Journal of Statistical Planning and Inference 160 (2015) 60–74

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

On using the hypervolume indicator to compare Pareto fronts: Applications to multi-criteria optimal experimental design

Yongtao Cao [a], Byran J. Smucker [b], Timothy J. Robinson [c,*]

[a] Department of Mathematics, Indiana University of Pennsylvania, Indiana, PA, USA

[b] Department of Statistics, Miami University, Oxford, OH, USA

[c] Department of Statistics, University of Wyoming, Laramie, WY 82071, USA

ARTICLE INFO

Article history:
Received 7 April 2014
Received in revised form 16 September 2014
Accepted 11 December 2014
Available online 13 January 2015

Keywords:
Pareto front
Multi-objective optimization
Design of experiments
Point exchange

ABSTRACT

The Pareto approach to optimal experimental design simultaneously considers multiple objectives by constructing a set of Pareto optimal designs while explicitly considering tradeoffs between opposing criteria. Various algorithms have been proposed to populate Pareto fronts of designs, and evaluating and comparing these fronts – and by extension the algorithms that produce them – is crucial. In this paper, we first propose a framework for comparing algorithm-generated Pareto fronts based on a refined hypervolume indicator. We then theoretically address how the choice of the reference point affects comparisons of Pareto fronts, and demonstrate that our approach is Pareto compliant. Based on our theoretical investigation, we provide rules for choosing reference points when two-dimensional Pareto fronts are compared. Because theoretical results for three-dimensional fronts are difficult to obtain, we propose an empirical rule for the three-dimensional case by making an analogy to the rules for two dimensions. We also consider the use of our procedure in evaluating the progress of a front-constructing algorithm, and illustrate our work with two examples from the literature.

1. Introduction

Most experiments are conducted with multiple, competing objectives in mind. Therefore, designing under a single criterion may be inadequate. For instance, Gilmour and Trinca (2012) show via examples that the traditional D-optimal designs allow no ability to estimate pure error; on the other hand, the best design for estimating pure error performs very poorly in terms of the D-criterion. In situations like this, the final choice of an experimental design should reflect appropriate compromise across the criteria of interest. But choosing a design based upon the simultaneous optimization of multiple design criteria is often a difficult problem. Without *a priori* knowledge about the interdependencies between the criteria, the conventional compound design and constrained design approach (e.g. Cook and Wong, 1994) for solving multiple-criteria optimal design problems could lead to relatively poor solutions (Coello Coello et al., 2007; Das and Dennis, 1997). Furthermore, the tradeoff between the objectives cannot be fully understood without simultaneously considering all criteria.

The Pareto front approach (Park, 2009; Lu et al., 2011; Sambo et al., 2014) not only accounts for the varying interest and importance of the various objectives simultaneously but also provides the most insight about the tradeoffs between the

* Corresponding author.
E-mail address: tjrobin@uwyo.edu (T.J. Robinson).

http://dx.doi.org/10.1016/j.jspi.2014.12.004

alternative choices, which in turn enables better decision making. This procedure involves finding a set of *Pareto optimal designs* and then using the experimenter's evaluation of the existing trade-offs between the designs to ultimately choose between them. The criterion vectors associated with the Pareto optimal set of designs is known as the *Pareto front*. The shape of the Pareto front provides useful information about the amount of tradeoff between the different criteria and how much compromise is needed from some criteria to improve others. Critical to this approach is the assumption that the Pareto front has been sufficiently populated. The true Pareto front, however, is rarely known and hence any algorithm used to generate a set of designs (e.g. the exchange algorithms of Lu et al., 2011; Sambo et al., 2014, or the multi-objective evolutionary algorithm of Park, 2009) merely results in an approximation of the true Pareto front. The quality of this approximation depends upon (1) the proximity of the points on the approximated front to the points on the true Pareto front; and (2) the diversity of the points on the approximated front, where more diversity is typically better. These characteristics are important both in offline settings, in which one compares multiple fronts produced by competing algorithms, and online settings, in which a front is evaluated as it evolves with the rate of this evolution potentially used as a termination criterion. In this article, we are concerned with how Pareto fronts are evaluated and compared, rather than with algorithm development.

