d$ let $u_n(x)$ denote the length of a longest chain in $\mathscr{X}_n$ consisting of points less than x with respect to $\leqq$. When $X_i = (g_1(x^i), \dots, g_d(x^i)) \in \mathbb{R}^d$ for $i = 1, \dots, n$, we have

$$x^i \in \mathscr{F}_1 \iff u_n(X_i) = 1,$$

and in general

$$x^i \in \mathscr{F}_k \iff u_n(X_i) = k,$$

provided all X_i are distinct. This observation is critical; it says that studying the shapes of the Pareto fronts $\mathscr{F}_1, \mathscr{F}_2, \dots$ is equivalent to studying the level sets of the longest chain function u_n. Notice in Figure 1, that the points on each Pareto front are connected by a staircase curve that represents the jump set of u_n.

The problem of studying the asymptotics of $\ell(n)$ can be traced back to Ulam's problem [23], which was first tackled by Hammersley [24]. He showed that for $X_1, \dots, X_n$ independent and uniformly distributed on $[0,1]^2$ we have $\ell(n) \sim c\sqrt{n}$ almost surely, and he conjectured that $c = 2$. This conjecture was later verified by Vershik and Kerov [25] and Logan and Shepp [26]. Bollobás and Winkler [27] extended Hammersley's results to dimensions $d \geq 3$, showing that there exist positive constants c_d such that $\ell(n) \sim c_d n^{\frac{1}{d}}$ almost surely, and $c_d \nearrow e$ as $d \to \infty$. Deuschel and Zeitouni [28] considered non-uniformly distributed points. They showed that 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, we have $\ell(n) \sim 2\overline{J}\sqrt{n}$ 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.

Our work is most closely related to [28]. In particular, we extend their results to higher dimensions and to densities f on arbitrary domains, and more importantly, we connect the variational problem to a Hamilton-Jacobi partial differential equation, which allows us to design a fast non-dominated sorting algorithm.

B. Continuum limit

Our main result is the following continuum limit:

Theorem 1: Let $X_1, \dots, X_n$ be *i.i.d.* random variables on $\mathbb{R}^d$ with density function $f : \mathbb{R}^d \to [0, \infty)$. Suppose there exist $\Omega \subset \mathbb{R}^d_+$ open and bounded, with Lipschitz boundary, such that f is continuous on $\overline{\Omega}$ and $f = 0$ on $\mathbb{R}^d \setminus \overline{\Omega}$. Then there exists a positive constant c_d such that

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

where $U \in C^{0,\frac{1}{d}}([0,\infty)^d)$ is the unique Pareto-monotone viscosity solution of the Hamilton-Jacobi equation

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

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

Here, $\mathbb{R}_+ = (0,\infty)$, U_{x_i} refers to the partial derivative of U in the i^{th} coordinate direction, and by Pareto-monotone, we mean that $x \leqq y \implies U(x) \leq U(y)$. The set Ω is the domain of the random variables $X_1,\dots,X_n$, and the constants c_d are the same as those given by Bollobás and Winkler [27]. In particular, $c_1 = 1$, $c_2 = 2$ and $c_d \nearrow e$ as $d \to \infty$. The proof of Theorem 1 will appear in the SIAM Journal on Mathematical Analysis in 2014 [35].

Let us say a few words about the notion of viscosity solutions of Hamilton-Jacobi equations. In general, there does not exist a continuously differentiable function U satisfying a Hamilton-Jacobi equation like (P), due to the possibility of crossing characteristics. On the other hand, it is often the case that there exist infinitely many almost everywhere differentiable functions U satisfying (P) at every point of differentiability. These solutions are called *almost everywhere solutions*. The notion of viscosity solution selects the "physically correct" almost everywhere solution. For more details on viscosity solutions, we refer the reader to the standard references [39], [40]. We should note that when f is a product density, i.e., $f(x) = f_1(x_1)\cdots f_d(x_d)$, U has a familiar form; it is given by the d^{th}-root of the cumulative distribution function

$$U(x) = F(x)^{\frac{1}{d}} = \left(\int_{0\leqq y\leqq x} f(y)\,dy\right)^{\frac{1}{d}}. \tag{3}$$

