An Optimization Framework for Adaptive Questionnaire Design

by

J. Abernethy
T. Evgeniou
and
JP. Vert

2004/91/TM

An Optimization Framework for Adaptive Questionnaire Design

Jacob Abernethy[1]
Toyota Technological Institute at Chicago,
Chicago, IL 60637, USA
jabernethy@tti-c.org

Theodoros Evgeniou
Technology Management,
INSEAD,
Bd de Constance, Fontainebleau 77300, France
Tel: +33 (0)1 60 72 45 46
Fax: +33 (0)1 60 74 55 01
theodoros.evgeniou@insead.edu

Jean-Philippe Vert
Ecole des Mines de Paris
35 rue Saint-Honoré, Fontainebleau 77300, France
jean-philippe.vert@mines.org

[1]The work within this paper was performed while the author was at INSEAD, Bd de Constance, Fontainebleau, 77300, France.

An Optimization Framework for Adaptive Questionnaire Design

Abstract

We propose a general approach for adaptively designing questionnaires for conjoint analysis customized at the individual level. At each step the next question presented to an individual is designed on the fly and computationally fast based on the responses the individual has given to all previous choice questions. Our framework also encompasses recent polyhedral adaptive conjoint methods as a special case. Within our framework we develop a novel conjoint analysis method that is in the spirit of recently proposed conjoint estimation methods. We test the proposed method on widely used simulation data and compare the effectiveness of the designed questionnaires with a standard orthogonal design, a random design, and a polyhedral adaptive conjoint questionnaire under varying conditions. The results show that the proposed method leads to individual-specific questionnaires and estimations of individual utilities that are significantly more accurate than what is estimated with the other methods and questionnaires when there is high response error. We finally show that further significant improvements are achieved when we use a hybrid individual-specific and aggregate customization method that we also develop within our general framework for conjoint analysis.

Keywords: Choice Models, Marketing Research, Data Mining, Regression And Other Statistical Techniques, Marketing Tools.

1 Introduction

A central problem in marketing, for example in the area of conjoint analysis, is the design of questionnaires that can effectively capture the preferences of individuals. The goal is to elicit the preferences of individuals *as accurately* as possible while asking *as few* questions as possible, either in order to avoid subjects' fatigue or simply because it is not practical to have answers to many questions. Intuitively this can be better achieved if the questions asked are well-designed adaptively, so that every new question is built based on the answers provided by an individual to previous questions. This would also lead to choice questionnaires that are customized for each individual, since the answers provided to the questions differ from individual to individual. Furthermore each question needs to be designed computationally efficiently since the questions are constructed on the fly, while individuals respond. This can be very useful, for example, for online market research, an emerging trend.

In this paper we present a novel general approach for conjoint analysis focusing on the design of individually adapted conjoint questionnaires. It is based on the iterative optimization of cost functions motivated by statistical learning theory and Support Vector Machines (SVM) (Vapnik 1998), much like recently proposed conjoint *estimation* methods (Cui and Curry 2004; Evgeniou et al 2004) developed *only* for the estimation of preference models from existing data and *not* for the design of questionnaires. The iterative optimization approach we present can also be seen as a generalization of recently developed adaptive polyhedral conjoint analysis methods (Toubia et al 2003; Toubia et al

2004). Within our general framework we develop a novel conjoint analysis method that is computationally very efficient and can be used to design questionnaires in real time customized adaptively for each individual. We believe that the proposed framework is general enough so that it can also be used by researchers to develop other methods and new theory in the area of conjoint analysis and possibly other marketing problems.

There has been a lot of work on designing questionnaires that can efficiently elicit the preferences of individuals (Arora and Huber 2001; Huber and Zwerina 1996; Sandor and Wedel 2001; Toubia et al 2003; Toubia et al 2004). Typically in conjoint studies the same questions, for example questions coming from an orthogonal design (Huber and Zwerina 1996), are given to all individuals. There has also been work on customizing the questionnaires at the individual level using responses from other individuals, often called *aggregate customization* (Arora and Huber 2001; Huber and Zwerina 1996; Sandor and Wedel 2001). Finally, recently Toubia et al (2003; 2004) proposed an approach to customizing questionnaires at the individual level using *only* the responses the same individual gave to previous questions. Their approach is based on polyhedral optimization methods and has shown very promising results.

