Olivier Toubia is a graduate student at the Marketing Group and the Operations Research Center, Massachusetts Institute of Technology, E56-345, 38 Memorial Drive, Cambridge, MA 02142, (617) 253-0159, fax (617) 253-7597, toubia@mit.edu.

Duncan I. Simester is an Associate Professor, Sloan School of Management, Massachusetts Institute of Technology, E56-305, 38 Memorial Drive, Cambridge, MA 02142, (617) 258-0679, fax (617) 253-7597, simester@mit.edu.

John R. Hauser is the Kirin Professor of Marketing, Sloan School of Management, Massachusetts Institute of Technology, E56-314, 38 Memorial Drive, Cambridge, MA 02142, (617) 253-2929, fax (617) 253-7597, jhauser@mit.edu.

This research was supported by the Sloan School of Management and the Center for Innovation in Product Development at M.I.T. This paper may be downloaded from mitsloan.mit.edu/virtual_customer as of February 7, 2001. That website also contains (1) open source code to implement the methods described in this paper, (2) open source code for the simulations described in this paper, (3) demonstrations of web-based questionnaires based on the methods in this paper, and (4) related papers on web-based interviewing methods. All authors contributed fully and synergistically to this paper. We wish to thank Limor Weisberg for creating the graphics in this paper.

The sample sizes in the simulations in this draft are in the process of being increased to 1,000 respondents for most figures and twenty sets of 200 for the aggregate analyses. This should make the lines in the graphs smoother and improve the aggregate analyses, but should not change the basic insights.
Fast Polyhedral Adaptive Conjoint Estimation

Abstract

Web-based customer panels and web-based multimedia capabilities offer the potential to get information from customers rapidly and iteratively based on virtual product profiles. However, web-based respondents are impatient and wear out more quickly. At the same time, in commercial applications, conjoint analysis is being used to screen large numbers of product features. Both of these trends are leading to a demand for conjoint analysis methods that provide reasonable estimates with fewer questions.

In this paper we propose and test new adaptive conjoint analysis methods that attempt to reduce respondent burden while simultaneously improving accuracy. We draw on recent “interior-point” developments in mathematical programming which enable us to quickly select those questions that narrow the range of feasible partworths as fast as possible. We then use recent centrality concepts (the analytic center) to estimate partworths. Not only are these methods efficient – they run with no noticeable delay in web-based questionnaires, but they have the potential to provide estimates of the partworths with fewer questions than extant methods.

After introducing these “polyhedral” algorithms we implement one such algorithm and test it with Monte Carlo simulation against benchmarks such as efficient (fixed) designs and Adaptive Conjoint Analysis (ACA). While no method dominates in all situations, the polyhedral algorithm appears to hold significant potential when (a) profile comparisons are more accurate than the self-explicated importances used in ACA, (b) when respondent wear out is a concern, and (c) when the product development and marketing teams wish to screen many features quickly. We also test a hybrid method that combines polyhedral question selection with ACA estimation and show that it, too, has the potential to improve predictions in many contexts. The algorithm we test helps to illustrate how polyhedral methods can be combined effectively and synergistically with the wide variety of existing conjoint analysis methods.

We close with suggestions on how polyhedral algorithms can be used in other preference measurement contexts (e.g., choice-based conjoint analysis) and other marketing problems.
**A Changing Landscape**

Conjoint analysis continues to be one of the most widely used quantitative research methods in marketing (e.g., Mahajan and Wind 1992) with well over 200 commercial applications per year in such sectors as consumer packaged goods, computers, pharmaceutical products, electric vehicles, MBA job choice, hotel design, performing arts, health care, and a wide variety of business-to-business products (Cattin and Wittink 1982, Currim, Weinberg, and Wittink 1981, Green and Krieger 1992, Parker and Srinivasan 1976, Urban, Weinberg, and Hauser 1996, Wind, Green, Shifflet, and Scarbrough 1989, Wittink and Cattin 1989). It has proven both reliable and valid in a variety of circumstances and has been adapted to a wide variety of uses, including product design, product line design, and optimal positioning. It is safe to say that there is strong academic and industrial interest in the continued improvement of conjoint methods.

However, with the advent of new communications and information technologies, the landscape is changing and changing rapidly. These technologies are impacting conjoint analysis in at least two ways. The first is the applications’ demands. Product development (PD) is changing to a more dispersed, distributed, and global activity with cross-functional teams spread throughout the world (Wallace, Abrahamson, Senin, and Sferro 2000). These changes place a premium on rapidity as many firms now focus on time-to-market as a key competitive advantage (Anthony and Mckay 1992, Cooper and Kleinschmidt 1994, Cusumano and Yoffie 1998, Griffin 1997, Ittner and Larcher 1997, McGrath 1996, Smith and Reinertsen 1998). In today’s world many PD teams now use spiral processes, rather than stage-gate processes, in which PD cycles through design and customer testing many times before the final product is launched and in which customer input is linked to engineering requirements with just-in-time or turbo processes (Cusumano and Selby 1995, McGrath 1996, Smith and Reinertsen 1998, Tessler and Klein 1993). As a result, PD teams routinely demand customer preference information more frequently and more rapidly.

Conjoint analysis is also being used to address larger and more complex problems. In a classic example, Wind, et. al. (1989) used hybrid conjoint analysis to design a service with 50 attributes. In the last ten years this trend has accelerated with problems in which PD teams face decisions on almost 10,000 separate parts (Eppinger 1998, Eppinger, Whitney, Smith, and Gebala 1994, Ulrich and Eppinger 2000). Although we are unlikely to see a conjoint analysis with thousands of variables, we have heard of

---

commercial applications with over 120 parameters. Applications now routinely require a higher information-to-question ratio.

At the same time, new technologies are providing new capabilities. With the advent of the Internet and the worldwide web, researchers are developing highly visual and interactive stimuli, which demonstrate products, product features, and product use with vivid multimedia tools (Dahan and Srinivasan 2000, Dahan and Hauser 2000). Furthermore, new on-line panels can reach customers more quickly and at a fraction of the cost of previous recruiting procedures. For example, NFO Worldwide, Inc. is developing a balanced panel of over 500,000 web-enabled respondents, DMS, Inc., a subsidiary of AOL, uses “Opinion Place” to recruit respondents dynamically and claims to be interviewing over 1 million respondents per year, Knowledge Networks has recruited 100,000 Internet enabled respondents with random digit dialing methods and provides them with web access if they do not already have it, Greenfield Online, Inc. has an online panel of 3 million respondents, and Harris Interactive, Inc. has an online panel of 6.5 million respondents (Buckman 2000). General Mills now claims to do 60% of their market research on-line (Marketing News, 2000). If these developments can be exploited effectively, PD teams will use conjoint analysis more often and earlier in the PD process.

Computing power and Internet bandwidth also mean that more complex computations can be used to present virtual product profiles that are realistic and customized. With today’s bandwidth we can create these profiles on the server and send them to a local Internet-connected PC in the customers’ home or place of business.

However, communication and information technology is a double-edged sword. Conjoint analysis has long been concerned with respondent burden and researchers have a long track record of developing methods to reduce this burden (see next section for examples). Web-based interviewing requires even more attention to respondent burden. In central location interviewing (e.g., Sawtooth 1996), respondents were remarkably tolerant of multiple questions. The literature (and our experience) indicates that respondents who complete a survey on the web are much less tolerant. This may reflect the culture of the web, the ease with which the respondent can switch tasks, or perhaps, less commitment by over-tasked respondents (Black, Harmon, MacElroy, Payne, Willke 2000). In our experience, interest and reliability drop off very rapidly after eight to twelve repetitive conjoint analysis tasks. The most direct evidence is seen in response rates. For example, in a commercial application to the design of a new I-Zone camera, the yield from the home-based web survey was less than half that from the same survey administered in a central (mall) location (McArdle 2000), although the proportion of valid and usable responses was high under both approaches. We expect the move to web-based surveys to create an even greater demand for conjoint analysis methods that gather more information with fewer questions.
New Polyhedral Methods for Conjoint Analysis

In this paper we propose and test new adaptive conjoint analysis methods that attempt to reduce respondent burden while simultaneously improving accuracy. Because the methods make full use of high-speed computations and adaptive, customized local web pages, they are ideally suited for Internet panels and for use in the new PD processes. Specifically we interpret the problem of selecting questions and estimating parameters as a mathematical program and estimate the solution to the program using recent developments based on the interior points of polyhedra. These techniques provide the potential for accurate estimates of partial utilities from far fewer questions than required by extant methods.

Our goals are two-fold. First, we investigate whether polyhedral methods have the potential to enhance the effectiveness of existing conjoint methods. Second, by focusing on a widely studied marketing problem we hope to illustrate the recent advances in mathematical programming and encourage their applications in the marketing literature.

Because the methods are new and adopt a different estimation philosophy, we use Monte Carlo experiments to explore the properties of the proposed polyhedral methods. The Monte Carlo experiments explore the conditions under which polyhedral methods are likely to do better or worse than extant methods. We have some reason for optimism. Conjoint analysis methods such as Linmap (Srinivasan and Shocker 1973a, 1973b) successfully use classical linear programming to obtain estimates by placing constraints on the feasible set of parameters.