Theorem 1 provides a new tool for studying the asymptotic properties of non-dominated sorting. For example, in [35], we used Theorem 1 to show that non-dominated sorting is asymptotically stable under bounded random perturbations. Furthermore, convexity (or lack thereof) is a crucial property of Pareto fronts [1]. When the Pareto fronts are non-convex, linear scalarization—a popular approach to multi-objective optimization based on minimizing a linear combination of the objectives—will only find Pareto-optimal solutions on the convex hull of the Pareto front, and will neglect equally good solutions on non-convex portions. Since the Pareto fronts converge to the level sets of U, the asymptotic convexity of the Pareto fronts is related to the quasiconcavity of U; thus an interesting problem is to characterize the class of density functions f for which the solution U of (P) is quasiconcave.

III. Numerics

Evidently, Theorem 1 reduces the problem of non-dominated sorting (in the asymptotic regime) to solving a Hamilton-Jacobi equation. We now show how to exploit this to design a fast approximate non-dominated sorting algorithm.

A. Numerical scheme

When f is a product density, U can be computed efficiently by numerically integrating (3). Our main result here is that for general densities, f, there is a similarly efficient numerical scheme for computing U. The numerical scheme is as follows:

$$\text{(S)}\quad \begin{cases} \displaystyle\prod_{i=1}^{d} h^{-1}(U_h(x) - U_h(x-he_i)) = f(x), & x \in h\mathbb{N}^d \\ U_h(x) = 0, & x \in h\mathbb{N}_0^d \setminus h\mathbb{N}^d. \end{cases}$$

Here, $U_h : h\mathbb{N}_0^d \to \mathbb{R}$ is the numerical solution on a grid of spacing $h > 0$, $e_1,\dots,e_d$ are the standard basis vectors in $\mathbb{R}^d$, and $\mathbb{N}_0 = \{0,1,2,\dots\}$. The choice of backward difference quotients gives what is called an 'upwind scheme' due to the fact that information propagates along coordinate axes in the definition of Pareto fronts (1). There numerical solution U_h satisfying (S) is computed recursively in the following way: Given $U_h(x-he_1),\dots,U_h(x-he_d)$, we compute $U_h(x)$ by solving the algebraic equation given in (S). There are in general d solutions, for given values of $U_h(x-he_1),\dots,U_h(x-he_d)$. We obtain the Pareto-monotone viscosity solution of (P) by selecting the unique $U_h(x)$ satisfying

$$U_h(x) \geq \max\Big(U_h(x-he_1),\dots,U_h(x-he_d)\Big).$$

The algebraic equation in (S) can be solved in general by an efficient binary search. In dimension $d = 2$, the equation is quadratic, and we can explicitly write

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

Computing U_h involves visiting each grid point exactly once in any sweeping pattern that respects the componentwise partial order $\leqq$, and therefore has linear complexity.

Our main numerical result is the following

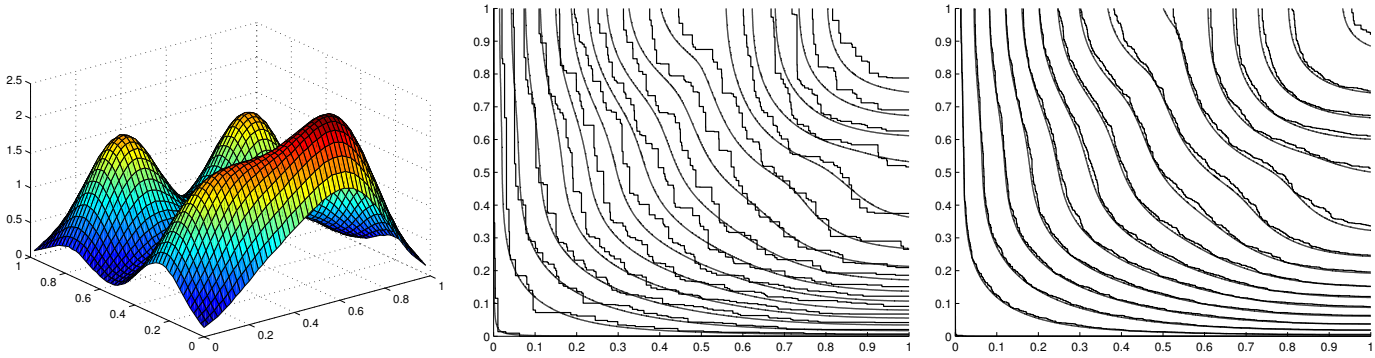
> *Theorem 2:* Assume f satisfies the hypotheses from Theorem 1. Then $U_h \to U$ uniformly on $[0,\infty)^2$ as $h \to 0$.

