the british psychological society
promoting excellence in psychology

British Journal of Mathematical and Statistical Psychology (2020), 73, 375–396

www.wileyonlinelibrary.com

Combining diversity and dispersion criteria for anticlustering: A bicriterion approach

Michael J. Brusco[1]*, J. Dennis Cradit[1] and Douglas Steinley[2]
[1]Florida State University, Tallahassee, Florida, USA
[2]University of Missouri, Columbia, Missouri, USA

Most partitioning methods used in psychological research seek to produce homogeneous groups (i.e., groups with low intra-group dissimilarity). However, there are also applications where the goal is to provide heterogeneous groups (i.e., groups with high intra-group dissimilarity). Examples of these anticlustering contexts include construction of stimulus sets, formation of student groups, assignment of employees to project work teams, and assembly of test forms from a bank of items. Unfortunately, most commercial software packages are not equipped to accommodate the objective criteria and constraints that commonly arise for anticlustering problems. Two important objective criteria for anticlustering based on information in a dissimilarity matrix are: a *diversity* measure based on within-cluster sums of dissimilarities; and a *dispersion* measure based on the within-cluster minimum dissimilarities. In many instances, it is possible to find a partition that provides a large improvement in one of these two criteria with little (or no) sacrifice in the other criterion. For this reason, it is of significant value to explore the trade-offs that arise between these two criteria. Accordingly, the key contribution of this paper is the formulation of a bicriterion optimization problem for anticlustering based on the diversity and dispersion criteria, along with heuristics to approximate the Pareto efficient set of partitions. A motivating example and computational study are provided within the framework of test assembly.

1. Introduction

In most cluster analysis applications, the goal is to partition a set of objects into homogeneous and well-separated groups. That is, objects within the same group should be similar to one another (low intra-group dissimilarity), whereas objects in different groups should be diverse (high inter-group dissimilarity). Ward's (1963) hierarchical clustering method, *K*-means clustering (MacQueen, 1967; Steinley, 2006), and finite mixture modelling approaches for cluster analysis (Steinley & Brusco, 2011) are generally well suited for producing partitions with homogeneous and well-separated groups. Although much less common, there are also practical situations where the goal is to partition a set of objects into groups with high intra-group dissimilarity and low inter-group dissimilarity. That is, the objects within a group should be as diverse as possible, whereas pairs of groups should be very similar in composition. This type of partitioning is

*Correspondence should be addressed to Michael J. Brusco, Department of Business Analytics, Information Systems, and Supply Chain, Florida State University, Tallahassee, FL 32306-1110, USA (email: mbrusco@business.fsu.edu).

DOI:10.1111/bmsp.12186

known as *anticlustering*, a term originally coined by Späth (1986) who framed the problem as one of *anti-K*-means partitioning, whereby the objective was to maximize rather than minimize the within-group sum-of-squares criterion.

There are different types of scientific applications that make use of anticlustering principles. For example, Gillet, Wild, Willett, and Bradshaw (1998, p. 550) discuss applications pertaining to the selection of molecules from chemical databases where the goal is to choose subsets of molecules that are as diverse as possible. Fernández, Kalcsics, Nickel, and Ríos-Mercado (2010) report a fascinating application of anticlustering related to the location of waste recycling stations in Germany as part of the Waste Electric and Electronic Equipment (WEEE) directive of the European Commission. The nature of the anticlustering problem in this application hinged on the fact that the stations should be assigned to groups (companies) so that, for any given group, their stations were well dispersed across Germany.

The potential of anticlustering for psychological research was mentioned by Arabie and Hubert (1992) and, more recently, by Steinley (2006). In educational psychology, anticlustering is closely related to problems that arise in the context of test assembly, where the objective is to assign test items to multiple versions of an exam (Belov, 2008; Van der Linden, 2005, Ch. 6). For example, Sanders and Verschoor (1998, pp. 215–216) describe a problem where the goal is to assign a set of test items to produce parallel tests of equal length so as to maximize the minimum test reliability coefficient. The domain of test assembly also commonly involves the conversion of difficulty and discrimination statistics into pairwise dissimilarities among test items, which are subsequently used to facilitate the production of parallel test forms that have a high diversity of items within the test forms but low diversity across forms (Armstrong, Jones, & Wu, 1992; Wang, Zheng, Zheng, Su, & Li, 2016). Gierl, Daniels, and Zhang (2017) have recently underscored the importance of constructing fair (comparable difficulty level) parallel test forms within the context of on-demand testing for undergraduate psychology students.

There are also psychologically relevant applications in the educational and developmental psychology arena with respect to the assignment of students to groups (Baker & Powell, 2002; Hübscher, 2010). In these situations, the goal is typically to form student groups using a variety of general criteria such as age, gender, race, and income, or possibly context-related criteria such as courses taken, mathematics skills, or verbal skills. Similar problems occur in the assignment of participants to focus groups in consumer psychology, as well as industrial/organizational psychology contexts related to the formation of project workgroups or teams (Bhadury, Mighty, & Damar, 2000; Colarelli & Boos, 1992; Kelly, 2007). Anticlustering problems also arise in the selection of stimulus subgroups in experimental psychology (Brusco & Stahl, 2001; Brusco & Steinley, 2006; Coldren & Haaf, 2000; Compton, 2003; Derryberry, 1991; Lees & Neufeld, 1999). A sampling of anticlustering applications in psychology and other disciplines is provided in Table 1 (see Glover, Kuo, & Dhir, 1998, for additional examples).

In our research, we focus on what are arguably two of the most prominent and basic types of anticlustering criteria. The first criterion, which is especially popular, is known as *maximum diversity* (Bhadury *et al.*, 2000; Gallego, Laguna, Martí, & Duarte, 2013; Glover, Hersh, & McMillan, 1977; Glover *et al.*, 1998; Mingers & O'Brien, 1995; Miyaji, Ohno, & Mine, 1988; O'Brien & Mingers, 1997; Weitz & Lakshminarayanan, 1997, 1998). Maximum diversity partitioning seeks to assign objects to groups with the goal of maximizing the sum of the within-group sums of dissimilarities. In this sense, the goals of maximum diversity partitioning can be characterized as the *antithesis* of those typically associated with hierarchical clustering methods such as average linkage and Ward's

Table 1. Applications of anticlustering

Psychological applications	Other applications
1. Assigning students to groups	1. Assigning meeting attendees to dinner tables
2. Assigning test items to different exam versions	2. Sub-typing of viruses
3. Assigning workers to project teams	3. Analysis of microarray data
4. Assigning stimulus objects to subgroups	4. Very large-scale integration in circuitry
5. Assigning consumers to focus groups	5. Location of facilities

method and partitioning methods such as *K*-means. The second criterion is known as *maximum dispersion* (Fernández, Kalcsics, & Nickel, 2013), where the goal is to assign objects to groups so as to maximize the minimum dissimilarity between any two items in the same group. Accordingly, maximum dispersion partitioning is a monotone invariant clustering procedure with a goal that is essentially the antithesis of those typically associated with complete-linkage hierarchical clustering (McQuitty, 1960) and minimum diameter partitioning (Brusco & Steinley, 2014).[1]

The diversity criterion is the sum (across all groups) of all of the pairwise dissimilarities between objects within the same group. Thus, diversity can be viewed as a measure of the *total* dissimilarity in the partition (just as the total within-cluster sum of squares is the index for *K*-means clustering). By contrast, the dispersion criterion is not a sum, but the minimum dissimilarity across any two objects in the same group. Accordingly, dispersion is more of a *localized* measure of the *worst-case* pairwise dissimilarity across all groups (just as partition diameter is a worst-case measure in complete-linkage or minimum-diameter partitioning). Both criteria are important, yet they are not necessarily concordant. One could have a partition that maximizes the diversity criterion, yet has two very similar objects in the same cluster (thus a low level of dispersion). Likewise, one could have a partition that maximizes the dispersion criterion, yet has a total level of dissimilarity that is far from maximal (i.e., high diversity is not achieved). Ideally, we would like a partition that makes total pairwise dissimilarity large (i.e., high diversity), yet also keeps the worst-case pairwise dissimilarity large (i.e., high dispersion).

Given that the two most popular types of anticlustering objective criteria can be perceived as *reversals* of traditional clustering criteria, it is reasonable to postulate that anticlustering problems might be tackled using suitable transformations of the dissimilarity data in conjunction with commercially available software programs employing standard clustering methods. Unfortunately, such an approach is impractical for several reasons: (1) data transformation can result in a partitioning problem that is discordant with the original problem[2]; (2) for anticlustering to have practical merit in psychological applications (e.g., assigning students to groups, workers to teams, test questions to exams, or stimulus items to subsets), it is common to impose constraints on group size, and such constraints are typically not feasible in commercial software programs; (3) commercial

[1] Although not considered in this paper, one could also propose an anticlustering criterion that is the antithesis of single-linkage hierarchical clustering, whereby the criterion would be to choose the linkage that maximizes the maximum dissimilarity between any two objects not in the same group.