A popular measure of the quality of an approximated Pareto front is the front's *hypervolume* (Zitzler and Thiele, 1998), which measures the size of the space enclosed by all solutions on the Pareto front and a user-defined reference point (for a formal definition see Section 2). This measure of Pareto front quality has gained increasing interest in recent years and has become the standard offline indicator of the performance of multi-objective optimization algorithms (Zitzler et al., 2008). It has also been used as an online indicator to guide the optimization process (Knowles et al., 2003; Zitzler and Künzli, 2004; Emmerich et al., 2005; Beume et al., 2007; Bader and Zitzler, 2011). Its success and popularity are due to the fact that it simultaneously accounts for proximity and diversity and is strictly Pareto compliant. This means that whenever one Pareto front approximation dominates another, the hypervolume of the former is greater than that of the latter. One significant drawback to this measure is that its magnitude is dependent upon an arbitrarily chosen reference point. We will return to this point, in detail, in Sections 3 and 4.

Though the hypervolume measure is a well-established indicator of a front's quality, it has only recently been introduced to the statistics literature. Lu and Anderson-Cook (2012) develop a hypervolume-like indicator within the context of optimal experimental design, but there are several issues with their proposed measure: (1) they use different reference points for different approximate Pareto fronts which leads to unfair comparisons; (2) they choose the reference point in a way that does not permit a contribution to the hypervolume by all points; (3) Pareto compliance is not maintained because dominated points are used to compare an approximate front to the reference front; and (4) when used in an online setting, their proposed procedure can suggest decreases in Pareto front quality even as a front evolves in the context of a front construction algorithm. These issues are explained in more detail in Sections 2.2 and 4.1.

In this paper we address the aforementioned issues and propose an improved hypervolume-based measure for use in Pareto optimal design. In Section 2, we review the notion of Pareto optimal design, describe the computation of the hypervolume indicator, and explain in more detail the problems with the outstanding versions of the measure as well as our solutions to those problems. We also propose an interpretable scalar metric for describing how well a Pareto front is approximated and illustrate how the proposed indicator can be used in comparing competing Pareto fronts. In Section 3, we develop theoretical properties regarding the influence of the reference point on the calculation of the hypervolume indicator. Guidance is provided for choosing the reference point, in the presence of two criteria, based on our theoretical investigations, and we suggest a similar approach for the three-dimensional case. In Section 4, we illustrate our proposed procedure in both an offline and online setting in the context of multi-objective optimal experimental design, consider the influence of the reference point in a three-dimensional example, and explore the reasons for unintuitive decreases in the online uses of the hypervolume measure. In Section 5 we provide a recap and discussion.

Though we proceed with experimental design as the context, we note that our results and conclusions are more generally applicable to any multi-objective optimization setting in which the Pareto approach is employed and similar algorithms are used.

2. The hypervolume indicator and procedure for comparing discrete Pareto fronts

Without loss of generality, assume that the goal of a general multiple-criteria design optimization problem is to simultaneously maximize $C \geq 2$ design criteria. Let $\mathbf{f}(\xi) = (f_1(\xi), f_2(\xi), \ldots, f_C(\xi))'$ denote the $C \times 1$ vector of criterion values for design ξ. Let Ξ denote the *search space* of all feasible designs. A design $\xi_1 \in \Xi$ is said to dominate $\xi_2 \in \Xi$ if $f_j(\xi_1) \geq f_j(\xi_2)$ for all $j \in \{1, \ldots, C\}$ and there exists at least one $j \in \{1, \ldots, C\}$ for which $f_j(\xi_1) > f_j(\xi_2)$. In this case, the criterion vector $\mathbf{f}(\xi_1)$ is said to dominate the criterion vector $\mathbf{f}(\xi_2)$ and we write $\xi_1 \succ \xi_2$. If $f_j(\xi_1) \geq f_j(\xi_2)$ for all $1 \leq j \leq C$, we say ξ_1 weakly dominates ξ_2 and we write $\xi_1 \succeq \xi_2$. Henceforth, the criteria vector corresponding to a particular design is referred to as a *point* in the criterion space. A design is *Pareto optimal* if and only if no other design dominates it and its corresponding criterion vector is a *non-dominated* vector. The set of Pareto optimal designs constitutes the *Pareto optimal set* and the corresponding criterion vectors are said to be on the *Pareto front* or *frontier*. A good overview of the Pareto-related concepts is available in Coello Coello et al., 2007.

Given the experimental design setting, we treat every Pareto front as finite and discrete. We then assume that a given point on the Pareto front can be written as the ordered pair $(f_1(\xi), f_2(\xi)) \in \mathbb{R}^2$ in two dimensions or the ordered triplet $(f_1(\xi), f_2(\xi), f_3(\xi)) \in \mathbb{R}^3$ in three dimensions, where $f_1(\xi), f_2(\xi)$ and $f_3(\xi)$ correspond to design criterion values 1, 2 and