In this paper we present a general framework for developing individually-customized questionnaires based on optimization theory like the conjoint estimation methods of (Cui and Curry 2004; Evgeniou et al 2004; Srinivasan and Shocker 1973; Srinivasan 1998) and the conjoint analysis methods of (Toubia et al 2003; Toubia et al 2004). Within this general framework we develop a novel conjoint analysis method, and particularly a novel method for designing conjoint questionnaires which we experimentally compare with existing questionnaires such as an orthogonal design, a random questionnaire, and the design developed using the method of Toubia et al (2004). The specific method we show is for choice-based type conjoint analysis, but similar methods can be developed for example for metric based conjoint analysis or other questionnaire design problems.

Although the focus of this paper is on the design of questionnaires for a single individual, we also discuss how to follow a *hybrid* approach between *aggregate customization* and *individual-specific customization* within our general framework. We show experimentally that such an approach leads to further significant improvements which indicate that future work on such hybrid individual specific and aggregate customization conjoint analysis methods is very promising. We leave this for future work.

The paper is organized as follows. In section 2 we first present our general framework and then develop a novel conjoint analysis method within this framework which we test experimentally in section 3. In section 4 we discuss how to extend the proposed conjoint analysis method, always within our general framework, to develop hybrid individual specific and aggregate customization conjoint methods which we believe, as our experiments indicate, that are very promising. Finally we conclude in section 5.

2 A General Framework for Conjoint Analysis

We consider conjoint analysis, in a simple view, having two key building blocks:

1. A method for designing a questionnaire to be given to the individuals, and

2. An estimation method for estimating the utility functions of the individuals based on the responses to the questionnaire.

We first discuss (2), the general framework for the estimation methods, and then we shift to the general framework for the questionnaire design methods which is our main focus. We finally develop a novel conjoint analysis method within the proposed framework, which we test experimentally in the next section.

2.1 Notation

We make the standard assumption (Ben-Akiva and Lerman 1985; Srinivasan and Shocker 1973) that the utility functions are linear functions. We note that an important advantage of our approach is that it can be easily and computationally efficiently generalized to the case of highly non-linear utility functions, for example estimating all interactions among product attributes, using the kernel transformations as done in the Support Vector Machines (SVM) literature (Cortes and Vapnik 1995; Vapnik 1998) and in (Cui and Curry 2004; Evgeniou et al 2004). For simplicity we do not discuss this here.

We denote products with $\mathbf{x}$ and the true utility of an individual with $\bar{\mathbf{w}}$. Therefore the utility that the individual assigns to a product $\mathbf{x}$ is $U(\mathbf{x}) = \bar{\mathbf{w}} \cdot \mathbf{x}$. For simplicity we consider first the case that we design questions where the individual has to choose among two products. So question i is to choose the preferred product among $\{\mathbf{x}_{i1}, \mathbf{x}_{i2}\}$. We discuss how to generalize this later. The goal is to design questions (pairs of products) for an individual adaptively so that the $i+1$ question is designed using the information we have about the choices the individual has made to the first i questions. To simplify notation we assume that for each i the first product $\mathbf{x}_{i1}$ is the preferred one - we can rename the products otherwise.

2.2 A General Framework for Conjoint Estimation

2.2.1 Support Vector Machine (SVM) Type Conjoint Estimation

In (Cui and Curry 2004; Evgeniou et al 2004) a conjoint estimation method is proposed in the spirit of SVM (Cortes and Vapnik 1995; Vapnik 1998) a statistical method that is widely and successfully used in the machine learning and data mining areas. We briefly review it here since what we will present below is a generalization of that method.