[2] Data transformations are potentially less problematic for monotone invariant clustering procedures such as complete linkage, but can be problematic for the diversity criterion. For example, a partition that maximizes the within-group sums of dissimilarities is not necessarily the same as the partition that minimizes the within-group sums of the inverse of the dissimilarities.

software packages do not allow for the easy assessment of trade-offs among two or more criteria.

In this paper, we present a method that (1) can be directly applied to the raw dissimilarity data without transformation, (2) directly handles the cardinality constraints, and (3) allows for examination of trade-offs pertaining to the diversity and dispersion criteria. Expanding on point (3), it is important to recognize that, in some instances, there might be perfect (or near-perfect) concordance of the partitions obtained by maximum diversity and maximum dispersion partitioning. However, in many other instances, there might be considerable antagonism, whereby the partition that maximizes diversity results in a poor value for dispersion and vice versa. In light of the fact that a researcher is generally unaware of the extent of concordance/antagonism, there is a need for methods that consider both criteria so as to investigate the results for each criterion. Accordingly, we formulate a bicriterion combinatorial optimization that explores the trade-offs between the diversity and dispersion criteria. General approaches for approximating the Pareto efficient (non-dominated) set for bicriterion optimization problems (Ehrgott, 2005) are described. Two specific procedures, a pairwise interchange heuristic and an iterated local search heuristic (Lourenço, Martin, & Stützle, 2003, 2010), are developed for Pareto set approximation for the bicriterion anticlustering problem.

Section 2 presents formulations of the maximum diversity and maximum dispersion partitioning problems that can be used for anticlustering. A mathematical formulation of a bicriterion combinatorial optimization problem that considers both criteria is also provided. The iterated local search heuristic for anticlustering is described in Section 3. A numerical example in the context of test assembly is provided in Section 4. A computational study comparing the bicriterion multiple-restart (multistart) pairwise interchange and iterated local search heuristics is presented in Section 5. The paper concludes in Section 6 with a brief summary and a discussion of limitations and possible extensions.

2. Diversity and dispersion criteria for anticlustering

2.1. Diversity measures

We assume that there is a set, $S = \{1, 2, \ldots, N\}$, consisting of N objects identified by their index number. These objects could represent students, patients, employees, test questions, or any other set of items to be placed into groups. The number of groups is designated by G, and the groups are indexed by the letter g (and, when also necessary, h). A partition of the object set (S) is indicated by $\pi = \{S_1, \ldots, S_G\}$, where S_g contains the subset of objects assigned to group g, for all $1 \leq g \leq G$. The set of all possible partitions of the N objects into G groups is designated by Π. The number of partitions in Π, which can be computed as a Stirling number of the second kind (see Brusco & Stahl, 2005, p. 18), is enormous for practical values of N and G. For the moment, we will assume that all partitions in Π are permissible; however, there are often practical constraints on the assignment of objects to groups that result in the infeasibility of a large proportion of the partitions in Π.

We also assume that an $N \times N$ non-negative symmetric matrix, $\mathbf{D} = [d_{ij}]$, of pairwise dissimilarity measurements is available for the N objects. The dissimilarity measures frequently correspond to pairwise Euclidean distances (or squared Euclidean distances) between items computed across a set of variables or attributes. However, in other

applications, the dissimilarity measures might be obtained from correlation analysis, paired-comparison similarity judgements, sorting tasks, or other approaches.

The most common measure of diversity is based on a within-group sum of dissimilarities measure that is computed as follows:

$$f_1(S_g) = \sum_{(i<j)\in S_g} d_{ij}, \text{ for } 1 \leq g \leq G. \tag{1}$$

The goal is to seek the partition, π, that maximizes the sum of the $f_1(S_g)$ values across all groups:

$$\max_{\pi\in\Pi}\{F_1(\pi)\}, \quad \text{where } F_1(\pi) = \sum_{g=1}^{G} f_1(S_g). \tag{2}$$

Although the unrestricted version of maximum diversity partitioning is a viable optimization problem in its own right, in practice it is far more common to encounter cardinality constraints on the grouping process (see Brusco & Stahl, 2001; Gallego *et al.*, 2013). That is, the number of objects in each cluster is assumed to be a prespecified size, M_g, for $1 \leq g \leq G$. The prespecification of group sizes (which are often equal) is common in applications such as the formation of student groups or the psychologically familiar context of parallel test assembly (Van der Linden, 2005). In this latter context, N would be the total number of available test items, G would be the number of parallel test forms, and M_g would be the number of items for each test. The d_{ij} values would be measures of dissimilarity between pairs of items. These dissimilarities could be obtained in a variety of ways, with one popular measure being a reliability-index distance that is based on the difficulty and discrimination statistics for the items (Armstrong *et al.*, 1992; Wang *et al.*, 2016).

2.2. Dispersion measure

The primary dispersion within-group measurement is the *minimum* dissimilarity between any two items in the group:

$$f_2(S_g) = \min_{(i<j)\in S_g}\{d_{ij}\}, \text{ for } 1 \leq g \leq G. \tag{3}$$

The goal is to seek the partition, π, that maximizes the minimum of the $f_2(S_g)$ values across all groups:

$$\max_{\pi\in\Pi}\{F_2(\pi)\}, \quad \text{where } F_2(\pi) = \min_{1\leq g\leq G}\{f_2(S_g)\}. \tag{4}$$

The problem posed by equation (4) has been described as the *unrestricted maximum dispersion* grouping problem (Fernández *et al.*, 2013). However, the problem can also be described as an anticlustering problem because, as is the case for Späth's (1986) anticlustering version of *K*-means, the goal is to maximize (rather than minimize) intra-group heterogeneity. Finally, the problem posed by equation (4) is a monotone invariant anticlustering problem because any order-preserving transformation of the input dissimilarities will not affect the solution obtained by the partitioning method.

2.3. Bicriterion combinatorial optimization model

Ideally, we would like to find a partition, π, that simultaneously maximizes both $F_1(\pi)$ and $F_2(\pi)$. Unfortunately, such a partition frequently does not exist and it is necessary to establish a bicriterion combinatorial optimization model that allows for trade-offs between the criteria. Specifying ω_1 and ω_2 as the priorities for $F_1(\pi)$ and $F_2(\pi)$, respectively, the bicriterion formulation of the model is

$$\max_{\pi \in \Pi} \{Z(\pi)\} = \omega_1 F_1(\pi) + \omega_2 F_2(\pi). \tag{5}$$

When considering the bicriterion optimization problem, it would be beneficial for an analyst to have information pertaining to the Pareto efficient (or *non-dominated*) set of partitions, which we denote as *P*. A partition (π) dominates another partition (π') if the diversity and dispersion measures for π equal or exceed those of π', that is, $F_1(\pi) \geq F_1(\pi')$ and $F_2(\pi) \geq F_2(\pi')$, and strictly exceed those of π' for at least one of the two measures, that is, $F_1(\pi) > F_1(\pi')$ and/or $F_2(\pi) > F_2(\pi')$.

There are at least three possible approaches to generating non-dominated solutions (see Ehrgott, 2005): the weighted-sum method; the constraint method; and the direct method. The *weighted-sum* or scalarization method (Brusco & Steinley, 2009) converts the problem into one with a single, weighted criterion function. When using the *constraint* method (see Brusco & Cradit, 2005), the value of one criterion is constrained to some threshold while the other criterion is maximized. The *direct* method (Delattre & Hansen, 1980; Ferligoj & Batagelj, 1992), as the name implies, uses a direct search of the solution space for the optimization problem and updates the Pareto efficient set (adding and removing solutions) during the search process.