As representative of the breadth of conjoint analysis methods, we compare the proposed method to two benchmarks: (1) efficient fixed designs in which the questions are chosen to yield orthogonality and balance and (2) Sawtooth’s adaptive conjoint analysis (ACA) in which the questions are adapted based on respondents’ answers. We further explore five issues that are relevant to web-based applications. Specifically:

1. Accuracy vs. the number of questions. Because web-based conjoint surveys place a premium on a small number of questions, we explore how rapidly the estimates converge to their true values as the number of questions increases. The simulations highlight situations in which the polyhedral methods obtain the same accuracy in five-to-ten fewer questions than the benchmark methods – this may be the difference between an enjoyable questionnaire and one that the respondent sees as a burden.

2. Self-explicated questions. Both ACA and hybrid conjoint analysis (HCA) use self-explicated questions in which the respondent states the importance of features directly. These importances are then updated with data from the comparison or ranking of product profiles (revealed importances). We explore how the various methods perform based on the relative accuracy of answers to the self-explicated vs. the revealed importance questions. The results suggest that under many reasonable situations the
polyhedral methods can skip the self-explicated stage and still perform as well as ACA. This has the potential to dramatically reduce the length of a web-based survey.

3. Question selection vs. partworth estimation. Because the polyhedral methods can be used for both question selection and partworth estimation, we test hybrid methods that facilitate comparison of each component (question selection and estimation) with traditional procedures and which suggest how polyhedral methods complement existing procedures.

4. Respondent wear out. We explore what happens if respondents wear out and their attention degrades as the number of questions increases.

5. Individual vs. population estimates. The polyhedral methods (and many conjoint analysis methods) seek to provide estimates of customer preferences that vary by respondent. However, such heterogeneous estimates should not compromise the accuracy of population estimates. Thus, we also compare the alternative methods on their abilities to estimate population averages.

The paper is structured as follows. We begin with a review of conjoint analysis (fixed designs) and ACA. For perspective, we briefly review previous attempts to simplify the task and improve accuracy. We then describe the polyhedral methods and introduce the interior-point algorithms. (We provide the detailed mathematics in an appendix.) We follow with the design, results, and interpretation of the Monte Carlo experiments and discuss empirical issues implied by these experiments. We close with a discussion and comparison of the applicability of the polyhedral methods to conjoint analysis and to other marketing problems.

**Conjoint Analysis: Methods to Reduce Respondent Burden**

As early as 1978, Carmone, Green, and Jain (p. 300) found that most applications demanded a dozen or more features, but that it was difficult for customers to rank more than a dozen profiles. Many researchers have documented that the respondents’ task can be burdensome and have suggested that accuracy degrades as the number of questions increases (Bateson, Reibstein, and Boulding, 1987, Green, Carroll, and Goldberg 1981, p. 34, Green, Goldberg, and Montemayor 1981, p. 337, Huber, Wittink, Fiedler, and Miller 1993, Lenk, DeSarbo, Green, and Young 1996, p. 183, Malhotra 1982, 1986, p. 33, Moore and Semenik 1988, Srinivasan and Park 1997, p. 286). The academic and commercial response was immediate and continues today. When appropriate, efficient experimental designs are used so that the respondent need consider only a small fraction of all possible product profiles (Addelman 1962, Kuhfeld, Tobias, and Garratt 1994). Tradeoff analysis presents respondents with two attributes at a time and has them evaluate the reduced sets (Jain, Acito, Malhotra, and Mahajan 1979, Johnson 1974, Segal 1982). Two stages can be introduced in which respondents eliminate unacceptable products, unacceptable attributes, or use prior sorting tasks to simplify the evaluation task (Acito and Jain 1980, Green, Krieger,


Each of these methods, when used carefully and responsibly, reduces the respondents’ burden and is feasible in large commercial applications. We do not propose to replace any of these research streams, but, rather, provide new capabilities that enhance these methods.

**Adaptive Conjoint Analysis – a Brief Exposition**

Before we provide details on the polyhedral methods, we briefly review the most widely used and studied adaptive method. ACA is of great interest to both academics and industry. For example, in 1991 Green, Krieger and Agarwal (p. 215) stated that “in the short span of five years, Sawtooth Software’s Adaptive Conjoint Analysis (ACA) has become one of the industry’s most popular software packages for collecting and analyzing conjoint data,” and go on to cite a number of academic papers on ACA. Although accuracy claims vary, ACA appears to predict reasonably well in many situations and is the primary adaptive method to estimate individual-level parameters (Johnson 1991, Orme 1999). ACA uses four sections, including:

1. Unacceptability task. The respondent is asked to indicate unacceptable levels which are subsequently deleted from the survey tasks. However, this step is often skipped because respondents can be too quick to eliminate levels (Sawtooth 1996).

2. Self-explicated task. If the rank-order preference for levels of a feature are not known a priori (e.g., color), the respondent ranks the levels of a feature. The respondent then states the relative
importance (on a 4-point scale) of improving the product from one feature level to another (e.g., adding automatic film ejection to an instant camera).

3. Paired-comparison task. The respondent states his or her preference for pairs of partial profiles in which two or three features vary (and all else is assumed equal). This is the adaptive stage because the specific pairs are chosen by an heuristic algorithm designed to increase the incremental information yielded by the next response. In particular, based on “current” estimates of partworths, the respondent is shown pairs of profiles that are nearly equal in utility. Constraints ensure the overall design is nearly orthogonal (features and levels are presented independently) and balanced (features and levels appear with near equal frequency).

4. Calibration concepts. Full profiles are presented to the respondents who evaluate them on a purchase intention scale.

The estimation (updating) of the partworths occurs after each paired comparison question. If \( \hat{\mathbf{s}} \) is the vector of prior self-explicated partworth estimates, \( X \) is the design matrix for the pairs (chosen adaptively), \( \mathbf{a} \) is the vector of respondent answers to the pairs, and \( I \) is the identity matrix, then the updated estimates, \( \tilde{\mathbf{u}} \), are obtained by minimizing the following least-squares norm.

\[
\left\| \begin{bmatrix} I \\ X \end{bmatrix} \tilde{\mathbf{u}} - \begin{bmatrix} \hat{\mathbf{s}} \\ \mathbf{a} \end{bmatrix} \right\|_2^2
\]

Early versions of ACA used Equation 1 directly, however, later versions decompose the estimates into components due to the priors, the paired responses, and predicted paired responses. Little or no weight is placed on the last component. Finally, after the last paired-comparison question, the calibration concepts are regressed (logit transformation) on the components of utility due to the pairs and due to the self-explicated task. The resulting weights are used in forecasting.

We now describe the polyhedral question selection and partworth estimation procedures. We begin with a conceptual description that highlights the geometry of the parameter space and then introduce the interior-point methods based on the “analytic center” of a polyhedron.

**Information and the Polyhedral Feasible Sets**

In this section we illustrate the concepts with a 3-parameter problem because 3-dimensional spaces are easy to visualize and explain. The methods generalize easily to realistic problems that contain ten, twenty, or even one hundred parameters. Indeed, relative to existing methods, the polyhedral methods are most useful for large numbers of parameters. By a parameter, we refer to a partworth that needs to be estimated. For example, twenty two-level features require ten parameters because, without loss of generality, we can set to zero the partworth of the least preferred feature. Similarly, ten three-level features also require twenty parameters. Interactions among features require still more parameters.
We focus on paired comparison tasks in which the respondent is shown two profiles and asked to provide an interval-scaled paired-comparison (metric) preference rating. We choose this focus because (1) this task enables a direct comparison to ACA, (2) this task is common in computer-aided interviewing, (3) it has proven reliable in previous studies (Reibstein, Bateson, and Boulding 1988), and (4) its use in practice and in the literature is exceeded only by the full profile task (see Cattin and Wittink 1982; and Wittink and Cattin 1989 for applications surveys). It is also consistent with growing evidence that carefully collected metric data provide valid and reliable parameter estimates (Carmone, Green, and Jain 1978, Currim, Weinberg, and Wittink 1981, Hauser and Shugan 1980, Hauser and Urban 1979, Huber 1975, Leigh, MacKay, and Summers 1984, Malhotra 1986, Srinivasan and Park 1997, and Wittink and Cattin 1981). We feel this is a practical decision for a first test of polyhedral algorithms. Polyhedral methods can also be developed for other respondent tasks including tasks where the respondent simply chooses among the pairs or chooses from a set of profiles.

Suppose that we have three features of a camera – light control (automatic or one-step), picture size (postage stamp or 3" square), and focusing (automatic or manual). If we scale the least desirable level of each feature to zero we have three non-negative parameters to estimate, \( u_1, u_2, \) and \( u_3 \), reflecting the additional utility (partworth) associated with the most desirable level of each feature.\(^2\) If, without loss of generality, we impose a constraint that the sum of the parameters does not exceed 100, then the feasible region for the parameters is a 3-dimensional bounded polyhedron occupying all of the space between the origin and the plane \( u_1 + u_2 + u_3 = 100 \).\(^3\) See example in Figure 1a. Each of the points in this space represents a feasible set of partial utilities given the boundary constraints.