This result guarantees that the numerical solutions U_h are arbitrarily good approximations to the viscosity solution of (P) for $h > 0$ sufficiently small. The theorem does not guarantee any rate of convergence. In [36], we give some numerical evidence indicating that $U_h = U + O(h^{\frac{1}{d}})$. The proof of Theorem 2 can be found in a recent preprint by the authors [36]. Figure 2 compares the Pareto fronts to the level sets of the numerical solution U_h of (P) computed via the scheme (S). Here, we chose f to be the multi-modal density depicted in Figure 2(a) and solved (S) on a 1000×1000 grid ($h = 0.001$).

To give an idea of the computational complexity, it takes approximately one quarter of a second to solve (S) on a 1000×1000 grid using an average laptop. In practice, such a fine grid is unnecessary, and we have found that a grid on the order of 100×100 ($h = 0.01$) to be sufficient. Of course, the complexity of solving (S) on a fixed grid grows exponentially fast in dimension d, hence the scheme (S) is only applicable in relatively small dimensions, i.e., $d = 2,3,4$. It is thus a very interesting problem to numerically solve (P) in higher dimensions, and we leave this to future work.

B. Fast approximate sorting

If the data $X_1,\dots,X_n$ are drawn *i.i.d.* from a smooth density function f, and n is large enough so that $n^{-\frac{1}{d}}u_n$ is well approximated by $c_d d^{-1} U$, then it is reasonable to consider the following approximate non-dominated sorting algorithm: 1)



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

Fig. 2. Comparison of the exact Pareto fronts and the continuum approximation via the level sets of U_h—the numerical solution of (P)—for the density f depicted in (a). In each case, we show 15 equally spaced Pareto fronts and the corresponding level sets of U_h. We used a 1000×1000 grid for solving the PDE, which corresponds to $h = 0.001$.

Since the density f is rarely known in practice, we first form an estimate $\widehat{f}$ of f using a small (random) subset of the samples $X_1, \dots, X_n$. We opt for a simple histogram estimator, but there are of course other more accurate options available [41], [42]. 2) Use the numerical scheme (S) and the estimate $\widehat{f}$ to solve (P) on a fixed grid of size h. This yields a numerical solution $\widehat{U}_h$. (3) Evaluate $\widehat{U}_h$ at each sample $X_1, \dots, X_n$ to yield approximate Pareto ranks. In summary we have

Algorithm 1: Fast approximate non-dominated sorting

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

1b) 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., for $x \in h\mathbb{N}_0^d$ we have

$$\widehat{f}_h(x) = \frac{1}{kh^d} \cdot \#\Big\{ Y_i : x \leqq Y_i \leqq x + h(1, \dots, 1) \Big\}. \quad (4)$$

2) Compute $\widehat{U}_h$ on $h\mathbb{N}_0^d \cap [0,1]^d$ via (S).

3) Evaluate $\widehat{U}_h(X_i)$ for $i = 1, \dots, n$ via interpolation.

The final evaluation step can be viewed as an interpolation, and the specific form of interpolation is not all that important; d-linear interpolation is sufficient, and in [36] we use a marginally more accurate interpolation algorithm based adapting the scheme (S) to subgrid resolution. For simplicity of presentation, we assumed that $X_1, \dots, X_n \in [0,1]^d$, but the scheme can be easily adapted to any compact hypercube.