Given the answers to n choice questions $\{(\mathbf{x}_{11}, \mathbf{x}_{12}), (\mathbf{x}_{21}, \mathbf{x}_{22}), \ldots, (\mathbf{x}_{n1}, \mathbf{x}_{n2})\}$, where we assume the first product $\mathbf{x}_{i1}$ is always chosen, Cui and Curry (2004) and Evgeniou et al (2004) estimate the utility function $\mathbf{w}$ as the minimizer of the following cost function:

$$\min_{\mathbf{w}} \sum_{i=1..n} \theta(1 - \mathbf{w} \cdot (\mathbf{x}_{i1} - \mathbf{x}_{i2})) \cdot (1 - \mathbf{w} \cdot (\mathbf{x}_{i1} - \mathbf{x}_{i2})) + \lambda \|\mathbf{w}\|^2, \tag{1}$$

where $\theta(x) = 1$ if $x > 0$ and 0 otherwise. It turns out that this is a quadratic optimization problem which has a unique solution and can be solved efficiently (Cortes and Vapnik 1995; Vapnik 1998). In (Cui and Curry 2004; Evgeniou et al 2004) the estimation method is written in a different, but equivalent way, namely as:

$$\min_{\mathbf{w},\xi_i} \sum_{i=1..n} \xi_i + \lambda\|\mathbf{w}\|^2$$

subject to:

$$\mathbf{w} \cdot \mathbf{x}_{i1} \geq \mathbf{w} \cdot \mathbf{x}_{i2} + 1 - \xi_i$$

for $\forall i \in \{1, \ldots, n\}$, and

$$\xi_i \geq 0. \tag{2}$$

It is easy to see that problems (1) and (2) are equivalent. We use formulation (1) to show our generalization more clearly below.

We note that the cost function (1) that we optimize consists of two parts. The first part measures the error made by the estimated utility function $\mathbf{w}$ on the existing data $(\mathbf{x}_{i1}, \mathbf{x}_{i2})$. It is easy to see that if:

$$\mathbf{w} \cdot \mathbf{x}_{i1} \geq \mathbf{w} \cdot \mathbf{x}_{i2} + 1$$

then there is no "penalty" for question i, and otherwise the penalty is

$$1 - (\mathbf{w} \cdot \mathbf{x}_{i1} - \mathbf{w} \cdot \mathbf{x}_{i2}).$$

Constant 1 plays a scaling role for $\mathbf{w}$. The second part, $\|\mathbf{w}\|^2$, controls the complexity of the solution $\mathbf{w}$ and incorporating it is necessary in order to avoid two key problems: sensitivity to noise, and the curse of dimensionality (Vapnik 1998). The parameter λ of the cost function is a user-defined parameter (below we set it by default to $\frac{1}{n}$) that can be also chosen using, for example, cross-validation or a validation set (Evgeniou et al 2004; Vapnik 1998). It defines a trade off between the first part of the cost function, measuring the error made on the n questions, and the second part, measuring the complexity of the solution. Incorporating the complexity control $\|\mathbf{w}^2\|$ with a trade off defined by a *non-zero and non-infinity* parameter λ is crucial: it leads to models that do not suffer from the curse of dimensionality and that are robust to noise, as discussed and also shown experimentally by (Cui and Curry 2004; Evgeniou et al 2004; Vapnik 1998) and many others. Consistent with past work, our experiments below also show that the advantage of the proposed conjoint analysis method relative, for example, to the method of Toubia et al (2004) is greater when there is high response error - we discuss this below. A lot of theoretical work on this type of methods is done in the area of machine learning and SVM. We refer the reader for example to www.kernelmachines.org for more information.

2.2.2 A General Conjoint Estimation Method

One can generalize the estimation method (1) by simply using a different penalty function for the errors made on the past questions and different complexity costs. The general problem (1) can be written as:

$$\mathbf{min_w} \; V(\mathbf{w}, \{(\mathbf{x}_{i1}, \mathbf{x}_{i2})\}) + \lambda \Phi(\mathbf{w}), \tag{3}$$

where $V(\mathbf{w}, \{(\mathbf{x}_{i1}, \mathbf{x}_{i2})\})$ is some function which measures the error that the utility function $\mathbf{w}$ makes on the n responses (choice among products $\mathbf{x}_{i1}$ and $\mathbf{x}_{i2}$), and Φ is a complexity cost function, such as the square norm used above for SVM.