Suppose that we ask the respondent to evaluate a pair of profiles that vary on one or more features and the respondent says (without error) (1) that he or she prefers profile \( C_1 \) to profile \( C_2 \) and (2) provides a rating, \( a \), to indicate his or her preference. This introduces an equality constraint that the utility associated with profile \( C_1 \) exceeds the utility of \( C_2 \) by an amount equal to the rating. If we define \( \bar{u} = (u_1, u_2, u_3)^T \) as the 3×1 vector of parameters, \( \bar{z}_l \) as the 1×3 vector of product features for the left profile, and \( \bar{z}_r \) as the 1×3 vector of product features for the right profile, then, for additive utility, this equality constraint can be written as \( \bar{z}_l \bar{u} - \bar{z}_r \bar{u} = a \). With one question, the vector, \( \bar{u} \), can not be determined uniquely. However, we can characterize what we have learned from this question and answer.

\(^2\) In this example, we assume preferential independence which implies an additive utility function. We can handle interactions by relabeling features. For example, a 2x2 interaction between two features is equivalent to one four-level feature. We hold this convention throughout the paper.

\(^3\) We might also impose a constraint by scaling the largest to 100, in which case the feasible region would be a cube. Also in Figure 1a, the hyperplane could intersect the feasible polyhedron in a single point – a degenerate case handled easily.
FAST POLYHEDRAL ADAPTIVE CONJOINT ESTIMATION

Figure 1. Respondent’s Answers Affect the Feasible Region

Specifically, if we define \( \bar{x} = \bar{z}_i - \bar{z}_r \), then \( \bar{x} \) is a 1×3 vector describing the difference between the two profiles in the question. Thus, \( \bar{x} \bar{u} = a \) defines a hyperplane through the polyhedron in Figure 1a. The only feasible values of \( \bar{u} \) are those that are in the intersection of this hyperplane and the polyhedron. The new feasible set is also a polyhedron, but it is reduced by one dimension (2-dimensions rather than 3-dimensions). Because smaller polyhedra mean fewer parameter values are feasible, questions that reduce the size of the initial polyhedron as fast as possible lead to more precise estimates of the parameters.

However, in any real problem we expect the respondent’s answer to contain error. We can model this error as a probability density function over the parameter space (as in standard statistical inference). Alternatively, we can incorporate imprecision in a response by treating the equality constraint \( \bar{x} \bar{u} = a \) as a set of two inequality constraints: \( a - \delta \leq \bar{x} \bar{u} \leq a + \delta \). In this case, the hyperplane defined by the question-answer pair has “width.” The intersection of the initial polyhedron and the “fat” hyperplane is now a three-dimensional polyhedron as illustrated in Figure 1b. Naturally, we can ask more questions. Each question, if asked carefully, will result in a hyperplane that intersects a polyhedron resulting in a smaller polyhedron – a “thin” region in Figure 1a or a “fat” region in Figure 1b. Each new question-answer pair slices the polyhedron in Figure 1a or 1b yielding more precise estimates of the parameter vector \( \bar{u} \).

Also, we can easily incorporate prior information about the parameters by imposing constraints on the parameter space. For example, if \( u_m \) and \( u_h \) are the medium and high levels, respectively, of a feature, then we can impose the constraint \( u_m \leq u_h \) on the polyhedron. Previous research suggests that these types of constraints enhance estimation (Johnson 1999, Srinivasan and Shocker 1973a, 1973b). We
now example question selection by dealing first with the case in which subjects respond without error (Figure 1a). We then describe how to modify the algorithm to handle error as in Figure 1b.

**Question Selection**

The question selection task describes the design of the profiles that respondents are asked to compare. ACA selects profiles so that they are nearly equal in utility (coupled with balancing criteria) and similar methods have been proposed for selecting profiles in CBC (Huber and Zwerina 1996). With polyhedral estimation, we select the question that is likely to reduce the size of the feasible set the fastest.

Consider for a moment a 20-dimensional problem (without errors in the answers). As in Figure 1a, a question-based constraint reduces the dimensionality by one. That is, the first question reduces a 20-dimensional set to a 19-dimensional set; the next question reduces this set to an 18-dimensional set and so on until the twelfth question which reduces a 9-dimensional set to an 8-dimensional set (8 dimensions = 20 parameters – 12 questions). Without further restriction, the feasible parameters are generally not unique – any point in the 8-dimensional polyhedron is still feasible. However, the 8-dimensional set might be quite small and we might have a very good idea of the partworths. For example, the first twelve questions might be enough to tell us that some features, say 3, 7, 13, and 19, have large partworths and some features, say features 2, 8, 11, 17, and 20, have very small partworths. If this holds across respondents then, during an early phase of a stage-gate or spiral PD process, the PD team might feel they have enough information to focus on these key features.

Although the polyhedral algorithm is most effective in high dimensional spaces, it is hard to visualize 20-dimensional polyhedra. Instead, to illustrate the polyhedral selection criteria, we assume that we have already asked enough questions such that the remaining feasible set is easy to visualize. Specifically, by generalizing our notation slightly to \( q \) questions and \( p \) parameters, we define \( \vec{a} \) as the \( q \times 1 \) vector of answers and \( X \) as the \( q \times p \) matrix with rows equal to \( \vec{x} \) for each question. (Recall that \( \vec{x} \) is a \( 1 \times p \) vector.) Then the respondent’s answers to the first \( q \) questions define a \((p-q)\)-dimensional hyperplane given by the equation \( X\vec{u} = \vec{a} \). This hyperplane intersects the initial \( p \)-dimensional polyhedron to give us a \( p-q \) dimensional polyhedron. In the example of \( p=20 \) parameters and \( q=18 \) questions, the result is a 2-dimensional polyhedron that is easy to visualize. One such 2-dimensional polyhedron is illustrated by the dark region in Figure 2a.

Our task is now to select questions such that we reduce the 2-dimensional polyhedron as fast as possible. Mathematically, we select a new question vector, \( \vec{x} \), and the respondent answers this question with a new rating, \( a \). We add the new question vector as the last row of the question matrix and we add the new answer as the last row of the answer vector. While everything is really happening in \( p \)-dimensional space, the net result is that the new hyperplane will intersect the 2-dimensional polyhedron in
a line segment (i.e., a 1-dimensional polyhedron). The slope of the line will be determined by \( \bar{x} \) and the intercept by \( a \). We illustrate two potential question-answer pairs in Figure 2a. The slope of the line is determined by the question, the specific line by the answer, and the remaining feasible set by the line segment within the polyhedron. In Figure 2a, one of the question-answer pairs \((\bar{x}, a)\) reduces the feasible set more rapidly than the other question-answer pair \((\bar{x}', a')\). Figure 2b repeats a question-answer pair \((\bar{x}, a)\) and illustrates an alternative answer to the same question \((\bar{x}', a'')\).

**Figure 2. Choice of Question (2-dimensional slice)**

If the polyhedron is elongated as in Figure 2, then, in most cases, the questions that will result in small remaining feasible sets are those that imply line segments that are perpendicular to the longest “axis” of the polyhedron. Also, because the longest “axis” is in some sense a bigger target, it is more likely that the respondent’s answer will select a hyperplane that intersects the polyhedron. From analytical geometry we know that hyperplanes (line segments in Figure 2) are perpendicular to their defining vectors \((\bar{x})\), thus, we can reduce the feasible set as fast as possible (and make it more likely that answers are feasible) if we choose question vectors that are parallel to the longest “axis” of the polyhedron. In Figure 2b, the two line segments are perpendicular to this “axis.”

If we can develop an algorithm that works in any \(p\)-dimensional space, then we can generalize this intuition to any question, \(q\), such that \(q \leq p\). (We address later the cases where the respondent’s
answers contain error and where \( q > p \).) After receiving answers to the first \( q \) questions, we find the longest vector of the \((p-q)\)-dimensional polyhedron that describes the set of feasible parameter values. We then ask the question based on a vector that is parallel to this “axis.” The respondent’s answer creates a hyperplane that intersects the polyhedron to produce a new polyhedron that we expect to be as small as possible. Later in the paper we use Monte Carlo simulation to determine if this question-selection method produces reasonable estimates of the unknown parameters.

**Intermediate Estimates of Part-worths and Updates to those Estimates**

Polyhedral geometry also gives us a means to estimate the parameter vector, \( \hat{u} \), when \( q \leq p \). Recall that, after question \( q \), any point in the remaining polyhedron is consistent with the answers the respondent has provided. If we impose a diffuse prior that any feasible point is equally likely, then we would like to select the point that minimizes the expected absolute error. This point is the center of the feasible polyhedron. The smaller the feasible set, either due to better question selection or more questions (higher \( q \)), the more precise the estimate. If there were no respondent errors, then the estimate would converge to its true value when \( q = p \) (the feasible set becomes a single point, with zero dimensionality). For \( q > p \) the same point would remain feasible. As we will discuss, this changes when responses contain error.