The three approaches to generating non-dominated solutions are, essentially, different strategies for accomplishing the same goal of finding the Pareto efficient set. If it were possible to directly evaluate all possible partitions of the *N* objects into *G* groups, then the complete Pareto set could be obtained via this direct approach. Likewise, if it were possible to maximize one of the two criteria (e.g., diversity) subject to a threshold constraint for every possible value for the other criterion (e.g., dispersion), then the complete Pareto set could be generated using the constraint method. The weighted-sum approach corresponds to the formulation of equation (5) using a pair of non-negative weights such that $0 < \omega_1 < 1$, and $\omega_2 = 1 - \omega_1$. A globally optimal solution to this formulation provides a non-dominated partition. There are two major limitations associated with the weighted-sum approach. First, the set of feasible weights is infinite and the *a priori* selection of weights can be extraordinarily difficult, particularly in circumstances where the criterion functions differ markedly in scale. Second, the weighted-sum approach produces only *supported* Pareto efficient solutions, but bicriterion combinatorial optimization problems might also have *unsupported* Pareto efficient solutions that cannot be identified with any weighting scheme (see Ehrgott, 2005). To illustrate, consider a Pareto set that consists of three solutions (A, B, C). Solution A has values of $F_1(\pi) = 80$ and $F_2(\pi) = 40$. Solution B has values of $F_1(\pi) = 40$ and $F_2(\pi) = 80$. Solution A has a greater weighted sum than B for $\omega_1 > \omega_2$, and solution B has a greater weighted sum than A for $\omega_1 < \omega_2$. The two solutions have the same weighted sum when $\omega_1 = \omega_2 = 0.5$. The third solution in the Pareto set, solution C, has values of $F_1(\pi) = 50$ and $F_2(\pi) = 50$. Clearly C is not dominated by either A or B, its $F_2(\pi)$ value is larger than that of A, and its $F_1(\pi)$ value is larger than that of B. However, there are no values of ω_1 and ω_2 such that C will have the highest weighted sum. At $\omega_1 = \omega_2 = 0.5$, the

weighted sums for A, B, and C are 60, 60, and 50, respectively. Increasing ω_1 will improve solution A's weighted sum relative to C, whereas decreasing ω_1 will improve solution B's weighted sum relative to C. Thus, in this example, A and B are supported Pareto efficient solutions, whereas C is an unsupported Pareto efficient solution.

For bicriterion combinatorial optimization problems that are small or have particular structures that can be exploited, it is sometimes possible to generate the Pareto efficient set of solutions using mathematical programming solutions or specially designed algorithms (see, for example, Brusco, 2015, 2017; Delattre & Hansen, 1980). Such examples, however, are the exceptions, not the rule. Larger, more complex problems generally require the use of heuristic procedures to *approximate* the Pareto efficient set (see Brusco, Doreian, Steinley, & Satornino, 2013; Brusco & Steinley, 2012). Given the computational challenges of the maximum diversity problem, as well as the fact that some anticlustering problems can have hundreds of objects, we restrict our focus to the *approximation* of the Pareto efficient set for the bicriterion anticlustering problem. In particular, in the next section, we describe a pairwise interchange heuristic for approximating the Pareto efficient set, as well as an extension of this heuristic within the framework of iterated local search.

3. Bicriterion local search heuristic

3.1. A multistart bicriterion pairwise interchange heuristic

We developed a multiple restart (multistart) bicriterion pairwise interchange (MBPI) heuristic for approximating the Pareto efficient set for the bicriterion anticlustering problem. Although the MBPI algorithm seeks to maximize a weighted sum of the diversity and dispersion criteria, it is best classified as a direct algorithm for the bicriterion optimization problem because all partitions generated are evaluated with respect to the incumbent Pareto set (P) and required adjustments are made to P if necessary. The MBPI algorithm is a stand-alone program written in Fortran 95 and compiled using GNU Fortran. The algorithm was implemented on a desktop computer with an Intel® Core™ i7-6700T CPU (Intel Corporation, Santa Clara, CA, USA) at 2.8 GHz with 16 GB of RAM. The pseudocode for the heuristic is provided in Figure 1.

The required input is a dissimilarity matrix (**D**), as well as all relevant parameter information such as the number of objects (N), the number of groups (G), and the desired size for each group (M_g, for $1 \leq g \leq G$). The desired number of restarts for the MBPI, R, is provided by the user as input. The set of Pareto efficient partitions is initialized to the null set, $P = \varnothing$. Each restart of the MBPI heuristic begins with the selection of a pair of nonnegative weights (ω_1 and ω_2) for the diversity and dispersion criteria such that $\omega_1 + \omega_2 = 1$. The value of ω_1 is randomly selected from a set of prespecified options based on a uniform distribution and $\omega_2 = 1 - \omega_1$. For the analyses reported later in this paper, ω_1 was selected from among 10 options (.000001, .00001, .0001, .001, .01, .1, .5, .99, .999, .999999), with equal probability for each option.

The next step is the generation of a feasible partition (π) of the N objects into G groups of the appropriate sizes. This is accomplished by generating the proper number of cluster labels for each group and then randomly assigning the labels to the objects. The resulting partition serves as the initial partition for the pairwise interchange search process for maximizing $Z(\pi)$ in equation (5), which runs until convergence. Because the MBPI heuristic uses only exchanges of group memberships for two randomly selected objects not in the same group as the local search process for generating trial solutions, the number

```
INPUT the dissimilarity matrix, D, and a desired number of restarts for the algorithm, R.
SET the initial Pareto set as P = ∅.
For r = 1 to R
        SELECT a convex combination of weights: 0 ≤ ω1 ≤ 1 and ω2 = 1 - ω1
        CONSTRUCT an initial partition, π = {S1,…,SG}, by randomly assigning group
        labels to each item.
        EXECUTE LOCAL SEARCH
        Set Z* = ω1F1(π) + ω2F2(π)
        Flag = 0
        While Flag = 0
                Set Flag = 1
                For i = 1 to N-1
                        For j = i+1 to N
                                set g : i ∈ Sg
                                set h : j ∈ Sh
                                If g ≠ h then
                                        Modify π
                                        Set Sg = Sg ∪ {j} \ {i}
                                        Set Sh = Sh ∪ {i} \ {j}
                                        Update P as appropriate
                                        Compute Z = ω1F1(π) + ω2F2(π)
                                        If Z > Z* then
                                                Set Z* = Z
                                                Set Flag = 0
                                        Else
                                                Set Sg = Sg ∪ {i} \ {j)
                                                Set Sh = Sh ∪ {j} \ {i)
                                        End if
                                End if
                        Next j
                Next i
        End While
Next r
RETURN P.
```

Figure 1. Pseudo-code for the MBPI heuristic.

of objects in each group never changes during the algorithm. Thus, feasibility with respect to group size is preserved by the generation of the initial partition in conjunction with the search process. Each trial partition that is generated by the pairwise interchange search process is evaluated for inclusion in the Pareto efficient set (P). If a trial partition, π', is not dominated by any $\pi \in P$, then π' is added to P (i.e., $P = P \cup \pi'$). When P is augmented in this manner, it is also necessary to determine if the newly added partition dominates any of the other partitions in P. That is, any partition, $\pi \in P$, that is dominated by π' is removed from P (i.e., $P = P \backslash \pi$).

The pairwise interchange search process continues until there is no further interchange that will improve $Z(\pi)$. Upon convergence of the pairwise search process, r is incremented (i.e., $r \leftarrow r + 1$) and the next restart of the heuristic repeats the process of the selection of a set of weights, the construction of new initial partition, and the pairwise interchange search process to maximize $Z(\pi)$. The MBPI algorithm continues in

this manner until $r = R$. Throughout the search process, P is evaluated for update after generation of every trial solution in each restart.

3.2. A bicriterion iterated local search heuristic

It is well recognized in the discrete optimization literature that multistart local search heuristics commonly fail to identify globally optimal partitions as problem size increases (Johnson & McGeoch, 1997; Lourenço *et al.*, 2003, 2010; Schreiber & Martin, 1999). The crux of the problem is the propensity for multistart local search heuristics to produce solutions with an average percentage deviation from the mean that increases with problem size. The tails of the distribution for criterion function values also tend to flatten as problem size increases and it becomes less and less likely that the multistart approach will find a global optimum (or even a near-global optimum) in the tail corresponding to the best criterion values.

Lourenço *et al.* (2003, 2010) formalized iterated local search (ILS) as a general solution strategy for mitigating the limitations of the multistart approach. Rather than using multiple restarts where the initial solution for each restart is randomly constructed, an ILS procedure conducts a vigorous search of the neighbourhood of an existing (incumbent) solution. This is accomplished via a two-phase process whereby the incumbent solution is randomly perturbed to a solution that is not locally optimal and, subsequently, is returned to local optimality by the local search heuristic. Brusco, Shireman, Steinley, Brudvig, and Cradit (2017) recently demonstrated the effectiveness of ILS relative to multistart relocation heuristics in the context of Gaussian model-based partitioning. One of the key aspects of their findings is that it is often advantageous to use the multistart location heuristic to provide a good initial solution that can subsequently be refined by ILS. We adopt a similar approach here in our bicriterion implementation of ILS (i.e. BILS) for anticlustering.