Different utility estimation methods can be defined simply by choosing appropriate V and Φ functions. For example we saw above how the SVM type estimation methods are special cases of (3). We consider here two other special cases of (3):

- Choosing the square loss function:

$$V(\mathbf{w}, \{(\mathbf{x}_{i1}, \mathbf{x}_{i2})\}) = \sum_{i=1}^{n} (1 - \mathbf{w} \cdot (\mathbf{x}_{i1} - \mathbf{x}_{i2}))^2$$

  and having, like in SVM, $\Phi(\mathbf{w}) = \|\mathbf{w}\|^2$ leads to least square estimation with complexity control, which is known as Regularization Networks (RN) (Tikhonov and Arsenin, 1977; Vapnik, 1998; Evgeniou et al., 2000):

$$\mathbf{min_w} \sum_{i=1 \ldots n} (1 - \mathbf{w} \cdot (\mathbf{x}_{i1} - \mathbf{x}_{i2}))^2 + \lambda \|\mathbf{w}\|^2 \tag{4}$$

  Below we will use this estimation method because first, as we note below, it leads to very fast adaptive questionnaire designs, and second the cost functional is convex and twice differentiable which is required for the questionnaire design approach we develop below. We will be noting this method with RN in what follows.

- It turns out (we show the derivation in the Appendix) that the estimation method of Toubia et al (2004) can be written, in the special case that there is no response error, as:

$$\mathbf{min}_{\mathbf{w}} \quad -\sum_{i=1}^{n'} \ln(\mathbf{w} \cdot (\mathbf{x}_{i1} - \mathbf{x}_{i2})) + \lambda \sum_{i=1}^{p} (w_i - \frac{1}{\lambda} \ln\left(\frac{1}{w_i}\right)), \tag{5}$$

where p is the dimensionality of $\mathbf{w}$ (e.g. the number of attributes of the products) and w_i is the i^{th} coordinate of $\mathbf{w}$. The summation is taken not over all questions n, but over a subset of them n' (note the slight abuse of notation here) which are the "toughest ones" (see (Toubia et al 2004)). We show in the appendix how to also write the estimation method of Toubia et al (2004) in the case of response error. We also discuss in the appendix why it may be that the estimation method of Toubia et al (2004) is sensitive to response error, as we experimentally observe.

2.3 A General Framework for Conjoint Questionnaire Design

In this section we propose a novel general approach for conjoint questionnaire design, applicable in conjunction to the general conjoint estimation method outlined above. This approach is based on the analysis of the effect of a new question on the utility function estimated. Here we consider estimation methods that solve the problem (3):

$$\min_{\mathbf{w}} R_n(\mathbf{w}) = V(\mathbf{w}, \{(\mathbf{x}_{i1}, \mathbf{x}_{i2})\}) + \lambda\Phi(\mathbf{w}), \tag{6}$$

where the error term can be decomposed as a sum of errors:

$$V(\mathbf{w}, \{(\mathbf{x}_{i1}, \mathbf{x}_{i2})\}) = \sum_{i=1}^{n} v\left(\mathbf{w} \cdot (\mathbf{x}_{i1} - \mathbf{x}_{i2})\right),$$

for some function $v : \mathbf{R} \to \mathbf{R}$.

From now on, we write $\mathbf{z}_i = (\mathbf{x}_{i1} - \mathbf{x}_{i2})$ to describe the difference vector between the two proposed products for question i. Notice that the cost function R_n only depends on the $\mathbf{z}_i$'s. Furthermore, when products $\mathbf{x_i}$ are real-valued, we need to only determine $\mathbf{z}_{n+1}$ to generate the next question, and $\mathbf{x}_{(n+1)1}$ and $\mathbf{x}_{(n+1)2}$ can be chosen arbitrarily, so long as their difference is $\mathbf{z}_{n+1}$. Below we develop the theory for real-valued products, and we discuss how to then design products that have binary attributes (or levels), as typically used in practice, in section 2.5 and in the Appendix.