Philosophically, the proposed polyhedral methods make maximum use of the information in the constraints and then take a central estimate based on what is still feasible. Carefully chosen questions shrink the feasible set rapidly. The centrality criterion has proven to be a surprisingly good estimate in a variety of engineering problems including, for example, finding the center of gravity of a solid. More generally, the centrality estimate is similar in some respects to the proven robustness of linear models, and in some cases, to the robustness of equally-weighted models (Dawes and Corrigan 1974, Einhorn 1971, Huber 1975, Moore and Semenik 1988, Srinivasan and Park 1997).

**Interior-point Algorithms and the Analytical Center of a Polyhedron**

To select questions and obtain intermediate estimates our heuristics require that we solve two non-trivial mathematical programs. First, we must find the longest “axis” of a polyhedron (to select the next question) and second, we must find the polyhedron’s midpoint (to provide a current estimate). If we were to define the longest “axis” of a polyhedron as the longest line segment in the polyhedron, then one method to find the longest “axis” would be to enumerate the vertices of the polyhedron and compute the distances between the vertices. However, solving this problem requires checking every boundary, which is computationally difficult. In practice, solving the problem would impose noticeable delays between questions. Also, the longest line segment in a polyhedron may not capture the concept of a longest “axis.” Finding the midpoint of the polyhedron is even more difficult and computationally demanding.
Fortunately, recent work in mathematical programming literature has led to extremely fast algorithms based on projections within the interior of polyhedrons (much of this work started with Karmarkar 1984). Interior-point algorithms are now routinely used to solve large problems and have spawned many theoretical and applied generalizations. One such generalization uses bounding ellipsoids. In 1985, Sonnevend demonstrated that a bounded polyhedron can be approximated by proportional ellipsoids, centered at the analytic center of the polyhedron. The analytic center is the point in the polyhedron that maximizes the geometric mean of the distances to the boundaries of the polyhedron. It is a central point that approximates the midpoint of the polyhedron and finds practical use in engineering and optimization. Furthermore, the axes of ellipsoids are well-defined and might capture the concept of an “axis” of a polyhedron. For more details see Freund (1993), Nesterov and Nemirovskii (1994), Sonnevend (1985a, 1985b), and Vaidja (1989).

**Figure 3. Bounding Ellipsoid and the Analytical Center (2-dimensions)**

We illustrate the proposed process in Figure 3 using the same two-dimensional polyhedron depicted in Figure 2. The algorithm proceeds in four steps. The mathematics are in the appendix; we provide the intuition here. We first find a point in the interior of the polyhedron. This is a standard linear programming (LP) problem and runs quickly. Then, following Freund (1993) we use Newton’s method to make the point more central. This is a well-formed problem and converges quickly to yield the analytic center as illustrated by the black dot in Figure 3. We next find a bounding ellipsoid based on a
formula that depends on the analytic center and the question-matrix, $X$. We then find the longest axis of the ellipsoid (diagonal line in Figure 3) with a quadratic program that has a closed-form solution. The next question, $\tilde{x}$, is based on the vector most nearly parallel to this axis.

Analytically, this algorithm works in higher dimensional spaces. For example, Figure 4 illustrates the algorithm when $p - q = 3$, that is, when we are trying to reduce a 3-dimensional feasible set to a 2-dimensional feasible set. Figure 4a illustrates a polyhedron based on the first $q$ questions. Figure 4b illustrates a bounding 3-dimensional ellipsoid, the longest axis of that ellipsoid, and the analytic center. The longest axis defines the question that is asked next which, in turn, defines the slope of the hyperplanes that intersect the polyhedron. One such hyperplane is shown in Figure 4c. The respondent’s answer selects the specific hyperplane and the intersection of the selected hyperplane and the 3-dimensional ellipsoid is a new 2-dimensional polyhedron, such as that in Figure 3. This process applies (in higher dimensions) from the first question to the $p^{th}$ question. For example, the first question implies a hyperplane that cuts the first $p$-dimensional polyhedron such that the intersection yields a $(p-1)$-dimensional polyhedron (Figure 1a for $p = 3$).

While the polyhedral algorithm exploits complicated geometric relationships, it runs extremely fast. For example, we have implemented the algorithm to select questions for a web-based conjoint analysis. Based an example with ten two-level features, respondents notice no delay in question selection nor any difference in speed versus a fixed design. For a demonstration see the website listed on the cover page of this paper. Because there is no guarantee that the polyhedral algorithm will work well with the conjoint task, we use simulation to examine how well the analytical center approximates the true parameters and how quickly ellipsoid-based questions reduce the feasible set of parameters.
Inconsistent Responses and Error-modeling with Polyhedral Estimation

Figures 2, 3, and 4 illustrate the geometry when respondents answer without error. However, real respondents are unlikely to be perfectly consistent. It is more likely that, for some \( q < p \), the respondent’s answers will be inconsistent and the polyhedron will become empty. That is, we will no longer be able to find any parameters, \( \tilde{u} \), that satisfy the equations that define the polyhedron, \( X\tilde{u} = \tilde{a} \). Thus, for real applications, we extend the polyhedral algorithm to address response errors. Specifically, so that the algorithm will continue to work when then answers become inconsistent, we adjust the polyhedron in a minimal way so that some parameter values are still feasible. We do this by modeling errors, \( \delta \), in the
respondent’s answers such that $\tilde{a} - \delta \leq X\tilde{u} \leq \tilde{a} + \delta$. Review Figure 1b. We then choose the minimum errors such that these constraints are satisfied. As it turns out, this is analogous to regression analysis, except that we substitute the “minmax” criterion for the least-squares criterion. This same modification covers the case of $q > p$.

To implement this policy we use a two-stage algorithm. In the first stage we treat the responses as if they occurred without error – the feasible polyhedron shrinks rapidly and the analytical center is a working estimate of the true parameters. However, as soon as the feasible set becomes empty, we adjust the constraints by adding or subtracting “errors,” where we choose the minimum errors, $\|\tilde{\delta}\|$, for which the feasible set is non-empty. The analytical center of the new polyhedron becomes the working estimate and $\tilde{\delta}$ becomes an index of response error. We also switch from the polyhedral question-selection procedure to the standard method of balancing features and levels. As with all of our heuristics, the accuracy of our error-modeling method is tested with simulation.

**Addressing Other Practical Implementation Issues**

In order to apply polyhedral estimation to practical problems we have to address several implementation issues. We note that other solutions to these problems may yield more or less accurate parameter estimates, and so the performance of the polyhedral method in the Monte Carlo simulations is a lower bound on the performance of the class of polyhedral methods.

**Product profiles with discrete features.** In most conjoint analysis problems, the features are specified at discrete levels, e.g., picture sizes of “postage stamp” or “3” square.” This constrains the elements of the $\tilde{x}$ vector to be 1, –1, or 0, depending on whether the left profile, the right profile, neither profile, or both profiles have the “high” feature. In this case we choose the vector that is most nearly parallel to the longest axis of the ellipsoid. Because we can always recode multi-level features or interacting features as binary features, the geometric insights still hold even if we otherwise simplify the algorithm.

**Restrictions on question design.** Sawtooth (1996) suggests that: “Most respondents can handle three attributes after they’ve become familiar with the task. Experience tells us that there does not seem to be much benefit from using more than three attributes. (p. 7).” We incorporate this constraint by restricting the set of questions over which we search when finding a question-vector that is parallel to the longest axis of the ellipse.

---

4 Technically, we use the “$\infty$-norm” rather than the “2-norm.” Exploratory simulations suggest that the choice of the norm does not have a large impact on the results. Nonetheless, this is one area suggested for future research.
First question. Unless we have prior information before any question is asked, the initial polyhedron of feasible utilities is defined by the boundary constraints. Because the boundary constraints are symmetric, the polyhedron is also symmetric and the polyhedral methods offer little guidance in the choice of a respondent’s first question. We choose the first question so that it helps improve estimates of the population means by balancing the frequency with which each attribute level appears in the set of questions answered by all respondents. In particular, for the first question presented to each respondent we choose attribute levels that appeared infrequently in the questions answered by previous respondents.

Programming. The optimization algorithms used for the simulations are written in MatLab and are available at the website on the cover page of this paper. We also provide the simulation code and demonstrations of web-based applications. All code is open-source.

Monte Carlo Experiments

The polyhedral methods for question selection and partworth estimation are new and untested. Although interior-point algorithms and the centrality criterion have been successful in many engineering problems, we are unaware of any application to conjoint analysis (or any other marketing problem). Thus, we turn to Monte Carlo experiments to identify circumstances in which polyhedral methods may contribute to the effectiveness of current conjoint methods.

Monte Carlo simulations offer at least two advantages over field tests involving actual customers. First, they can be repeated at little cost in a relatively short time period. This facilitates comparison of different techniques in a range of contexts. By varying parameters we can evaluate modifications of the techniques and hybrid combinations. We can also evaluate performance after varying characteristics of the respondents, including the heterogeneity and reliability of their responses. Second, simulations overcome the problem of identifying the correct answer. In studies involving actual customers, the true partial utilities are unobserved. This introduces problems in evaluating the performance of each method. In simulations the true partial utilities are constructed in advance and we simulate responses by adding error to these true measures. We then compare how well the methods identify the true utilities from the noisy responses.