It is natural to wonder if one can obtain a rate of convergence for $\widehat{U}_h \to U$. In [36], we show formally that the following estimate should hold with high probability:

$$\|\widehat{U}_h - U\|_{L^\infty([0,1]^d)} \leq C \left(k^{-\frac{1}{2d}} h^{-1} + h^{\frac{1}{d}} \right). \quad (5)$$

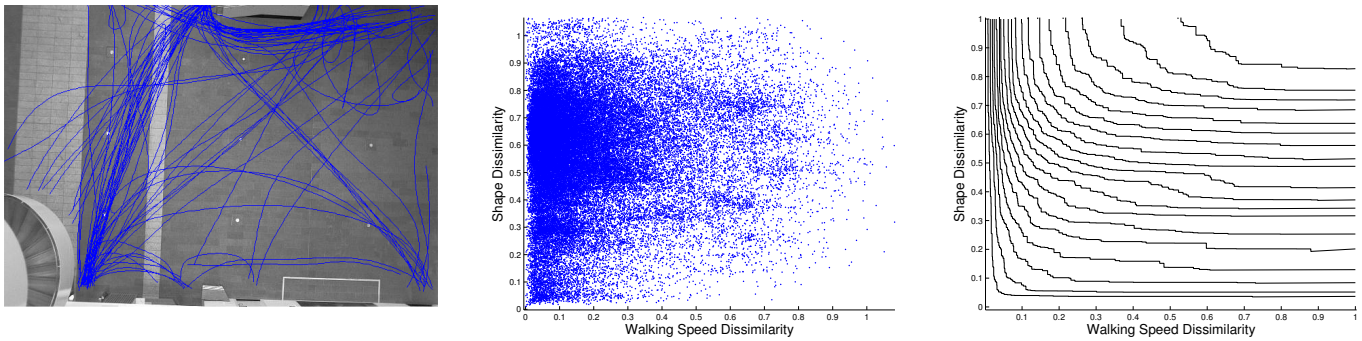
Here, $C > 0$ is a constant independent of h and k. We intend to prove (5) rigorously in a future work. The first term on the right hand side in (5) arises from the effects of random errors (variance) due to an insufficient number of samples k, whereas the second term is due to the effect of non-random errors (bias) due to insufficient grid resolution h. This decomposition is in some ways analogous to the mean integrated squared error decomposition in the theory of non-parametric regression and image reconstruction [43]. The estimate (5) is of course useful in choosing appropriate values of k and h in Algorithm 1.

Notice that Steps 1) and 2) in Algorithm 1 require only $O(k + h^{-d})$ operations. Therefore, Algorithm 1 can be considered sublinear in the following sense: For any $\varepsilon > 0$, by choosing k and h appropriately, and fixing them independent of n, we can obtain an algorithm that requires $O(1)$ operations, as $n \to \infty$, to compute an estimate $\widehat{U}_h$ of U satisfying $\|\widehat{U}_h - U\|_{L^\infty([0,1]^d)} \leq \varepsilon$ with high probability. Of course, evaluating $\widehat{U}_h$ at each sample $X_1, \dots, X_n$ requires $O(n)$ operations.

C. Anomaly detection

We now demonstrate Algorithm 1 on a large-scale real-world dataset from an anomaly detection problem [37]. 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 [44]. Figure 3(a) shows 100 of the over 100,000 trajectories captured from the overhead camera.

The approach to anomaly detection employed in [37] utilizes multiple criteria to measure the dissimilarity between trajectories, and combines the information using a Pareto-front method, i.e., non-dominated sorting. The database consists of a collection of $M = 110{,}035$ trajectories, and the criteria are 1) a walking speed dissimilarity, and 2) a trajectory shape dissimilarity. The walking speed dissimilarity is the L^2 distance between the velocity histograms of two trajectories, and the shape dissimilarity is the L^2 distance between the trajectories, assuming a uniform walking speed. There is then a Pareto point $X_{i,j} \in \mathbb{R}^2$ for every pair of trajectories, yielding $\binom{M}{2} \approx 6 \times 10^9$ Pareto points. Figure 3(b) shows an example of 50,000 Pareto points and Figure 3(c) shows the respective



(a) Example trajectories (b) 50,000 Pareto points (c) Pareto fronts

Fig. 3. (a) Example pedestrian trajectories, (b) Plot of 50,000 of the approximately 6×10^9 Pareto points, (c) Depiction of 30 evenly spaced Pareto fronts associated to the 50,000 points in (b).

