Nonlinear Analysis 141 (2016) 88–108

Contents lists available at ScienceDirect

Nonlinear Analysis

www.elsevier.com/locate/na

A direct verification argument for the Hamilton–Jacobi equation continuum limit of nondominated sorting[☆]

Jeff Calder

Department of Mathematics, University of California, Berkeley, United States

ARTICLE INFO

Article history:
Received 1 December 2015
Accepted 30 March 2016
Communicated by Enzo Mitidieri

MSC:
35D40
60F15
60C05

Keywords:
Continuum limit
Hamilton–Jacobi equation
Viscosity solution
Stochastic homogenization
Nondominated sorting
Longest chain problem

ABSTRACT

Nondominated sorting is a combinatorial algorithm that sorts points in Euclidean space into layers according to a partial order. It was recently shown that nondominated sorting of random points has a Hamilton–Jacobi equation continuum limit. The original proof, given in Calder et al. (2014), relies on a continuum variational problem. In this paper, we give a new proof using a direct verification argument that completely avoids the variational interpretation. We believe this may be generalized to apply to other stochastic homogenization problems for which there is no obvious underlying variational principle.

1. Introduction

Many problems in science and engineering require the sorting, or ordering, of large amounts of multivariate data. Since there is no canonical linear criterion for sorting data in dimensions greater than one, many different methods for sorting have been proposed to address various problems (see, e.g., [7,57,46,16]). Many of these algorithms abandon the idea of a linear ordering, and instead sort the data into layers according to some set of criteria.

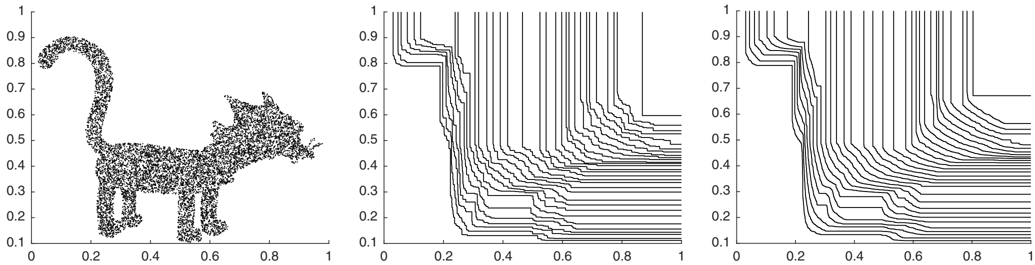
We consider here *nondominated sorting*, which arranges a set of points in Euclidean space into layers by repeatedly removing the set of minimal elements. Let $\leqq$ denote the coordinatewise partial order on $\mathbb{R}^d$ defined by

$$x \leqq y \iff x_i \leq y_i \quad \text{for all } i.$$

☆ The research described in this paper was partially supported by National Science Foundation grant DMS-1500829.

E-mail address: jcalder@berkeley.edu.

http://dx.doi.org/10.1016/j.na.2016.03.023
0362-546X/



(a) *i.i.d.* samples. (b) $n = 10^4$ points. (c) $n = 10^6$ points.

Fig. 1. Examples of Pareto fronts corresponding to *i.i.d.* random variables $X_1, \dots, X_n$ drawn from the distribution depicted in (a). In (b) and (c), we show 30 evenly spaced Pareto fronts for $n = 10^4$ and $n = 10^6$, respectively.

The *first nondominated layer*, also called the *first Pareto front* and denoted $\mathcal{F}_1$, is exactly the set of minimal elements of S with respect to $\leqq$, and the deeper fronts are defined recursively as follows:

$$\mathcal{F}_k = \text{Minimal elements of } S \setminus \bigcup_{i<k} \mathcal{F}_i.$$

This peeling process eventually exhausts the entire set S, and the result is a partition of S based on Pareto front index, which is often called *Pareto depth* or *rank*. Fig. 1 gives an illustration of nondominated sorting of a random set S.

Nondominated sorting is widely used in multi-objective optimization, where it is the basis of the popular and effective genetic and evolutionary algorithms [16,21,22,15,60]. Of course, multi-objective optimization is ubiquitous in engineering and scientific contexts, such as control theory and path planning [49,42,48], gene selection and ranking [59,29,28,30,18–20], data clustering [27], database systems [39,51,31], image processing and computer vision [50,14], and some recent machine learning problems [31–33].

Nondominated sorting is also equivalent to the longest chain problem, which has a long history in probability and combinatorics [63,26,9,17]. A *chain* in $\mathbb{R}^d$ is a finite sequence of points that is totally ordered with respect to $\leqq$. Let $X_1, \dots, X_n$ be n distinct points in $\mathbb{R}^d$ and define

$$U_n(x) = \ell(\{X_1, \dots, X_n\} \cap [0, x]), \tag{1.1}$$

where $\ell(\mathcal{O})$ denotes the length of a longest chain in the finite set $\mathcal{O} \subseteq \mathbb{R}^d$. The notation $[0, x]$ is a special case of the more general interval notation

$$[x, z] = \left\{ y \in \mathbb{R}^d : x \leqq y \leqq z \right\} = \prod_{i=1}^{d} [x_i, z_i]$$

that we shall use throughout the paper. Let $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \dots$ denote the Pareto fronts obtained by applying nondominated sorting to $S := \{X_1, \dots, X_n\}$. Then $x \in \mathcal{F}_1$ if and only if there are no other points $y \in S$ with $y \leqq x$, i.e., $U_n(x) = 1$. A point $x \in S$ is on the second Pareto front $\mathcal{F}_2$ if and only if all points $y \in S$ satisfying $y \leqq x$ are on the first front, and one such point exists. For any such $y \in \mathcal{F}_1$, $\{y, x\}$ is a chain of length $\ell = 2$ in $S \cap [0, x]$, and we see that $x \in \mathcal{F}_2 \iff U_n(x) = 2$. Peeling off successive Pareto fronts and repeating this argument yields

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

Hence, the Pareto fronts $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \dots$ are embedded into the level sets (or jump sets) of the longest chain function U_n, as depicted in Fig. 1.

In [12], we proved the following continuum limit for nondominated sorting.