Monte Carlo experiments have enjoyed a long history in the study of conjoint techniques, providing insights on interactions, robustness, continuity, attribute correlation, segmentation, new estimation methods, new data collection methods, post analysis with HB methods, and comparisons of ACA, CBC, and other conjoint methods. Although we focus on two benchmarks, ACA and efficient

---

fixed designs, there are many comparisons in the literature of these methods to other methods. (See reviews and citations in Green 1984, Green and Srinivasan 1978, 1990.)

The Monte Carlo experiments focus on five issues: (1) relative accuracy vs. the number of questions, (2) relative performance as the accuracy of self-explicated and paired-comparison data vary, (3) question selection vs. estimation including hybrid methods, (4) respondent wear out, and (5) relative performance on individual vs. population estimates. We begin by describing the design of the Monte Carlo experiments and then provide the results and interpretations.

**Design of the Experiments**

We focus on a design problem involving ten features, where the PD team is interested in learning the incremental utility contributed by each feature. We follow convention and scale to zero the partworth of the low level of a feature. This results in a total of ten parameters to estimate ($p = 10$). We feel that this $p$ is sufficient to illustrate the qualitative comparisons. We anticipate that the polyhedral methods are particularly well-suited to solving problems in which there are a large number of parameters to estimate relative to the number of responses from each individual ($q < p$). However, we would also like to investigate how well the methods perform under typical situations when the number of questions exceeds the number of parameters ($q > p$). In particular, Sawtooth (1996) recommends that the total number of questions be approximately three times the number of parameters – in our case 10 self-explicated and 20 paired-comparison questions. Thus, we examine estimates of partworths for all $q$ up to and including 20 paired-comparison questions.

We simulate each respondent’s partworths by drawing independently and randomly from a uniform distribution ranging from zero to 100. We explored the sensitivity of the findings to this specification by testing different methods of drawing partworths, including beta distributions that tend to yield more similar partworths (inverted-U shape distributions) or more diverse partworths (U-shaped distributions). This sensitivity analysis yielded a similar pattern of results, suggesting that the qualitative insights are not sensitive to the choice of this distribution.

To simulate the response to each paired-comparison question, we calculate the true utility difference between each pair of product profiles by multiplying the design vector by the vector of true partworths: $\mathbf{x} \cdot \mathbf{u}$. We assume that the respondents’ answers equal the true utility difference plus a zero-mean normal response error with variance $\sigma^2_{pc}$. The assumption of normally distributed error is common in the literature and appears to be a reasonable assumption about response errors. (Wittink and Cattin 1981 report no systematic effects due to the type of error distribution assumed.) For each comparison, we simulate 500 respondents.
Benchmark Methods

We compare the polyhedral method against two benchmarks, ACA and efficient fixed designs. For ACA we use the self-explicated (SE) and paired-comparison (PC) stages in Sawtooth’s algorithm. Estimates of the partworths are obtained after each paired-comparison question by minimizing the least squares criterion described in Equation 1. For the SE data, we assume that respondents’ answers are unbiased but imprecise. In particular, we simulate response error in the SE questions by adding to the vector of true partworths, \( \bar{u} \), a vector of independent identically distributed normal error terms with variance \( \sigma^2_{se} \).

In the fixed-design algorithm, for a given \( q \), we select the design with the highest efficiency (Kuhfield 1999, Kuhfield, Tobias, and Garratt 1994, Sawtooth 1999). This selection is as if the designer of the fixed design knew a priori how many questions would be asked. In general, this algorithm will do better than randomly selecting questions from a twenty-question efficient design, and thus provides a fair comparison to fixed designs. The specific algorithm that we use for finding an efficient design is Sawtooth’s CVA algorithm (1999). To estimate the partworths we use least-squares estimates (Equation 1 with no SE questions). Naturally, least-squares estimation requires that \( q \geq p \), so we cannot report fixed-algorithm results for \( q < p \).

All three methods use the PC questions, but only ACA requires the additional SE questions. If the SE questions are extremely accurate, then little information will be added by PC questions and ACA will dominate. Indeed, accuracy might even degrade for ACA as the number of PC questions grows (Johnson 1987). On the other hand, if the SE questions are very noisy, then the accuracy of all three methods will depend primarily on the PC questions. These two situations bound any empirical experience, thus we report results for two conditions – highly accurate SE questions and highly noisy SE questions. To facilitate comparisons among methods, we hold constant the noise in the PC questions.

To test relative performance we plot the absolute accuracy of the parameter estimates (true vs. estimated) averaged across attributes and respondents. Although our primary focus is the ability to estimate respondent-specific partworths, we also investigate how well the methods estimate mean partworths for the population.

Results of the Monte Carlo Experiments

Our goal is to illustrate the potential of the polyhedral methods and, in particular, to find situations where they add incremental value to the suite of conjoint analysis methods. We also seek to identify situations where extant methods are superior. As in any simulation analysis we cannot vary all

---

6 As described earlier, Sawtooth allows modifications of Equation 1. Based on in initial tests we used the estimation procedure that gave the best ACA results and thus did not penalize ACA.
parameters of the problem, thus, in these simulations we vary those parameters that best illustrate the
differences among the methods. So that other researchers might investigate other parameters our
simulation code is available on our website.

We select a moderate error in the paired-comparison questions. Most of the qualitative
interpretations hold for other error variances (e.g., Figures 5 and 6). In particular, we select \( \sigma_{pc} = 30 \). This
is 5% of the range of the answers to the PC questions and 30% of their maximum standard deviation (9% of
their variance).\(^7\)

Figure 5 compares the polyhedral algorithm to a set of efficient fixed designs. (Recall that we
choose the most efficient design for each \( q \) in order to get an upper bound on fixed designs. Choosing
random questions from a 20-question efficient design would not do as well as the benchmark in Figure 5a.) For \( q < 15 \) the polyhedral algorithm yields lower estimation error than the efficient fixed designs.
However, as more degrees of freedom are added to the least-squares estimates, about 50% more than the
number of parameters, we begin to see the advantage of orthogonality and balance in question design –
the goal of efficiency. However, even after twenty questions the performance of the polyhedral algorithm
is almost as effective as the most efficient fixed design. This is reassuring, indicating that the polyhedral
algorithm’s focus on rapid estimates from relatively few questions comes at the little loss in accuracy
when respondents provide responses to additional questions. Another way to look at Figure 5a is
horizontally; in many cases of moderate \( q \) the polyhedral algorithm can achieve the same accuracy as an
efficient fixed design, but with fewer questions. This is particularly relevant in a web-based context.

Figure 5b compares the polyhedral algorithm to the two ACA benchmarks. In one benchmark we
add very little error (\( \sigma_{se} = 10 \)) to the SE responses making them three times as accurate as the PC
questions (\( \sigma_{pc} = 30 \)). In the second benchmark we make the SE questions relatively noisy (\( \sigma_{se} = 50 \)).

From our own experience and from the empirical literature, we expect that these benchmarks should
bound empirical situations. We label the benchmark methods: “ACA (accurate priors)” and “ACA (noisy
priors).”

\(^7\) Each partworth is uniformly distributed \([0,100]\) and up to three features vary per question, thus, the PC responses
can vary on \([-300,300]\) with standard deviation of 100 (sum of three uniform variables). This is a maximum
because, for the ACA and polyhedral algorithms and for some \( q \)’s, the optimization might select less than three
features to vary. A review of the conjoint simulation literature suggests that the median error percentage (29%)
reported in that literature. Furthermore, Johnson (1987, 4) suggests that, with a 25% simulated error, ACA
estimation error “increases only moderately” relative to estimates based on no response error. Some results depend
on the choice of this error percentage – for example, all methods do uniformly better for low error variances than for
high error variances. We leave to future papers the complete investigation of error-variance sensitivity.
Figure 5: Comparison of Polyhedral Methods to ACA and Fixed Designs

![Polyhedral Algorithm vs. Efficient Fixed Designs](image1)

(a) Comparison to fixed designs

![Polyhedral Algorithm vs. ACA](image2)

(b) Comparison to ACA

As expected, the accuracy of the SE responses determines the precision of the ACA predictions. The polyhedral algorithm outperforms the ACA method when the SE responses are noisy but does not perform as well when respondents are able to give highly accurate self-explicated responses. Interestingly, the accuracy of the ACA method initially worsens when the priors are highly accurate. (See also Johnson 1987.) This highlights the relative accuracy of the SE responses compared to the PC responses in this benchmark. Not until \( q > p \) does the efficiency of least-squares estimation begin to reduce this error. Once sufficient questions are asked, the information in the PC responses begins to outweigh measurement error and the overall accuracy of ACA improves. However, despite ACA’s ability to exploit accurate SE responses, the polyhedral algorithm (without SE questions) begins to approach ACA’s accuracy soon after \( q > p \). This ability to eliminate SE questions can be important in web-based interviewing if the SE questions add significantly to respondent wear out.