Pareto fronts. In [37], only 1,666 trajectories from one day were used, due to the computational complexity of computing the dissimilarities and non-dominated sorting.

The anomaly detection algorithm from [37] performs non-dominated sorting on the Pareto points $\{X_{i,j}\}_{1\le i<j\le M}$, and uses this sorting to define an anomaly score for every trajectory. Trajectories with anomaly scores higher than a specific threshold are deemed anomalous. Using Algorithm 1, we can approximate the non-dominated sorting of all $\binom{M}{2}$ Pareto points using only a small subset of size k. This allows us to efficiently train the algorithm with all of the training data, instead of just one day.

In practice, the numerical ranks assigned to each point are largely irrelevant, provided the relative orderings between samples are correct. Thus, to evaluate the accuracy of Algorithm 1 for the anomaly detection problem, we define the following accuracy score:

$$\text{Accuracy} = \text{Fraction of pairs } (X_i, X_j) \text{ correctly ordered.}$$

Figure 4 shows the accuracy scores for Algorithm 1 versus the number of subsamples k, and the CPU time used by Algorithm 1 and non-dominated sorting. Notice that we plotted $-\log(1-\text{Accuracy})$, since the values are very close to one. Due to the memory requirements for non-dominated sorting, we cannot sort datasets significantly larger than than 10^9 points. In order to have ground truth to compare against, we used only 44722 out of 110035 trajectories, yielding approximately 10^9 Pareto points. Note that a 500×500 grid was used for solving the PDE, and we show the CPU time for steps 1) and 2) (Solve PDE) separate from the time to execute all of Algorithm 1 to illustrate the sublinearity of this portion of the algorithm. We compared against the $O(n \log n)$ non-dominated sorting algorithm given in [29], [4].

We also compare Algorithm 1 against a naïve algorithm for fast non-dominated sorting. We call the algorithm *subset ranking* [36], and the idea is to randomly sample k points from $X_1, \ldots, X_n$, sort this small subset, and then extrapolate the Pareto ranks to the larger dataset $X_1, \ldots, X_n$. Subset ranking is fast—comparable to Algorithm 1—but there is no reason to expect it to be accurate. In Figure 4, we show the accuracy of subset ranking. Although it is not as accurate as Algorithm 1, the strong performance of subset ranking is quite surprising, and there is, to our knowledge, no rigorous justification for this. We note also that $\{X_{i,j}\}_{1\le i<j\le M}$ are *not i.i.d.*, since they are elements of a Euclidean dissimilarity matrix, so these results are also strong evidence that Theorem 1 holds for some special cases of non-*i.i.d.* random variables.