Figure 2 presents the pseudo-code for the BILS algorithm. The first $R/2$ iterations of the BILS algorithm are identical to those of the MBPI heuristic. That is, the MBPI heuristic is used to obtain an initial approximation of the Pareto set, P. For the remaining $R/2$ iterations of the BILS algorithm, rather than randomly generating an initial partition, one of the partitions from P is randomly selected. This partition is subsequently perturbed by examining each pair of objects and, if they are not currently in the same group, then a probabilistic decision is made as to whether or not their group membership assignments should be swapped. At each iteration, the probability of swapping the assignments is randomly selected from a uniform distribution on the interval $[\xi_1, \xi_2]$ (we used [.05, .10] in our analyses). Once the partition has been perturbed, it is restored to local optimality using the pairwise interchange heuristic.

One of the inherent advantages of ILS is that it is conceptually more straightforward and requires less parameterization than metaheuristics such as tabu search (Glover & Laguna, 1993), genetic algorithms (Goldberg, 1989), and simulated annealing (Cerny, 1985; Kirkpatrick, Gelatt, & Vecchi, 1983). The major parameter decision in our BILS implementation is the determination of the range for the neighbourhood size, $[\xi_1, \xi_2]$. Like the MBPI, the BILS algorithm is a stand-alone program written in Fortran 95 and compiled using GNU Fortran. The BILS algorithm was also implemented on the same hardware platform as the MBPI.

3.3. Selecting a partition from the Pareto set

Once an approximation of the Pareto set has been obtained (using either MBPI or BILS), there remains the issue of selecting a partition from the Pareto set. This decision is

```
SELECT a desired number of restarts, R and a neighborhood size range [ξ1, ξ2]
RUN R/2 restarts of the MBPI (Figure 1) and let P be the approximated Pareto set
For q = 1 to R/2
    SELECT weights: 0 ≤ ω1 ≤ 1 and ω2 = 1 - ω1
    CHOOSE a neighborhood size ξ from a uniform distribution of the interval [ξ1, ξ2].
    RANDOMLY SELECT a partition, π, from P
    PERTURB π into a neighboring partition
    For i = 1 to N-1
        For j = i+1 to N
            set g : i ∈ Sg
            set h : j ∈ Sh
            If g ≠ h then
                Generate a uniform random number r from the interval [0, 1]
                If r < ξ then
                    Set Sg = Sg ∪ {j} \ {i}
                    Set Sh = Sh ∪ {i} \ {j}
                End if
            End if
        Next j
    Next i
    RE-APPLY the local-search component of the MBPI using π, ω1, and ω2 as input.
Next q
RETURN π as the partition.
```

Figure 2. Pseudo-code for the BILS heuristic.

principally based on two factors: the nature of the trade-offs between the two criteria; and the decision-maker's assessment of the relative importance of the two criteria. There are several ways to facilitate an understanding of the trade-offs between the two criteria. A useful initial step is to plot the criterion function values in two-dimensional space. There is commonly a small subset of partitions that produce a convex hull for the plot and the partitions that define the hull can help to narrow the set of candidates when there are dozens or hundreds of partitions in the Pareto efficient set. Another option in such circumstances is to prune the Pareto set to a more manageable set of solutions for the decision analyst to consider (see Brusco, 2017).

There are also some straightforward rules for selecting a single partition from the Pareto efficient set. Perhaps the simplest rule is to choose the partition that is at the maximum distance from the origin. However, because the diversity measure is almost always dramatically larger than the dispersion measure, such a rule will lead to the selection of the partition with the largest value of the diversity criterion. To level the playing field, it is possible to standardize the criterion values in some fashion. One approach is to standardize the diversity and dispersion measures based on the percentage of the best-found criterion values for each measure (see Brusco *et al.*, 2013). For example, a partition might yield a diversity index that is 99% as large as the best-found diversity index and a dispersion index that is 95% as large as the best-found dispersion index. Unfortunately, there is also the potential for bias with this rule because, when using a percentage basis, there is typically much more variation for diversity than there is for dispersion.

There is no definitive rule for selecting a solution for the Pareto set that works well in all circumstances. A plot of the Pareto set is a good starting point and enables the decision-maker to focus on particular regions of the plot that are in concordance with the relative importance of the predictors for a given application. In Section 5.3 we demonstrate how the plot, along with some measures of deviation from the origin, can be used to facilitate the selection of a partition.

4. A motivating example

We consider a small numerical example to motivate the consideration of both diversity and dispersion in the anticlustering process. The data, which were originally collected by Tatsuoka (1990), measure the performance of 536 schoolchildren on $N = 20$ test items pertaining to the subtraction of fractions. These data have also been analysed by de la Torre and Douglas (2004) in the context of cognitive diagnosis. Our goal in this example is not a rigorous statistical modelling of the fraction subtraction data, but to show how a researcher might apply bicriterion anticlustering to raw exam performance data so as to form multiple parallel tests based on item difficulty (Gierl *et al.*, 2017) and/or item discrimination.

In our example, we compute measures of item difficulty and item discrimination for each of the 20 items. For each item, difficulty is measured as the percentage of children who answered the item correctly. Discrimination is a measure of how well an item distinguishes between the best-performing (top 25%) and worst-performing (bottom 25%) students. Following Armstrong *et al.* (1992) and Wang *et al.* (2016), we used the Euclidean distance between items based on their difficulty and discrimination estimates to obtain the dissimilarities between pairs of items (i.e., the d_{ij} values). However, the same type of analyses could be completed using item difficulty only (see, e.g., Gierl *et al.*, 2017).

We used the MBPI and BILS heuristics to approximate the Pareto efficient set of partitions of the 20 test items into $G = 2$ groups of size 10, $G = 4$ groups of size five, $G = 5$ groups of size four, and $G = 10$ groups of size two. Both algorithms were implemented using $R = 10{,}000$. There was no difference between the two heuristics with respect to the Pareto sets that were approximated. Table 2 reports the diversity and dispersion values for the Pareto sets for each of the different values of G. Because the number of test items ($N = 20$) is relatively small, we were able to verify that the *true* (or complete) Pareto efficient set was obtained for each value of G. This verification was accomplished in two stages. First, a modified version of a branch-and-bound algorithm for minimizing partition diameter (Brusco, 2003) was used to ensure that a partition with the maximum possible dispersion was contained in the Pareto set. Second, an adaptation of a branch-and-bound implementation of the constraint method (see Brusco & Stahl, 2005, pp. 82–84) was used to verify the remainder of the Pareto efficient set by maximizing diversity subject to a threshold constraint for dispersion.

The approximated Pareto sets for $G = 2$, $G = 4$ and $G = 5$ are small and relatively uninteresting. The approximated Pareto set for $G = 10$ is slightly larger and more useful for examining the dispersion–diversity trade-offs. Table 3 provides a more thorough breakdown of the two extreme partitions in the Pareto set for $G = 10$.

An examination of the results for $G = 10$ in Table 2 reveals that the diversity measures of the partitions in the approximated Pareto set fall within a very narrow range, $3.39615 \leq F_1(\pi) \leq 3.41539$, such that the largest value is only 0.6% greater than the smallest value. Contrastingly, the dispersion measures fall within a much broader range,

Table 2. Dispersion and diversity measures of the Pareto efficient sets for the subtraction data at different values of G

$G = 2$		$G = 4$		$G = 5$		$G = 10$	
Dispersion	Diversity	Dispersion	Diversity	Dispersion	Diversity	Dispersion	Diversity
.04851	22.83729	.08243	10.99748	.10919	8.56379	.25706	3.39615
.04675	22.85865	.08228	11.01964	.10687	8.57464	.25646	3.40079
		.07847	11.02255	.08022	8.58299	.22308	3.40548
						.20936	3.41193
						.13582	3.41435
						.08711	3.41539

Table 3. Extreme partitions in the approximated Pareto set ($G = 10$) for the subtraction data