For noisy SE responses, ACA’s accuracy never approaches that of the polyhedral algorithm, even when \( q=2p \), the number of questions suggested by Sawtooth. Summarizing, Figure 5 suggests that ACA is the better choice if the SE responses are highly accurate (and easy to obtain). The polyhedral algorithm is likely a better choice when SE responses are noisy or difficult to obtain. The selection of algorithms depends upon the researcher’s expectations about the context of the application. For example, for product categories in which customers often make purchasing decisions about features separately, perhaps by purchasing from a menu of features, we might expect more accurate SE responses. In contrast, if the features are typically bundled together, so that customers have little experience in evaluating the importance of the individual features, the accuracy of the SE responses may be lower. Relative accuracy of the two sets of questions may also be affected by the frequency with which customers purchase in the category and their consequent familiarity with the features.
Including Priors in Polyhedral Algorithms

Although the polyhedral methods were developed to gather preference information in as few questions as possible, Figure 5b suggests that if SE responses can be obtained easily and accurately, then they have the potential to improve the accuracy of adaptive conjoint methods. This is consistent with the conjoint literature, which suggests that both SE and PC questions add incremental information to conjoint estimates (Green, Goldberg, and Montemayor 1981, Huber, Wittink, Fiedler, and Miller 1993, Johnson 1999, Leigh, MacKay, and Summers 1984). This evidence raises the possibility that the precision of polyhedral methods can also be improved by incorporating SE responses.

To examine the effectiveness of including SE responses in polyhedral algorithms and to isolate the polyhedral question-selection method, we test a hybrid method that combines the polyhedral question selection method with the estimation procedure in Equation 1. That is, we use ACA’s estimation procedure, but replace ACA’s criteria of balance and orthogonality with polyhedral question selection. Figure 6a compares the two question-selection methods holding constant the estimation procedures and the noise level of the priors. This figure suggests that polyhedral question selection has the potential to improve ACA. We observe a similar pattern of results when comparing the methods under more accurate priors.

Figure 6: Including SE Responses in Polyhedral Algorithms

To examine whether SE responses always improve the polyhedral estimates, Figure 6b compares polyhedral estimates without priors (black line from Figure 5b) to estimates based on accurate priors and estimates based on noisy priors, retaining the same levels of noise in the PC responses. As in Figure 5b, the choice of method depends upon the accuracy of the SE responses. If SE responses are accurate and easy to obtain, then combining ACA estimation with polyhedral question selection yields more accurate forecasts than either ACA alone or the polyhedral algorithm alone. As the SE responses become noisy,
then the hybrid method becomes increasingly less accurate until, at a moderate noise level, it is better to ignore the SE responses altogether and use polyhedral question selection with analytic center estimation.

**Modeling Respondent Wear Out**

Both our own experience in web-based surveys and the published references cited earlier indicate that respondents are much less patient when completing surveys on the web compared to respondents recruited to a central location. (See, for example, Lenk, DeSarbo, Green, and Young 1996, p. 173). If respondents wear out as the length of the survey grows, we expect that the response error will be higher for later questions than for earlier questions. This would magnify the importance of the initial questions. Although we do not know for sure how fast respondents tire, we can simulate the effect by allowing the noise to grow as the number of questions increase. Our goal is to demonstrate the phenomenon and to investigate how it affects each method. We hope also to motivate empirical investigations into the shape of the wear out function.

We could find little precedent for the wear out function in the literature, so we assume a simple linear growth function. In particular, if $\epsilon$ denotes a draw of normal measurement error for the PC questions, then, in our wear-out analysis, we select $\epsilon(q) = \epsilon q/10$. Dividing by 10 ensures that the average of the error term is unchanged by this modification when only PC questions affect wear out. Because ACA includes both SE and PC questions, we consider two approaches for modeling ACA wear out. In the first approach we assume that only the PC questions affect wear out. In the second approach, we assume that both the SE and PC questions affect wear out and do so equally. In this second approach the error is given by $\epsilon'(q) = \epsilon(q+p)/10$. In any empirical situation we expect that these two approaches will bound the true wear-out phenomenon because we expect that the SE questions contribute to wear out, but not as much as the PC questions. We label the two ACA benchmarks as ACA (PC questions) and ACA (all questions), respectively. For ease of comparison we leave the variance of the error terms unchanged from the previous figures and assume that the response error in the SE questions is constant for all questions.

Figure 7 summarizes the simulation of respondent wear out. Initially, as more PC questions are answered, estimation accuracy improves. The new information improves the estimates even though that information becomes increasingly noisy. After approximately 10-12 questions the added noise overwhelms the added information and the estimates begin to degrade yielding a U-shaped function of $q$. The rate of degradation for the ACA benchmarks is slower – the U-shape begins to appear around question 18. The slower degradation can be explained, in part, because ACA uses the SE responses to reduce reliance on the increasingly inaccurate PC questions. This interpretation is supported by the results of further simulations not reported in Figure 7. The hybrid method of Figure 6 declines at a rate
similar to ACA, reflecting the inclusion of SE responses in the hybrid method. Efficient fixed designs, which do not use SE responses, decline at a rate similar to the polyhedral method.8

Figure 7: Modeling Respondent Wear Out

![Graph showing modeling of respondent wear out](image_url)

The exact amount of wear out and rate at which it grows is an empirical question. However, Figure 7 suggests that wear out can have an important impact on the researcher’s choice of methods. In particular, if the SE questions contribute to wear out, then the researcher should favor methods such as the polyhedral algorithm or efficient fixed designs that do not require the respondent to answer SE questions. Our own experience suggests that wear out does occur (e.g., Chan 1999 and McArdle 2000), but that some wear out can be mitigated by changing the type of question to preserve respondent interest. However, we do not yet know whether this is sufficient to compensate for asking additional questions.

Estimates of Utilities for the Population

The previous figures report the mean absolute error when predicting the partworths for each respondent. We feel this is appropriate because the PD team often needs respondent-specific partworths to explore product-line decisions and/or segmentation. However, if the population is homogeneous, then the PD team may seek to estimate a single set of partworths to represent the population’s preferences. To estimate partworths for the population we can aggregate across all of the respondents in the population.

To investigate this issue we draw twenty sets of population means and simulate 200 respondents for each set of population means. (1/26/01 draft: 5 groups of 50. Rest to be completed.) Data is pooled

---

8 We invite the reader to explore this wear out phenomenon with alternative levels of noise in the SE responses. We have found that the results vary with noise level in a manner analogous to the discussions of Figures 5 and 6. As the SE responses become more accurate, ACA and the hybrid method perform relatively better.
within each group of 200 respondents and OLS is used to estimate a single set of partworths for that group. We then calculate the average mean absolute error, averaging across all 20 groups. To simulate the population means for each group we draw independently and randomly from a uniform distribution ranging from 25 to 75. We then simulate the partworths for the 200 respondents in each group by adding heterogeneity terms to the vector of population means for each respondent where the heterogeneity terms are drawn from a uniform distribution ranging from $-25$ to $25$. This induces an overall triangle distribution ranging from 0 to 100 for each partworth. Finally, given the vectors of partworths for each individual, we proceed as before, adding measurement error to construct the PC and SE responses.

**Figure 8. Predictions of population-level partworths**

![Figure 8](image_url)

*Figure 8* suggests that the magnitudes of the MAE are much lower for population-level estimates than for respondent-level estimates. This reflects the much larger sample of data used to estimate the partworths. As expected all of the methods perform well, with the polyhedral method and efficient fixed designs offering slightly more accurate parameter estimates than ACA with noisy priors. Although we might improve all methods by selecting questions that vary optimally across respondents, it is reassuring that even without such modifications, all of the techniques yield relatively accurate estimates of the population utilities.

**Summary of the Results of the Monte Carlo Experiments**

Our simulations suggest that no method dominates in all situations, but that there are a range of relevant situations where the polyhedral algorithm or the ACA-polyhedral hybrid is a useful addition to the suite of conjoint analysis methods available to a product developer or a market researcher. If SE

---

responses can be obtained accurately (relative to PC responses) and with little respondent wear out, then
either ACA or the ACA-polyhedral hybrid method is likely to be most accurate. If new PC question
formats can be developed that engage respondents with visual, interactive media, such that respondents
might be willing and able to answer sufficient questions such that $q > 1.5p$, then efficient fixed designs
might perform better than adaptive methods such as ACA, the polyhedral algorithm, or a hybrid.

The real advantage of the polyhedral methods comes when the researcher is limited to relatively
few questions ($q < p$), when wear out is a significant concern, and/or when SE responses are noisy
relative to PC responses. We believe that these situations are becoming increasingly relevant to conjoint
analysis applications. With the advent of spiral product development (PD) processes, where the PD team
cycles through iterative designs many times before the product is launched, PD teams are seeking
methods that can screen large numbers of features quickly. They are willing to live with imperfect
accuracy since they only seek to discard features that are relatively unimportant. This is particularly
relevant in the fuzzy front end of PD. At the same time, more and more interviewing is moving to web-
based environments where respondents are becoming increasingly impatient. In these environments
researchers are increasingly interested in algorithms that reduce the number of questions that need be
asked for a given level of accuracy. Wear out is a well-known, but rarely quantified phenomenon. Web-
based interviewing might increase wear out because respondents are impatient, but the web’s ability to
use multimedia interactive stimuli might actually decrease wear out. This remains an empirical issue, but
the polyhedral methods show promise when impatience dominates the enhanced stimuli.