2.3.1 Intuition

Let us assume that n questions have been asked thus far, and we have calculated the cost function R_n based on the respondent's answers and have an estimated utility vector $\mathbf{w}_n$, the minimizer of R_n. In creating the next question, we consider two strategies:

1. We first require, as in (Toubia et al. 2004), *utility balance*. That is, we create a

question about two (or more) products $\mathbf{x}_{n+1,1}$ and $\mathbf{x}_{n+1,2}$ that have *equal utility* based on our *current* estimated utility function. This means that we require that $\mathbf{w}_n \cdot \mathbf{x}_{n+1,1} = \mathbf{w}_n \cdot \mathbf{x}_{n+1,2}$, or more simply, $\mathbf{w}_n \cdot \mathbf{z}_{n+1} = 0$. Therefore, based on our current estimate, the respondent is equally likely to choose either of the two products, and in this sense the question will be most informative.

2. Notice that, in determining $\mathbf{z}_{n+1}$, we are now only restricted to a hyperplane perpendicular to $\mathbf{w}_n$, $\{\mathbf{z} \mid \mathbf{w}_n \cdot \mathbf{z} = 0\}$, according to the first strategy. We need also to consider the direction of $\mathbf{z}$. The second strategy is to *choose a direction for* $\mathbf{z}_{n+1}$ *where the current loss function* $R_n(\mathbf{w})$ *is as flat as possible.* The direction along which $R_n(\mathbf{w})$ is as flat as possible is, intuitively, a direction along which the estimated utility function is not well localized. Therefore a new question along this direction can considerably change the estimated utility function, and in this sense the question will be most informative. We see below that this is a generalization of a similar strategy used by Toubia et al (2003; 2004) from which we got inspired.

We now discuss how to formalize this intuition.

2.3.2 A General Approach for Designing Questions Adaptively

Let us now formulate the two strategies, by first assuming that the loss function v and the complexity function Φ are strictly convex and twice differentiable. Then the function $R_n(\mathbf{w})$ is itself strictly convex and twice differentiable, and the estimated utility vector, i.e., the minimum of $R_n(\mathbf{w})$, is the only point $\mathbf{w}_n$ that satisfies:

$$\nabla R_n\left(\mathbf{w}_n\right) = 0$$

Around that minimum, the "flatness" (or convexity) of R_n in various directions is given by its second derivative matrix (Hessian):

$$\left[\nabla^2 R_n\right]_{i,j} := \frac{\partial^2 R_n}{\partial w_i \partial w_j}.$$

More precisely, the convexity along a direction $\mathbf{z}$ is given by $\mathbf{z}^\top \nabla^2 R_n \mathbf{z}$.

In order to find the direction of smallest convexity (strategy 2) orthogonal to $\mathbf{w}_n$ (strategy 1), we therefore must solve the following problem:

$$\mathbf{min}_{\mathbf{z}} \quad \mathbf{z}^\top \nabla^2 R_n\left(\mathbf{w}_n\right)\mathbf{z} \tag{7}$$

Subject to

$$\mathbf{z}^\top \mathbf{w}_n = 0,$$

$$\mathbf{z}^\top \mathbf{z} = 1,$$

where we added the scaling $\mathbf{z}^\top \mathbf{z} = 1$. By projecting the Hessian matrix onto the hyperplane orthogonal to $\mathbf{w}_n$ by the equation:

$$B_n := \left(I_p - \frac{\mathbf{w}_n \mathbf{w}_n^\top}{\mathbf{w}_n^\top \mathbf{w}_n} \right) \nabla^2 R_n \left(\mathbf{w}_n \right), \tag{8}$$

where p is the dimensionality of $\mathbf{w}$ (the number of attributes of the products) and I_p is the $p \times p$ identity matrix, this problem reduces to finding the eigenvector $\hat{\mathbf{z}}_n$ of B_n (and scaling it) with the *second* smallest eigenvalue ($\mathbf{w}_n$ being itself the direction of the eigenvector with the smallest eigenvalue).