	Diversity-maximizing partition				Dispersion-maximizing partition			
Group	Item	Diff.	Disc.	Distance	Item	Diff.	Disc.	Distance
1	1	.5317	.9179	0.0871091	4	.5280	.7239	0.2570636
1	18	.4683	.8582		17	.4328	.9627	
2	3	.5149	.9328	0.2093629	3	.5149	.9328	.2701619
2	4	.5280	.7239		5	.5840	.6716	
3	5	.5840	.6716	0.2903323	12	.7295	.7612	0.2786292
3	11	.4739	.9403		18	.4683	.8582	
4	10	.3825	.8806	0.3205482	2	.5765	.9403	0.2893235
4	16	.6959	.8134		19	.3097	.8284	
5	2	.5765	.9403	0.3338301	1	.5317	.9179	0.2840788
5	13	.2948	.7612		13	.2948	.7612	
6	7	.3993	.9179	0.3655241	7	.3993	.9179	0.3145027
6	12	.7295	.7612		16	.6959	.8134	
7	14	.7220	.8209	0.4123810	10	.3825	.8806	0.3447608
7	19	.3097	.8284		14	.7220	.8209	
8	9	.6474	.5746	0.4434219	9	.6474	.5746	0.4047476
8	17	.4328	.9627		11	.4739	.9403	
9	6	.7948	.6791	0.4762106	6	.7948	.6791	0.4762106
9	20	.3918	.9328		20	.3918	.9328	
10	8	.7481	.5448	0.4766818	8	.7481	.5448	0.4766818
10	15	.4515	.9179		15	.4515	.9179	
Minimum intra-group distance (dispersion)				0.0871091				0.2570636
Sum of intra-group distances (diversity)				3.4154021				3.3961603

$0.08711 \leq F_2(\pi) \leq 0.25706$, such that the largest value is nearly three times (295%) as large as the smallest value. The comparison of the diversity-maximizing partition, $F_1(\pi) = 3.41539$, $F_2(\pi) = 0.08711$, and dispersion maximizing partition, $F_1(\pi) = 3.39615, F_2(\pi) = 0.25706$, in Table 3 shows the nature of the trade-offs between diversity and dispersion. Most of the groups in the diversity-maximizing partition consist of two items that have markedly different measures of difficulty and discrimination. However, group 1 consists of two items (items 1 and 18) that have rather similar levels of

both difficulty (.5317, .4683) and discrimination (.9179, .8582). This clearly reveals the major limitation of focusing solely on the diversity measure. That is, in order to maximize the dissimilarity sum, there is the potential for one or more groups that have two or more items that are quite similar.

The results for the dispersion-maximizing partition in Table 3 reveal that a small sacrifice in diversity enables the circumvention of a group with similar levels of both difficulty and discrimination. That is, group 1 of the dispersion-maximizing partition consists of two items (items 4 and 17) that have more dissimilar levels of both difficulty (.5280, .4328) and discrimination (.7239, .9627) relative to the diversity-maximizing partition. Thus, in this particular example, it is likely that a decision-maker would prefer the dispersion-maximizing partition because the trivial reduction in diversity leads to a huge improvement in dispersion. However, in the absence of the bicriterion approach, this trade-off would not have been detected. It is also important to recognize another advantage of the bicriterion approach that pertains to tie-breaking. There are multiple other partitions that have the same maximum dispersion measure of .25706, yet have lower levels of diversity. A method that optimized only dispersion (ignoring diversity) could yield one of these other partitions and thus result in an unnecessary sacrifice of diversity.

5. Computational study

5.1. Test problems

The example in the previous section did not reveal any differences between the approximate Pareto sets obtained by the MBPI and BILS algorithms. Therefore, to provide a better comparison, we designed a computational study using a test-problem generation process from the cognitive diagnosis literature. The blueprint for the generation of the test problems stems from the work of Straat, Van der Ark, and Sijtsma (2013) and has also been used by Brusco, Köhn, and Steinley (2015). The foundation for the design used by Straat *et al.* (2013) is the two-dimensional graded response model (Samejima, 1969). However, like Brusco *et al.* (2015), we restrict our attention here to the dichotomous case with additive traits on the logit scale. This special case is known as the multidimensional two-parameter logistic model (MD2PL; Reckase, 2009).

Following Straat *et al.* (2013) and Brusco *et al.* (2015), we assume two latent traits; however, we use a slightly larger number of test items ($N = 50$). Three other design features were manipulated to generate the test problems. First, the two latent variables (θ_1, θ_2) were generated from a bivariate normal distribution using four different levels of correlation (ρ): $\rho = 0$ (uncorrelated), $\rho = .35$, $\rho = .7$, and $\rho = 1.0$ (unidimensional). Second, for each item j, the difficulty parameter (δ_j) was sampled from a continuous uniform distribution using either [–1.5, 1.5] or [–3, 3]. Third, the mean of the item discrimination parameters ($\bar{\alpha}$) was evaluated at three levels: $\bar{\alpha} = 1$, $\bar{\alpha} = 1.25$, and $\bar{\alpha} = 1.5$. For each of the three levels, the discrimination values (α_{jl}) were generated using a normal distribution with mean $\bar{\alpha}$ and standard deviation 0.1. These values were then refined by setting $\alpha_{j2} = 0$ for the first half of the items and $\alpha_{j1} = 0$ for the second half of the items.

All combinations of the three design features were considered, resulting in $4 \times 2 \times 3 = 24$ unique data sets, which were generated in MATLAB based on the MD2PL as follows:

$$P\left(X_j = 1 \mid (\theta_1, \theta_2)\right) = \frac{\exp\left(\sum_{l=1}^{2} \alpha_{jl}(\theta_l - \delta_j)\right)}{1 + \exp\left(\sum_{l=1}^{2} \alpha_{jl}(\theta_l - \delta_j)\right)}. \quad (6)$$

For each test data set, item difficulty and item discrimination were obtained in the same manner described for the motivating example in Section 4. For each of the 24 data sets, anticlustering solutions were produced for $G = 5$ groups of 10 items each, as well as for $G = 10$ groups of five items each, which yields a total of 48 test problems. For each test problem, MBPI and BILS were implemented using $R = 10{,}000$.

5.2. Computational results

The average computation time for MBPI was 220.10 s with a standard deviation of 14.10 s, whereas the average for BILS was 229.34 s with a standard deviation of 15.14 s. The similarity of the computation times for the two methods is largely attributable to the fact that both algorithms were allotted $R = 10{,}000$ restarts. However, unlike MBPI, the latter 5,000 restarts of BILS use randomly selected partitions from the approximated Pareto efficient set. Because each of these partitions is subjected to random perturbation, slightly more computation time is used for the BILS algorithm.

Table 4 provides a summary of the results for the 48 test problems. As a first step, we rationalized the Pareto efficient set approximations obtained by MBPI and BILS. That is, for each test problem, all the partitions obtained by MBPI that were dominated by a partition from BILS were discarded. Partitions obtained by BILS that were dominated by a partition from MBPI were also eliminated. This process results in a *synthesized* Pareto efficient set (MBPI + BILS) that contains only partitions that were not dominated by a partition produced by either method. Accordingly, the synthesized Pareto set can be considered the *best available approximation* because it considers the results obtained by both algorithms jointly. The individual performances of the MBPI and BILS methods can then be compared on the basis of how many partitions they obtained that are in the synthesized Pareto efficient set.

The results in Table 4 illustrate that MBPI and BILS yielded comparable performances. For the $G = 5$ test problems, the MBPI algorithm found an average of 85.0% of the partitions in the synthesized Pareto efficient set, whereas the corresponding figure was 82.5% for BILS. The MBPI algorithm perfectly matched the synthesized Pareto efficient set for 16 of the 24 (66.7%) test problems at $G = 5$. By contrast, the MBPI algorithm perfectly matched the synthesized Pareto efficient for only 13 of the 24 (54.2%) of the $G = 5$ test problems.

The approximated Pareto efficient sets tended to be somewhat larger for the $G = 10$ test problems. The relative performances of BILS and MBPI were still comparable; however, there was a modest advantage for BILS. For the $G = 10$ test problems, the BILS algorithm found an average of 87.7% of the partitions in the synthesized Pareto efficient set, whereas the corresponding figure was 80.2% for MBPI. The BILS algorithm perfectly matched the synthesized Pareto efficient set for 11 of the 24 (45.8%) test problems at $G = 10$, whereas the MBPI algorithm perfectly matched the synthesized Pareto efficient for 10 of the 24 (41.7%) $G = 10$ test problems.