IV. Conclusion

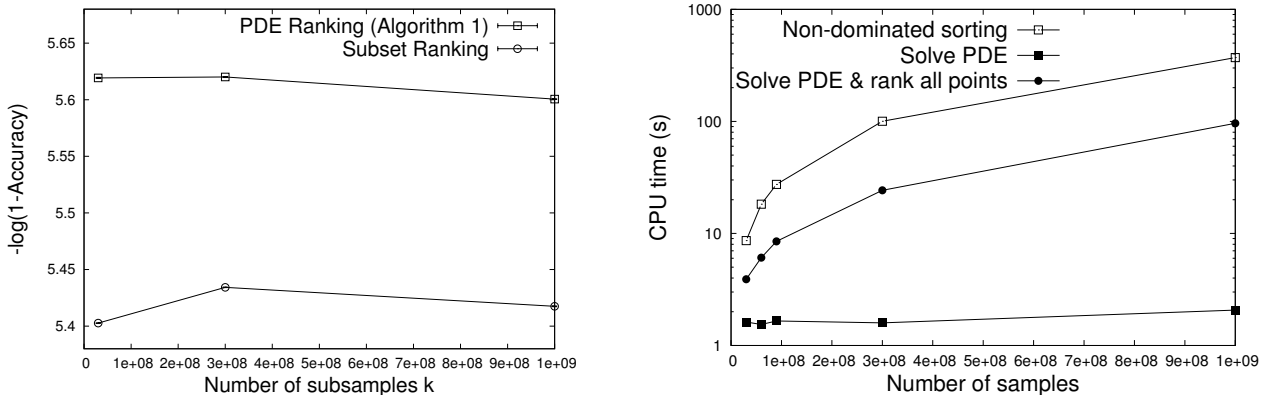
In conclusion, we have presented an overview of our recent work on a continuum limit for non-dominated sorting [35], [36]. We identified a Hamilton-Jacobi partial differential equation (PDE) for this continuum limit, and showed how to numerically solve the PDE efficiently. We presented a fast approximate non-dominated sorting algorithm based on numerically solving this PDE, and applied the algorithm to real-world data from an anomaly detection problem with favorable results. We expect this work to be useful in other multi-objective optimization problems as well; in particular, it seems very well-suited for big data problems in the streaming context [38].

Acknowledgment

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

References

[1] M. Ehrgott, *Multicriteria Optimization (2. ed.)*. Springer, 2005.
[2] K. Deb, *Multi-objective optimization using evolutionary algorithms*. Wiley, Chichester, UK, 2001.
[3] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.
[4] M. Jensen, "Reducing the run-time complexity of multiobjective EAs: The NSGA-II and other algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 5, pp. 503–515, 2003.
[5] C. Fonseca and P. Fleming, "Genetic algorithms for multiobjective optimization : formulation, discussion and generalization," *Proceedings of the Fifth International Conference on Genetic Algorithms*, vol. 1, pp. 416–423, Jul. 1993.



(a) Accuracy

(b) CPU time

Fig. 4. Accuracy scores for Algorithm 1 and subset ranking for sorting 10^9 Pareto points from the pedestrian anomaly detection problem versus the number of subsamples k. The proposed continuum approximation (PDE ranking via Algorithm 1) has significantly lower computational complexity than non-dominated sorting, and has higher accuracy compared to the similarly fast subset ranking algorithm presented in Section III-C. In b) we used $k = 10^7$ subsamples in Algorithm 1, and a 250×250 grid for solving the PDE.

[6] ——, "An overview of evolutionary algorithms in multiobjective optimization," *Evolutionary Computation*, vol. 3, no. 1, pp. 1–16, 1995.

[7] N. Srinivas and K. Deb, "Muiltiobjective optimization using nondominated sorting in genetic algorithms," *Evolutionary Computation*, vol. 2, no. 3, pp. 221–248, 1994.

[8] I. Mitchell and S. Sastry, "Continuous path planning with multiple constraints," in *IEEE Conference on Decision and Control*, vol. 5, Dec. 2003, pp. 5502–5507.

[9] A. Kumar and A. Vladimirsky, "An efficient method for multiobjective optimal control and optimal control subject to integral constraints," *Journal of Computational Mathematics*, vol. 28, no. 4, pp. 517–551, 2010.

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

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

[12] 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.

[13] A. Hero, "Gene selection and ranking with microarray data," in *IEEE International Symposium on Signal Processing and its Applications*, vol. 1, Jul. 2003, pp. 457–464.

[14] A. Hero and G. Fleury, "Pareto-optimal methods for gene ranking." *VLSI Signal Processing*, vol. 38, no. 3, pp. 259–275, 2004.

[15] 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.

[16] G. Fleury, A. Hero, S. Zareparsi, and A. Swaroop, "Gene discovery using Pareto depth sampling distributions," *Journal of the Franklin Institute*, vol. 341, no. 1–2, pp. 55–75, 2004.

[17] ——, "Pareto depth sampling distributions for gene ranking," in *IEEE International Symposium on Biomedical Imaging: Nano to Macro* , vol. 1, Apr. 2004, pp. 65–68.