Finally, the relative accuracy of SE vs. PC responses, and hence the choice of conjoint analysis
method, is likely to depend upon context. For complex products, for products where industrial design is
important, or for products with a high emotional content, then it might be easier for a respondent to make
a holistic judgment by comparing pairs of products than it would be for the respondent to evaluate the
products feature by feature. The new methods, where realistic, but virtual, products are presented to
respondents in a web-based environment, might also enhance the ability of respondents to make holistic
judgments. (The website listed on the cover page of this paper provides links to demonstrations of these
virtual-customer methods.)

Summary and Conclusions

We have proposed and tested a new conjoint analysis method designed specifically for web-based
interviewing. The method is designed to identify, using relatively few questions, features that have the
most influence on customer preferences. It uses advanced interior-point mathematical programming
methods to adaptively select question designs that constrain the set of feasible parameters. The method

(1989), Moore (1980), Moore and Semenik (1988), and Wittink and Montgomery (1979) all provide evidence that
uses centrality concepts to obtain reasonable estimates of individual preferences, even when individuals answer fewer questions than there are parameters to estimate.

We tested an example of a polyhedral algorithm by using a series of Monte Carlo simulations. The findings confirm that the method is particularly suited to web contexts, where subjects are less patient and more prone to wear out than in traditional central facility contexts. By isolating the impact of the question design component, we found that the relative accuracy of the polyhedral method is due, at least in part, to the design of the questions. Our simulations suggest that hybrid polyhedral question-selection methods could be used to enhance existing estimation methods.

To evaluate the contribution of the proposed polyhedral methods we had to make several practical implementation decisions. Examples include the choice of the first question and the procedures to design questions and estimate partworths when responses are inconsistent (the feasible set is empty). We recognize that as-yet undiscovered variations might yield more accurate predictions. In this respect, the performance of the specific polyhedral method evaluated in the Monte Carlo simulations should be interpreted as a lower bound on the potential performance of this class of methods. Like many new technologies, we expect its performance to improve with use and evolution (Bower and Christensen 1995, Christensen 1998). Under this interpretation, the performance of the specific method that we investigated is gratifying and suggests that this class of methods is worth investigating. For example, the judicious selection of profiles in the choice-based conjoint (CBC) task can also be viewed as the selection of hyperplanes that constrain the set of feasible parameters. In theory, with polyhedral methods we should be able to design an adaptive CBC algorithm that provides partworth estimates for each respondent.

More generally, we believe that the polyhedral methods used in this study are just one of a range of recent developments in the mathematical programming and optimization literatures that can contribute to our understanding of marketing problems. Because these methods can obtain near optimal solutions extremely fast, they might be used to select promotion variables dynamically as respondents navigate a website. Alternatively, they might be used to design ongoing experiments in which parameters of a website are varied in an optimal manner trading off current effectiveness for long-term improvement.

respondent-specific partworths predict better than population means.
References


Mathematics of Fast Polyhedral Adaptive Conjoint Estimation

Consider the case of \( p \) parameters and \( q \) questions where \( q \leq p \). Let \( u_j \) be the \( j \)th parameter of the respondent’s partworth function and let \( \mathbf{u} \) be the \( p \times 1 \) vector of parameters. Without loss of generality we assume binary features such that \( u_j \) is the high level of the \( j \)th feature. For more levels we simply recode the \( u \) vector and impose constraints such as \( u_m \leq u_h \). We handle these and other external constraints by adding slack variables, \( v_{hm} \geq 0 \), such that \( u_h = u_m + v_{hm} \). These new constraints are handled analogously to those imposed by the respondent’s answers.

Let \( \mathbf{z}_{i\ell} \) be the \( 1 \times p \) vector describing the left-hand profile in the \( i \)th paired-comparison question and let \( \mathbf{z}_{i\ell}^r \) be the \( 1 \times p \) vector describing the right-hand profile. The elements of these vectors are binary indicators taking on the values 0 or 1. Let \( X \) be the \( q \times p \) matrix of \( X_{ij} = \mathbf{z}_{i\ell}^r - \mathbf{z}_{i\ell} \) for \( i = 1 \) to \( q \). Let \( a_i \) be the respondent’s answer to the \( i \)th question and let \( \mathbf{a} \) be the \( q \times 1 \) vector of answers for \( i = 1 \) to \( q \). Then, if there were no errors, the respondent’s answers imply \( a_i X_{ij} = 0 \) for \( i = 1 \) to \( q \). These constraints (and any incorporated external constraints) form a polyhedron, \( \mathcal{P} = \{ \mathbf{u} \in \mathbb{R}^p \mid a_i X_{ij} = 0, \mathbf{u} \geq 0 \} \). We begin by assuming that \( \mathcal{P} \) is non-empty, that \( X \) is full-rank, and that no \( j \) exists such that \( u_j = 0 \) for all \( \mathbf{u} \in \mathcal{P} \). We later indicate how to handle these cases.

Finding an Interior Point of the Polyhedron

To begin the algorithm we first find a feasible interior point of \( \mathcal{P} \) by solving a linear program, \( \text{LP1} \). Let \( \mathbf{e} \) be a \( p \times 1 \) vector of 1’s and let \( \mathbf{0} \) be a \( p \times 1 \) vector of 0’s; the \( y_j \)’s and \( \theta \) are parameters of \( \text{LP1} \) and \( \mathbf{y} \) is the \( p \times 1 \) vector of the \( y_j \)’s. (When clear in context, inequalities applied to vectors apply for each element.) \( \text{LP1} \) is given by:

\[
\text{LP1} \quad \max \sum_{j=1}^{p} y_j, \quad \text{subject to: } X\mathbf{u} = \mathbf{0}, \quad \theta \geq 1, \quad \mathbf{u} \geq \mathbf{y} \geq \mathbf{0}, \quad \mathbf{y} \leq \mathbf{e}
\]

If \( (\mathbf{u}^*, \mathbf{y}^*, \theta^*) \) solves \( \text{LP1} \), then \( \theta^*/\mathbf{u}^* \) is an interior point of \( \mathcal{P} \) whenever \( \mathbf{y}^* > \mathbf{0} \). If there are some \( y_j \)’s equal to 0, then there are some \( j \)’s for which \( u_j = 0 \) for all \( \mathbf{u} \in \mathcal{P} \). If \( \text{LP1} \) is infeasible, then \( \mathcal{P} \) is empty. We address these cases later in this appendix.

Finding the Analytic Center

The analytic center is the point in \( \mathcal{P} \) that maximizes the geometric mean of the distances from the point to the faces of \( \mathcal{P} \). We find the analytic center by solving \( \text{OPT1} \).

\[
\text{OPT1} \quad \max \sum_{j=1}^{p} \ln(u_j), \quad \text{subject to: } X\mathbf{u} = \mathbf{a}, \quad \mathbf{u} > \mathbf{0}
\]

Freund (1993) proves with projective methods that a form of Newton’s method will converge rapidly for \( \text{OPT1} \). Indeed Freund’s algorithm is more general because it enables the \( u_j \)’s to be weighted, say to focus on features that are more important managerially. To implement Newton’s method we begin with the feasible point from \( \text{LP1} \) and improve it with a scalar, \( \alpha \), and a direction, \( \hat{d} \), such that \( \mathbf{u} + \alpha \hat{d} \) is close to the optimal solution of \( \text{OPT1} \). (\( \hat{d} \) is a \( p \times 1 \) vector of \( d_j \)’s.) We then iterate subject to a stopping rule.

We first approximate the objective function with a quadratic expansion in the neighborhood of \( \mathbf{u} \). That is:
FAST POLYHEDRAL ADAPTIVE CONJOINT ESTIMATION, APPENDIX

(A1) \[ \sum_{j=1}^{p} \ln(u_j + d_j) = \sum_{j=1}^{p} \ln(u_j) + \sum_{j=1}^{p} \left( \frac{d_j}{u_j - \frac{d^2_j}{2u_j^2}} \right) \]

If we define \( U \) as a \( p \times p \) diagonal matrix of the \( u_j \)'s, then the optimal direction solves OPT2:

(OPT2) \[ \max \quad \tilde{e}^T U^{-1} \tilde{d} - \left( \gamma_2 \right) \tilde{a}^T U^{-2} \tilde{d} \quad \text{subject to:} \quad X \tilde{d} = \tilde{0} \]

Newton’s method solves OPT1 quickly by exploiting an analytic solution to OPT2. To see this, consider first the Karush-Kuhn-Tucker (KKT) conditions for OPT2. They are (\( \tilde{z} \) is a \( p \times 1 \) vector that is a parameter of the KKT conditions that is unconstrained in sign):

(A2) \[ U^{-2} \tilde{d} - U^{-1} \tilde{e} = X^T \tilde{z} \]

(A3) \[ X \tilde{d} = \tilde{0} \]

Multiplying A2 on the left by \( XU^2 \), gives \( X \tilde{d} - XU\tilde{e} = XU^2 X^T \tilde{z} \). Applying A3 to this equation gives:

\[ -XU\tilde{e} = XU^2 X^T \tilde{z} \]

Since \( U\tilde{e} = \tilde{u} \) and since \( X\tilde{u} = \tilde{a} \), we have \( -\tilde{a} = XU^2 X^T \tilde{z} \). Because \( X \) is full rank and positive, we invert \( XU^2 X^T \) to obtain \( \tilde{z} = -(XU^2 X^T)^{-1} \tilde{a} \). Now replace \( \tilde{z} \) in A2 by this expression and multiply by \( U^2 \) to obtain \( \tilde{d} = \tilde{u} - U^2 X^T (XU^2 X^T)^{-1} \tilde{a} \).