It is important to acknowledge that, in most cases, the criterion function values obtained by MBPI and BILS exhibited only slight differences. For example, it was common for the two algorithms to find partitions with the same dispersion index, but slightly different (i.e., <0.1% difference) values for the diversity index. In such circumstances, the

Table 4. Results for simulated data from the MD2PL model: Number of solutions on the approximated Pareto set obtained by the multistart binary pairwise interchange (MBPI) and binary iterated local search (BILS) heuristics

ρ	δ	$\bar{\alpha}$	Results for $G = 5$			Results for $G = 10$		
			MBPI + BILS	BPI	BILS	MBPI + BILS	BPI	BILS
0	[−1.5,1.5]	1	5	5	4	14	9	12
0	[−1.5,1.5]	1.25	8	4	6	12	7	12
0	[−1.5,1.5]	1.5	4	4	3	15	12	14
0	[−3, 3]	1	5	5	1	19	13	15
0	[−3, 3]	1.25	5	5	3	16	13	15
0	[−3, 3]	1.5	2	2	2	10	10	9
.35	[−1.5,1.5]	1	5	5	5	10	5	8
.35	[−1.5,1.5]	1.25	5	3	4	11	6	8
.35	[−1.5,1.5]	1.5	11	5	10	18	13	18
.35	[−3, 3]	1	10	8	7	11	10	10
.35	[−3, 3]	1.25	6	6	6	10	10	10
.35	[−3, 3]	1.5	7	7	6	5	5	5
.7	[−1.5,1.5]	1	6	5	6	10	8	3
.7	[−1.5,1.5]	1.25	5	4	5	13	9	10
.7	[−1.5,1.5]	1.5	3	3	3	8	8	8
.7	[−3, 3]	1	7	7	7	10	9	9
.7	[−3, 3]	1.25	7	6	5	7	7	7
.7	[−3, 3]	1.5	4	4	4	4	4	4
1	[−1.5,1.5]	1	4	3	1	6	5	5
1	[−1.5,1.5]	1.25	1	1	1	8	8	8
1	[−1.5,1.5]	1.5	1	1	1	11	9	8
1	[−3, 3]	1	2	2	2	5	5	5
1	[−3, 3]	1.25	5	5	5	5	5	5
1	[−3, 3]	1.5	2	2	2	5	5	5

Note. The values in the MBPI + BILS columns are the number of partitions in the approximated Pareto set obtained from a synthesis of the partitions obtained by MBPI and BILS individually. The values in the MBPI and BILS columns are, respectively, the number of partitions obtained by MBPI and BILS that are included in the approximated set formed by the synthesis of the two algorithms.

algorithm with the inferior diversity index would fail to find the partition on the synthesized Pareto set as described above; however, it nonetheless found a partition that was only slightly worse than one in Pareto set.

5.3. Selecting a partition from the Pareto efficient set

To illustrate the process of selecting a partition from the Pareto efficient set, we consider the results for the test problem in Table 4 associated with $\rho = 0$, $\delta = [-3, 3]$, $\bar{\alpha} = 1$, and $G = 10$. The approximate Pareto set for this problem consists of 19 partitions. To provide some additional confidence in the approximation, we ran the BILS algorithm using $R = 40{,}000$, which required just over 1,000 s of computation time. Although there was some slight improvement in a few of the diversity indices, the Pareto set was largely the same and still consisted of 19 partitions. A plot of the diversity (vertical axis) and dispersion (horizontal axis) measures associated with these 19 partitions is presented in Figure 3. Table 5 provides the raw index measures for diversity and dispersion, as well as

normalized index values based on the range and the distance of each partition from the origin based on these normalized index values.

The line segments in Figure 3 illustrate the convex hull, which is comprised of five of the 19 partitions. The numbering of the partitions in Table 5 corresponds to the order of the points in Figure 1 when moving from right to left. The five partitions (1, 4, 12, 17, 19) are highlighted in bold font in Table 5. Partitions 4 and 17 on the convex hull appear promising because they correspond to large improvements in one criterion for a small sacrifice in the other. Partition 4 corresponds to a large improvement in diversity for a small sacrifice in dispersion, whereas Partition 17 enables a large gain in dispersion for a small loss in diversity.

To quantify the analysis of the Pareto set, we computed the range of diversity and dispersion values. The range for diversity was .10865 (i.e., from 40.09626 to 40.20491) and the range for dispersion was comparable at .08445 (i.e. from .07875 to .16320). We then used these ranges to compute, for each partition, the percentage of the range that was realized by their corresponding diversity and dispersion values. For example, consider partition 4, which has a diversity value (40.14005) that is 40.299% along the diversity range from 40.09626 to 40.20491 and a diversity value (.15911) that is 95.161% along the dispersion range from .07875 to .16320. Using these percentage values as coordinates, the distance of partition 4 from the origin is $\sqrt{.40299^2 + .95161^2} = 1.03343$.

Table 5. Dispersion and diversity values for the approximate Pareto efficient set for the ($\rho = 0$, $\delta = [-3, 3]$, $\bar{\alpha} = 1$, $G = 10$) test problem from Table 4, along with normalized measures based on range and the corresponding Euclidean distance from the origin

	Raw values		Range normalized		
Partition no.	Dispersion	Diversity	Dispersion	Diversity	Distance
1	**.16320**	**40.09626**	**1.00000**	**.00000**	**1.00000**
2	.16172	40.11235	.98251	.14811	0.99361
3	.16086	40.11584	.97234	.18021	0.98890
4	**.15911**	**40.14005**	**.95161**	**.40299**	**1.03343**
5	.15347	40.14291	.88483	.42939	0.98351
6	.15344	40.14398	.88452	.43917	0.98755
7	.14948	40.14438	.83761	.44292	0.94751
8	.14768	40.14650	.81625	.46243	0.93814
9	.14450	40.14954	.77860	.49038	0.92016
10	.14264	40.15026	.75654	.49698	0.90517
11	.14241	40.15287	.75388	.52104	0.91642
12	**.13637**	**40.16769**	**.68231**	**.65744**	**0.94751**
13	.12060	40.17461	.49559	.72107	0.87496
14	.11662	40.17499	.44846	.72460	0.85215
15	.10242	40.18039	.28028	.77432	0.82349
16	.10060	40.19454	.25881	.90451	0.94081
17	**.09560**	**40.20240**	**.19961**	**.97689**	**0.99708**
18	.08497	40.20396	.07370	.99129	0.99403
19	**.07875**	**40.20491**	**.00000**	**1.00000**	**1.00000**
Range	.08445	.10865			

Note. The rows highlighted in bold correspond to the partitions on the convex hull in Table 3. The "distance" column contains the distance of the partition from the origin when considering the range normalized values of dispersion and diversity.

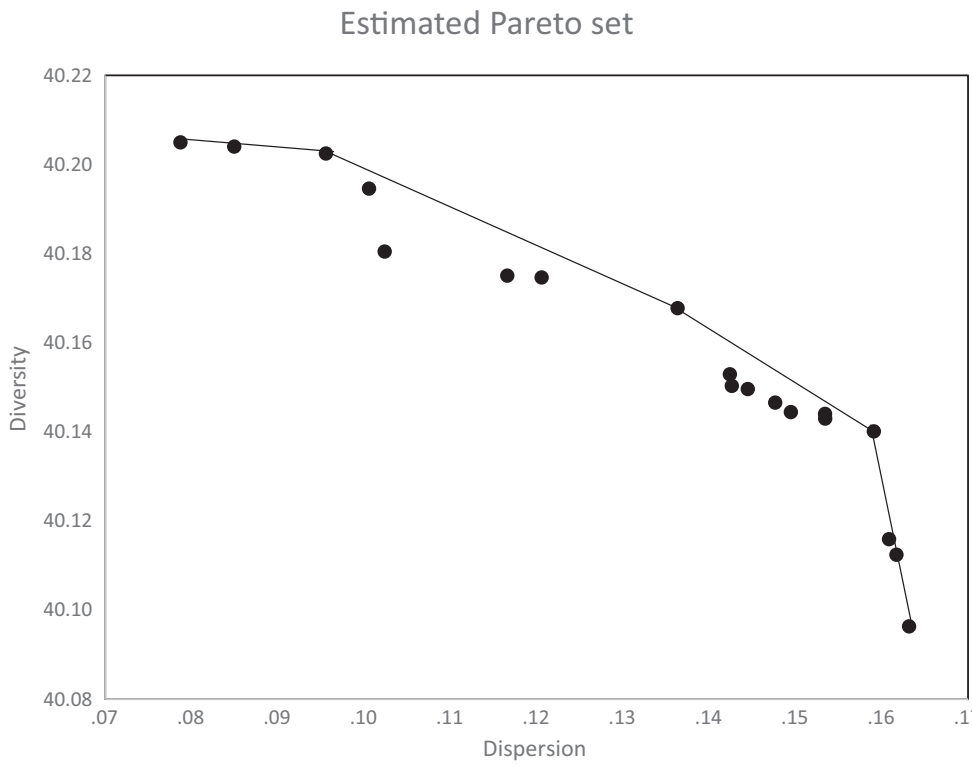

Figure 3. A plot of the Pareto efficient set for test problem ($\rho = 0$, $\delta = [-3, 3]$, $\bar{\alpha} = 1$, $G = 10$).

Notice that partition 4 is the greatest distance from the origin based on this measure. Partitions 1 and 19 are tied for second best with a distance from the origin of 1 based on the percentage measures. Partition 17 is next, with a distance of 0.99708.