[18] J. Handl and J. Knowles, "Exploiting the trade-off – The benefits of multiple objectives in data clustering," in *Evolutionary Multi-Criterion Optimization*. Springer, 2005, pp. 547–560.

[19] 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*, 2002, pp. 275–286.

[20] D. Papadias, Y. Tao, G. Fu, and B. Seeger, "Progressive skyline computation in database systems," *ACM Transactions on Database Systems (TODS)*, vol. 30, no. 1, pp. 41–82, 2005.

[21] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Communications on Pure and Applied Mathematics*, vol. 42, no. 5, pp. 577–685, 1989.

[22] T. Chan and L. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, Feb. 2001.

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

[24] J. Hammersley, "A few seedlings of research," in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1972, pp. 345–394.

[25] A. Vershik and S. Kerov, "Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tables," *Soviet Doklady Mathematics*, vol. 18, no. 527-531, p. 38, 1977.

[26] B. F. Logan and L. A. Shepp, "A variational problem for random Young tableaux," *Advances in Mathematics*, vol. 26, no. 2, pp. 206–222, 1977.

[27] B. Bollobás and P. Winkler, "The longest chain among random points in Euclidean space," *Proceedings of the American Mathematical Society*, vol. 103, no. 2, pp. 347–353, June 1988.

[28] J.-D. Deuschel and O. Zeitouni, "Limiting curves for i.i.d. records," *The Annals of Probability*, vol. 23, no. 2, pp. 852–878, 1995.

[29] S. Felsner and L. Wernisch, "Maximum k-chains in planar point sets: Combinatorial structure and algorithms," *SIAM Journal on Computing*, vol. 28, no. 1, pp. 192–209, 1999.

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

[31] R. Lou and M. Sarrafzadeh, "An optimal algorithm for the maximum three-chain problem," *SIAM Journal on Computing*, vol. 22, no. 5, pp. 976–993, 1993.

[32] G. Viennot, "Chain and antichain families, grids and Young tableaux," in *Orders: description and roles*, ser. North-Holland Mathematics Studies, 1984, vol. 99, pp. 409–463.

[33] M. Prähofer and H. Spohn, "Universal distributions for growth processes in 1+ 1 dimensions and random matrices," *Physical Review Letters*, vol. 84, no. 21, pp. 4882–4885, 2000.

[34] G. Adhar, "Parallel algorithms for chains and anti-chains of points on a plane," in *International Conference on Parallel and Distributed Systems (ICPADS)*, vol. 2, Dec. 2007, pp. 1–7.

[35] J. Calder, S. Esedoglu, and A. O. Hero, "A Hamilton-Jacobi equation for the continuum limit of non-dominated sorting," *To appear in the SIAM Journal on Mathematical Analysis*, 2014.

[36] J. Calder, S. Esedoğlu, and A. O. Hero, "A PDE-based approach to non-dominated sorting," *Submitted to SIAM Journal on Numerical Analysis*, 2013.

[37] K.-J. Hsiao, K. Xu, J. Calder, and A. Hero, "Multi-criteria anomaly detection using Pareto Depth Analysis," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 854–862. [Online]. Available: http://books.nips.cc/papers/files/nips25/NIPS2012_0395.pdf

[38] A. Gilbert and M. Strauss, "Analysis of data streams: Computational and algorithmic challenges," *Technometrics*, vol. 49, no. 3, pp. 346–356, 2007.

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

[40] 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*, vol. 27, no. 1, pp. 1–67, Jul. 1992.
[41] A. Tsybakov, *Introduction to nonparametric estimation*. Springer, 2009.
[42] D. Loftsgaarden and C. Quesenberry, "A nonparametric estimate of a multivariate density function," *The Annals of Mathematical Statistics*, pp. 1049–1051, 1965.
[43] A. P. Korostelev and A. B. Tsybakov, *Minimax theory of image reconstruction*. Springer-Verlag, New York, 1993.
[44] B. Majecka, "Statistical models of pedestrian behaviour in the forum," *Master's thesis, School of Informatics, University of Edinburgh*, 2009.