Thus stated, this strategy is very general and can be applied to any estimation procedure of the form (3) as long as (3) is convex and twice differentiable.

2.4 A Novel Method for Conjoint Questionnaire Design

The strategy outlined above can lead to a family of conjoint questionnaire design methods depending on the choice of the estimation method (3), as long as we choose V and Φ so that (3) is convex and twice differentiable. In particular:

- The questionnaire design method of Toubia et al (2003; 2004) can be seen in light of the general framework outlined here. As pointed out by Toubia et al (2003; 2004), the direction of smallest curvature of the cost function they use (equivalent to (5)), is the longest axis of the polyhedron that is defined by the constraints which the answers to the previous questions impose on the feasible space of solutions. This can be estimated through the computation of the eigenvectors of a matrix similar

(Toubia et al 2003; 2004) to the Hessian (8) defined above. We refer the reader to Toubia et al (2004) for the technical details of that method.

- For Regularization networks (RN), defined above as:

$$R_{RN}(\mathbf{w}) = \sum_{i=1}^{n} (\mathbf{w} \cdot (\mathbf{x}_{i1} - \mathbf{x}_{i2}) - 1)^2 + \lambda \mathbf{w}^\top \mathbf{w}.$$

it can be shown (see appendix) that the estimate $\mathbf{w}_n$ after n questions is

$$\mathbf{w}_n = \left(X_n^\top X_n + \lambda \mathbf{I}_p \right)^{-1} X_n^\top \mathbf{1}_n, \tag{9}$$

and the next question is designed along the direction of the *second* smallest eigenvector of:

$$\left(I_p - \frac{\mathbf{w}_n \mathbf{w}_n^\top}{\mathbf{w}_n^\top \mathbf{w}_n} \right) \left(X_n^\top X_n + \lambda \mathbf{I}_p \right), \tag{10}$$

where X_n is the matrix with n rows $X_{n(i)} = (\mathbf{x}_{i1} - \mathbf{x}_{i2})$, the “data matrix” after n questions.

We implemented this 2-steps-at-each-question method and show the experimental results below. It turns out that this is a *very fast method computationally*: notice that the solutions are in closed form and their estimation requires only an inversion of a matrix of size equal to the number of questions asked. This can be done in real time very fast even for a (practically) large number of questions.

To summarize, our conjoint analysis method consists of the following two steps at each iteration n:

1. *Step 1:* The utility function $\mathbf{w}_n$ after n questions are answered is:

$$\mathbf{w}_n = \left(X_n^\top X_n + \lambda \mathbf{I}_p\right)^{-1} X_n^\top \mathbf{1}_n,$$

2. *Step 2:* The next question (difference vector $\mathbf{z}_{n+1}$) is the second smallest eigenvector of:

$$\left(I_p - \frac{\mathbf{w}_n \mathbf{w}_n^\top}{\mathbf{w}_n^\top \mathbf{w}_n}\right)\left(X_n^\top X_n + \lambda \mathbf{I}_p\right)$$

scaled to have square norm 1.

We also fix $\lambda = \frac{1}{n}$ (see section 2.5 for a discussion on this issue).

- As a caveat, however, the question design strategy developed here can not be applied to the SVM cost function (1) which is not twice differentiable. However, we can still use, for example, the SVM estimation method for the questions designed using the proposed RN method.

Before turning to the experiments we briefly discuss now some practical implementation issues.

2.5 Practical Issues

There are a number of practical implementation issues to consider:

- **Designing the first question:** At the very beginning the only available information is typically limited. For example we may only know that the utility function is positive (or we know the least desirable level for each attribute). This, as also observed by Toubia et al (2004), makes the design of the first question random, as the Hessian (10) has many zero eigenvalues. Hence for the first question we design a random question for each individual ignoring the method proposed. This also leads to having the first question such that, across the individuals, all attributes of the products are considered.