Using the plot and the computed distances from the origin based on the range measures, we can provide some general guidelines. For a decision-maker who has a strong preference for dispersion, yet is willing to sacrifice a little dispersion for a big gain in diversity, the plot reveals that partition 4 is an excellent choice. This is supported by the fact that it has the largest distance from the origin in Table 5. For a decision-maker who has a strong preference for diversity, but is willing to sacrifice a little diversity for a big gain in dispersion, the plot and distance information suggest that partition 17 is a good choice. Partition 12 would be a viable selection for a decision-maker seeking a partition that provides a more equitable trade-off of diversity and dispersion on the approximated Pareto set.

6. Conclusions

6.1. Summary

Anticlustering, a term originally coined by Späth (1986), refers to partitioning problems that have the goal of maximizing rather than minimizing heterogeneity within groups. Two types of anticlustering problems are popular. The first, and more common, type is maximum diversity partitioning, which emphasizes the maximization of within-group

sums of dissimilarities. The second type is maximum dispersion partitioning, which focuses on the maximization of within-group minimum dissimilarities. The premise of our paper is that diversity and dispersion are both important and, ideally, quantitative analysts should have information for both measures when faced with an anticlustering application. Accordingly, we propose a bicriterion combinatorial problem that incorporates both diversity and dispersion measures. Bicriterion pairwise interchange and iterated local search heuristics were developed to approximate the Pareto efficient set for this problem. The methods can be applied to dissimilarity data such as distances for a set of variables, paired-comparison judgements, grouping tasks, and other measures. In this paper, we used a motivating example related to test assembly to illustrate the method. We also compared the performances of the multistart pairwise interchange heuristic and the bicriterion iterated local search heuristic in a simulation study conducted within the framework of the multidimensional two-parameter logistic model (Reckase, 2009). The results of this study demonstrated that, when constrained to approximately the same computation time, the two methods performed comparably. However, in practice, it would generally be advisable to run iterated local search after multistart pairwise interchange (regardless of the number of restarts) to see if further improvement in the Pareto set can be achieved.

There are two major benefits associated with the bicriterion model and method that we have developed. The first benefit pertains to tie-breaking. There are often a vast number of partitions that produce the same globally optimal dispersion index value. Although these partitions are "tied" with respect to dispersion, they can differ greatly with respect to their diversity index values. Because it focuses on finding non-dominated partitions, the bicriterion algorithm will select a partition from the tied set that maximizes (or, at least nearly maximizes) diversity. The problem of ties among partitions having the same diversity index is much less severe; however, it does occur. Therefore, the bicriterion approach can also sometimes be useful for selecting a partition with maximum dispersion from a set of partitions that have the same diversity index.

The second major benefit is related to the evaluation of trade-offs between diversity and dispersion. In the absence of a bicriterion perspective, it is difficult to know whether the relationship between diversity and dispersion is concordant or discordant. A concordant relationship implies that there are partitions that provide optimal (or near-optimal) values for both measures simultaneously. As discordance increases, the size of the Pareto set tends to increase and there are a variety of possible partitions that represent trade-offs between the two criteria. The bicriterion approach formalizes the nature of these trade-offs and can often enable analysts to find partitions whereby small sacrifices in one of the criteria lead to huge improvements in the other.

6.2. Limitations and extensions

There are limitations of our bicriterion anticlustering approach that should be acknowledged. First, and foremost, the bicriterion pairwise interchange and iterated local search algorithms do not guarantee that the true Pareto efficient set is recovered, but only provide an approximation. One possible approach to obtaining solutions guaranteed to be members of the Pareto efficient set is to use some mixed integer linear programming or branch-and-bound methods to obtain solutions for different sets of weights, or to optimize one of the criteria subject to constraint thresholds for the other. Unfortunately, such approaches are often computationally infeasible. In any case, our goal in this paper was not to extend the state of the art of mathematical programming methods for

maximum diversity or maximum dispersion partitioning, but rather to address the importance of incorporating both diversity and dispersion in the anticlustering process.

A second limitation is that there might be variants of the anticlustering problem that would require modifications of the MBPI and BILS algorithms. For example, our analysis in this paper has been restricted to problems where the group sizes are assumed known, which is often a reasonable assumption in applications such as parallel test assembly. However, there could be circumstances where the number of items in each group is not fixed, but allowed to fall within some particular range. For such applications, it would be necessary to either replace or augment the pairwise interchange algorithm with a transfer algorithm that evaluated the relocation of each individual item from its current group to one of the other groups. Whereas a pairwise interchange preserves the same number of items in each group, a transfer decreases the number of items in the 'moved-from' group by one and increases the number of items in the 'moved-to' group by one.

References

Arabie, P., & Hubert, L. J. (1992). Combinatorial data analysis. *Annual Review of Psychology*, *43*, 169–203. https://doi.org/10.1146/annurev.ps.43.020192.001125

Armstrong, R. D., Jones, D. H., & Wu, I.-L. (1992). An automated test development of parallel tests from a seed test. *Psychometrika*, *57*, 271–288. https://doi.org/10.1007/BF02294509

Baker, K. R., & Powell, S. G. (2002). Methods for assigning students to groups: A study of alternate objective functions. *Journal of the Operational Research Society*, *53*, 397–404. https://doi.org/10.1057/palgrave.jors.2601307

Belov, D. (2008). Uniform test assembly. *Psychometrika*, *73*, 21–38. https://doi.org/10.1007/s11336-007-9025-0

Bhadury, J., Mighty, E. J., & Damar, H. (2000). Maximizing workforce diversity in project teams: A network flow approach. *Omega*, *28*, 143–153. https://doi.org/10.1016/S0305-0483(99)00037-7

Brusco, M. J. (2003). An enhanced branch-and-bound algorithm for a partitioning problem. *British Journal of Mathematical and Statistical Psychology*, *56*, 83–92. https://doi.org/10.1348/000711003321645359

Brusco, M. J. (2015). A bicriterion algorithm for the allocation of cross-trained workers based on operational and human-resource objectives. *European Journal of Operational Research*, *247*, 46–59. https://doi.org/10.1016/j.ejor.2015.06.009

Brusco, M. J. (2017). Partitioning methods for pruning the Pareto set with application to multiobjective allocation of a cross-trained workforce. *Computers & Industrial Engineering*, *111*, 29–38. https://doi.org/10.1016/j.cie.2017.06.035

Brusco, M. J., & Cradit, J. D. (2005). Bicriterion methods for partitioning dissimilarity matrices. *British Journal of Mathematical and Statistical Psychology*, *58*, 319–332. https://doi.org/10.1348/000711005X63890

Brusco, M., Doreian, P., Steinley, D., & Satornino, C. B. (2013). Multiobjective blockmodeling for social network analysis. *Psychometrika*, *78*, 498–525. https://doi.org/10.1007/s11336-012-9313-1

Brusco, M. J., Köhn, H.-F., & Steinley, D. (2015). An exact method for partitioning dichotomous items within the framework of the monotone homogeneity model. *Psychometrika*, *80*, 949–967. https://doi.org/10.1007/s11336-015-9459-8

Brusco, M. J., Shireman, E., Steinley, D., Brudvig, S., & Cradit, J. D. (2017). Gaussian model-based partitioning using iterated local search. *British Journal of Mathematical and Statistical Psychology*, *70*, 1–24. https://doi.org/10.1111/bmsp.12084

Brusco, M. J., & Stahl, S. (2001). Compact integer programming models for extracting subsets of stimuli from confusion matrices. *Psychometrika*, *66*, 405–419. https://doi.org/10.1007/BF02294442

Brusco, M. J., & Stahl, S. (2005). *Branch-and-bound applications in combinatorial data analysis*. New York, NY: Springer.