According to Newton’s method, the new estimate of the analytic center, \( \tilde{u}' \), is given by
\[
\tilde{u}' = \tilde{u} + \alpha \tilde{d} = U(\tilde{e} + \alpha U^{-1} \tilde{d})
\]

There are two cases for \( \alpha \). If \( \|U^{-1} \tilde{d}\| < \gamma_2 \), then we use \( \alpha = 1 \) because \( \tilde{u} \) is already close to optimal and \( \tilde{e} + \alpha U^{-1} \tilde{d} > \tilde{0} \). Otherwise, we find \( \alpha \) with a line search.

Special Cases

If \( X \) is not full rank, \( XU^2 X^T \) might not invert. We can either select questions such that \( X \) is full rank or we can make it so by removing redundant rows. Suppose that \( \tilde{x}_k \) is a row of \( X \) such that
\[
\tilde{x}_k^T = \sum_{i=1}^{q} \beta_i \tilde{x}_i^T
\]

Then if \( a_k = \sum_{i=1,j \neq k}^{q} \beta_i a_i \), we remove \( \tilde{x}_k \). If \( a_k = \sum_{i=1}^{q} \beta_i a_i \), then \( \mathcal{P} \) is empty and we employ OPT4 described later in this appendix.

If in LP1 we detect cases where some \( y_j \)'s = 0, then there are some \( j \)'s for which \( u_j = 0 \) for all \( u \in \mathcal{P} \). In the later case, we can still find the analytic center of the remaining polyhedron by removing those \( j \)'s and setting \( u_j = 0 \) for those indices. If \( \mathcal{P} \) is empty we employ OPT4.

Finding the Ellipsoid and its Longest Axis

If \( \tilde{u} \) is the analytic center and \( \tilde{U} \) is the corresponding diagonal matrix, then Freund (1993, Theorem 2.1) demonstrates that \( \mathcal{E} \subseteq \mathcal{P} \subseteq \mathcal{E}_p \) where, \( \mathcal{E} = \{ \bar{u} \mid X\bar{u} = \bar{a}, (\bar{u} - \bar{u})^T \tilde{U}^{-2} (\bar{u} - \bar{u}) \leq 1 \} \) and \( \mathcal{E}_p \) constructed proportional to \( \mathcal{E} \) by replacing 1 with \( p \). Because we are interested only in the direction of the longest axis of the ellipsoids we can work with the simpler of the proportional ellipsoids, \( \mathcal{E} \). Let \( \tilde{g} = \tilde{u} - \tilde{u} \), then the longest axis will be a solution to OPT3.

(OPT3) \[ \max \quad \tilde{g}^T \tilde{g} \quad \text{subject to:} \quad \tilde{g}^T \tilde{U}^{-2} \tilde{g} \leq 1, \quad X \tilde{g} = \tilde{0} \]

OPT3 has an easy-to-compute solution based on the eigenstructure of a matrix. To see this we begin with the KKT conditions (where \( \phi \) and \( \gamma \) are parameters of the conditions).

(A4) \[ \tilde{g} = \phi \tilde{U}^{-2} \tilde{g} + X^T \tilde{\gamma} \]
\[(A5) \quad \phi(g^T \bar{U}^{-2} \bar{g} - 1) = 0\]

\[(A6) \quad g^T \bar{U}^{-2} \bar{g} \leq 1, \quad X \bar{g} = \bar{0}, \quad \phi \geq 0\]

It is clear that \(g^T \bar{U}^{-2} \bar{g} = 1\) at optimal, else we could multiply \(\bar{g}\) by a scalar greater than 1 and still have \(\bar{g}\) feasible. It is likewise clear that \(\phi\) is strictly positive, else we obtain a contradiction by left-multiplying \(A4\) by \(\bar{g}^T\) and using \(X \bar{g} = \bar{0}\) to obtain \(g^T \bar{g} = 0\) which contradicts \(g^T \bar{U}^{-2} \bar{g} = 1\). Thus, the solution to OPT3 must satisfy \(g = \phi \bar{U}^{-2} \bar{g} + X^T \bar{\omega}, \quad g^T \bar{U}^{-2} \bar{g} = 1, \quad X \bar{g} = \bar{0}, \quad \phi > 0\).

We rewrite \(A4-A6\) by letting \(I\) be the identify matrix and defining \(\eta = 1/\phi\) and \(\omega = \gamma / \phi\)

\[(A7) \quad (\bar{U}^{-2} - \eta I) \bar{g} = X^T \bar{\omega}\]

\[(A8) \quad g^T \bar{U}^{-2} \bar{g} = 1\]

\[(A9) \quad X \bar{g} = \bar{0}, \quad \phi > 0\]

We left-multiply \(A7\) by \(X\) and use \(A9\) to obtain \(X \bar{U}^{-2} \bar{g} = XX^T \bar{\omega}\). Since \(X\) is full rank, \(XX^T\) is invertible and we obtain \(\bar{\omega} = (XX^T)^{-1} \bar{U}^{-2} \bar{g}\) which we substitute into \(A7\) to obtain

\[(\bar{U}^{-2} - X^T (XX^T)^{-1} \bar{U}^{-2}) \bar{g} = \eta \bar{g}\]. Thus, the solution to OPT3 must be an eigenvector of the matrix, \(M \equiv (\bar{U}^{-2} - X^T (XX^T)^{-1} \bar{U}^{-2})\). To find out which eigenvector, we left-multiply \(A7\) by \(\bar{g}^T\) and use \(A8\) and \(A9\) to obtain \(\eta \bar{g}^T \bar{g} = 1\), or \(\bar{g}^T \bar{g} = 1/\eta\) where \(\eta > 0\). Thus, to solve OPT3 we maximize \(1/\eta\) by selecting the smallest positive eigenvector of \(M\). The direction of the longest axis is then given by the associated eigenvector of \(M\). We then choose the next question such that \(\bar{x}_{q+1}\) is most nearly collinear to this eigenvector subject any constraints imposed by the questionnaire design. (For example, in our simulation we require that the elements of \(\bar{x}_{q+1}\) be \(-1, 0, 1\).) The answer to \(\bar{x}_{q+1}\) defines a hyperplane orthogonal to \(\bar{x}_{q+1}\).

We need only establish that the eigenvalues of \(M\) are real. To do this we recognize that \(M = P \bar{U}^{-2}\) where \(P = (I - X^T (XX^T)X)\) is symmetric, i.e., \(P = P^T\). Then if \(\eta\) is an eigenvalue of \(M\), det \((P \bar{U}^{-2} - \eta I) = 0\), which implies that det \([\bar{U}(\bar{U}^{-1} P \bar{U}^{-2} - \eta I) \bar{U}^{-1}] = 0\). This implies that \(\eta\) is an eigenvalue of \(\bar{U}^{-1} P \bar{U}^{-2}\), which is symmetric. Thus, \(\eta\) is real (Hadley 1961, 240).

**Adjusting the Polyhedron so that it is non-Empty**

\(P\) will remain non-empty as long as respondents’ answers are consistent. However, in any real situation there is likely to be \(q < p\) such that \(P\) is empty. To continue the polyhedral algorithm, we adjust \(P\) so that it is non-empty. We do this by replacing the equality constraint, \(X \bar{u} = \bar{a}\), with two inequality constraints, \(X \bar{u} \geq \bar{a} + \bar{\delta}\) and \(X \bar{u} \leq \bar{a} - \bar{\delta}\), where \(\bar{\delta}\) is a \(q\times1\) vector of errors, \(\bar{\delta}\). We solve the following optimization problem. Our current implementation uses the \(\infty\)-norm where we minimize the maximum \(\bar{\delta}\), but other norms are possible.

\[(OPT4) \quad \min \|\bar{\delta}\| \quad \text{subject to: } \begin{align*}
X \bar{u} &\leq \bar{a} + \bar{\delta}, \\
X \bar{u} &\geq \bar{a} - \bar{\delta}, \\
\bar{u} &\geq \bar{0}, \\
\bar{\delta} &\geq \bar{0}
\end{align*}\]

At some point such that \(q > p\), extant algorithms will outperform OPT4 and we can switch to those algorithms. Alternatively, a researcher might choose to switch to constrained regression (norm-2) when \(q > p\). We leave these extensions to future research.