Brusco, M. J., & Steinley, D. (2006). Clustering, seriation, and subset extraction of confusion data. *Psychological Methods*, *11*, 271–286. https://doi.org/10.1037/1082-989X.11.3.271

Brusco, M. J., & Steinley, D. (2009). Cross-validation issues in multiobjective clustering. *British Journal of Mathematical and Statistical Psychology*, *62*, 349–368. https://doi.org/10.1348/000711008X304385

Brusco, M. J., & Steinley, D. (2012). A note on the estimation of the Pareto efficient set for multiobjective matrix permutation problems. *British Journal of Mathematical and Statistical Psychology*, *65*, 145–162. https://doi.org/10.1111/j.2044-8317.2011.02021.x

Brusco, M. J., & Steinley, D. (2014). Model selection for minimum-diameter partitioning. *British Journal of Mathematical and Statistical Psychology*, *67*, 471–495. https://doi.org/10.1111/bjmsp.12029

Cerny, V. (1985). A thermodynamical approach to the traveling salesman problem. *Journal of Optimization Theory and Applications*, *45*, 41–51. https://doi.org/10.1007/BF00940812

Colarelli, S. M., & Boos, A. L. (1992). Sociometric and ability-based assignment to work groups: Some implications for personnel selection. *Journal of Organizational Behavior*, *13*, 187–196. https://doi.org/10.1002/job.4030130208

Coldren, J. T., & Haaf, R. (2000). Asymmetries in infants' attention to the presence or absence of features. *Journal of Genetic Psychology*, *161*, 420–434. https://doi.org/10.1080/00221320009596722

Compton, D. L. (2003). The influence of item composition on RAN letter performance in first-grade children. *Journal of Special Education*, *37*, 81–94. https://doi.org/10.1037/0022-0663.95.2.225

de la Torre, J., & Douglas, J. (2004). Higher-order latent-trait models for cognitive diagnosis. *Psychometrika*, *69*, 333–353. https://doi.org/10.1007/BF02295640

Delattre, M., & Hansen, P. (1980). Bicriterion cluster analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *2*, 277–291. https://doi.org/10.1109/TPAMI.1980.4767027

Derryberry, D. (1991). The immediate effects of positive and negative feedback signals. *Journal of Personality and Social Psychology*, *61*, 267–278. https://doi.org/10.1037/0022-3514.61.2.267

Ehrgott, M. (2005). *Multicriteria optimization* (2nd ed.). Berlin, Germany: Springer.

Ferligoj, A., & Batagelj, V. (1992). Direct multicriteria clustering algorithms. *Journal of Classification*, *9*, 43–61. https://doi.org/10.1007/BF02618467

Fernández, E., Kalcsics, J., & Nickel, S. (2013). The maximum dispersion problem. *Omega*, *41*, 721–730. https://doi.org/10.1016/j.omega.2012.09.005

Fernández, E., Kalcsics, J., Nickel, S., & Ríos-Mercado, R. (2010). A novel maximum dispersion territory design model arising in the implementation of the WEEE-directive. *Journal of the Operational Research Society*, *61*, 503–514. https://doi.org/10.1057/jors.2009.70

Gallego, M., Laguna, M., Martí, R., & Duarte, A. (2013). Tabu search with strategic oscillation for the maximally diverse grouping problem. *Journal of the Operational Research Society*, *64*, 724–734. https://doi.org/10.1057/jors.2012.66

Gierl, M. J., Daniels, L., & Zhang, X. (2017). Creating parallel forms to support on-demand testing for undergraduate students in psychology. *Journal of Measurement and Evaluation in Education and Psychology*, *8*, 288–302. https://doi.org/10.21031/epod.305350

Gillet, V. J., Wild, D. J., Willett, P., & Bradshaw, J. (1998). Similarity and dissimilarity methods for processing chemical structure databases. *Computer Journal*, *41*, 547–558. https://doi.org/10.1093/comjnl/41.8.547

Glover, F., Hersh, G., & McMillan, C. (1977). *Selecting subsets of maximum diversity* (Management Science/Information Science Report No. 77-9). Boulder, CO: University of Colorado.

Glover, F., Kuo, C. C., & Dhir, K. S. (1998). Heuristic algorithms for the maximum diversity problem. *Journal of Information and Optimization Sciences*, *19*, 109–132. https://doi.org/10.1080/02522667.1998.10699366

Glover, F., & Laguna, M. (1993). Tabu search. In C. Reeves (Ed.), *Modern heuristic techniques for combinatorial problems* (pp. 70–141). Oxford, UK: Blackwell.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading, MA: Addison-Wesley.

Hübscher, R. (2010). Assigning students to groups using general and context-specific criteria. *IEEE Transactions on Learning Technologies*, *3*, 178–189. https://doi.org/10.1109/TLT.2010.17

Johnson, D. S., & McGeoch, L. A. (1997). The traveling salesman problem: A case study in local optimization. In E. H. L. Aarts & J. K. Lenstra (Eds.), *Local search in combinatorial optimization* (pp. 215–310). Chichester, UK: Wiley.

Kelly, P. (2007). Achieving desirable group-work outcomes through the group allocation process. *Team Performance Management*, *14*, 22–38. https://doi.org/10.1108/13527590810860186

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*, 671–680. https://doi.org/10.1126/science.220.4598.671

Lees, M. C., & Neufeld, R. W. J. (1999). Decision-theoretic aspects of stress arousal and coping personality. *Journal of Personality and Social Psychology*, *77*, 185–208. https://doi.org/10.1037/0022-3514.77.1.185

Lourenço, H. R., Martin, O. C., & Stützle, T. (2003). Iterated local search. In F. Glover & G. A. Kochenberger (Eds.), *Handbook of metaheuristics* (International Series in Operations Research & Management Science, Vol. *57*, pp. 320–352). Boston, MA: Kluwer Academic.

Lourenço, H. R., Martin, O. C., & Stützle, T. (2010). Iterated local search: Framework and applications. In M. Gendreau, & J.-Y. Potvin (Eds.), *Handbook of metaheuristics* (2nd ed., International Series in Operations Research & Management Science, Vol. *146*, pp. 363–397). New York, NY: Springer.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In L. M. LeCam, & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. *1*, pp. 281–297). Berkeley, CA: University of California Press.

McQuitty, L. L. (1960). Hierarchical linkage analysis for the isolation of types. *Educational and Psychological Measurement*, *20*, 55–67. https://doi.org/10.1177/001316446002000106

Mingers, J., & O'Brien, F. (1995). Creating student groups with similar characteristics: A heuristic approach. *Omega*, *23*, 313–321. https://doi.org/10.1016/0305-0483(95)00014-F

Miyaji, I., Ohno, K., & Mine, H. (1988). Solution method for partitioning students into groups. *European Journal of Operational Research*, *33*, 82–90. https://doi.org/10.1016/0377-2217(88)90256-1

O'Brien, F., & Mingers, J. (1997). A heuristic algorithm for the equitable partitioning problem. *Omega*, *25*, 215–223. https://doi.org/10.1016/S0305-0483(96)00046-1

Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: Springer.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society. Retrieved from https://www.psychometricsociety.org/sites/default/files/pdf/MN17.pdf

Sanders, P. F., & Verschoor, A. J. (1998). Parallel test construction using classic item parameters. *Applied Psychological Measurement*, *22*, 212–223. https://doi.org/10.1177/01466216980223002

Schreiber, G. R., & Martin, O. C. (1999). Cut size statistics of graph bisection heuristics. *SIAM Journal on Optimization*, *10*, 231–251. https://doi.org/10.1137/S1052623497321523

Späth, H. (1986). Anticlustering: Maximizing the variance criterion. *Control and Cybernetics*, *15*, 213–218.

Steinley, D. (2006). *K*-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, *59*, 1–34. https://doi.org/10.1348/000711005X48266

Steinley, D., & Brusco, M. J. (2011). Evaluating mixture-modeling for clustering: Recommendations and cautions. *Psychological Methods*, *16*, 63–79. https://doi.org/10.1037/a0022673

Straat, J. H., Van der Ark, L. A., & Sijtsma, K. (2013). Comparing optimization algorithms for item selection in Mokken scale analysis. *Journal of Classification*, *30*, 75–99. https://doi.org/10.1007/s00357-013-9122-y

Tatsuoka, K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Safto (Eds.), *Monitoring skills and knowledge acquisition* (pp. 453–488). Hillsdale, NJ: Erlbaum.

Van der Linden, W. J. (2005). *Linear models for optimal test design*. New York, NY: Springer.

Wang, S., Zheng, Y., Zheng, C., Su, Y.-H., & Li, P. (2016). An automated test assembly design for a large-scale Chinese proficiency test. *Applied Psychological Measurement*, *40*, 233–237. https://doi.org/10.1177/0146621616628503

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236–244. https://doi.org/10.2307/2282967

Weitz, R. R., & Lakshminarayanan, S. (1997). An empirical comparison of heuristic and graph theoretic methods for creating maximally diverse groups, VLSI design, and exam scheduling. *Omega*, *25*, 473–482. https://doi.org/10.1016/S0305-0483(97)00007-8

Weitz, R. R., & Lakshminarayanan, S. (1998). An empirical comparison of heuristic methods for creating maximally diverse groups. *Journal of the Operational Research Society*, *49*, 635–646. https://doi.org/10.1057/palgrave.jors.2600510

Received 26 January 2018; revised version received 22 June 2019