- **Designing questions with more than 2 products:** When more than two products per question are needed we consider not only the second eigenvector of the Hessian (8) (or (10) for our method) but also the third, fourth, etc. In the appendix we show, for example, how to do this in the case of four products per question, when we discuss how to transform real-valued questions into binary ones.

- **Choosing parameter λ in (3):** As discussed above (see also (Vapnik 1998)), the "trade off" parameter λ in (3) can be chosen in practice using a small validation set or using cross-validation. However, as we have no data about the individual before the first question, it is not possible to choose parameter λ this way. It should generally be chosen according to how much response error there is (Vapnik 1998). Clearly we cannot have this information before asking any questions. Therefore we make the following choice: we set λ to being $\frac{1}{n}$ where n is the number of questions. Parameter λ needs to decrease as the number of data increases (Vapnik 1998).

We note that one can also choose λ using data from another group of individuals with "similar" response error, when this is available (this would play the role of a validation set).

- **Designing questions with products with binary features:** The approach outlined above assumes that we can choose real-valued questions – the mathematical analysis is done in the continuous space. In practice, however, we may need to design questions with products whose attributes take only a few values (levels). This would require to generate questions with discrete values. For this purpose, we first generate real-valued products with the proposed method, and then we transform these products into discrete (e.g. binary) following exactly the approach of Toubia et al (2004). We describe this in the appendix and refer the reader to Toubia et al (2004) for more details. We chose to replicate the approach of Toubia et al (2004) for this issue in order to avoid any differences in the experimental comparison below that have to do solely with the transformation of the real-valued products into binary ones. We believe that further improvements can be achieved in this step, but this is beyond the scope of this work.

- **Adding extra constraints:** Finally, often it may be the case that we have information about the utility vectors that we can use to constrain the solution. One can do so using virtual examples, as discussed by Evgeniou et al (2004). We note that there are also other approaches and a large literature on how to incorporate constraints to RN and SVM type methods (see for example (Scholkopf et al, 1996)).

3 Experiments

We run Monte Carlo simulations (Carmone and Jain, 1978; Andrews et al 2002; Toubia et al 2004) to compare the performance of the proposed method with:

- a random questionnaire design;
- a standard orthogonal questionnaire design – we used the orthogonal design also used by (Toubia et al 2004);
- a questionnaire designed adaptively for each individual using the polyhedral method of Toubia et al (2004),

under varying conditions of response error (noise) and population heterogeneity. We used the experimental setup of Toubia et al (2003; 2004), which in turn was based on the simulation studies of Arora and Huber (2001). For completeness we briefly describe that setup here.

We designed questionnaires where each question consists of 4 products to choose from. Each product has 4 attributes, and each attribute has 4 levels. Therefore each product is a 16 dimensional binary vector with four 1's (a 1 for each attribute at its "active" level) and twelve 0's. We generated designs with 16 questions per individual for all methods so that the orthogonal design is well-defined. For each experiment we simulated 500 individuals.

The partworths for each individual were generated randomly from a Gaussian with mean $(-\beta, -\frac{1}{3}\beta, \frac{1}{3}\beta, \beta)$ for each of the four 4-level attributes. Parameter β is the magnitude that controls the noise (response error). We used a logistic error model (Ben-Akiva 1985). We followed (Toubia et al 2004) and we used $\beta = 2$ for high magnitude (low noise) and $\beta = 0.5$ for low magnitude (high noise). We modeled heterogeneity among the 500 individuals by varying the variance σ^2 of the Gaussian from which the partworths were generated. The covariance matrix of the Gaussian was a diagonal matrix with all diagonal elements being σ^2. We modeled high heterogeneity using $\sigma^2 = 2\beta$, and low heterogeneity using $\sigma^2 = 0.5\beta$, like in (Toubia et al 2004). As discussed in (Arora and Huber, 2001) and (Toubia et al, 2004) these parameters are chosen so that the range of average partworths and heterogeneity found in practice is covered.

Notice that for each of the four attributes the mean partworths are the smallest for the first level and the largest for the fourth level - in increasing order. Because the method of Toubia et al (2004) uses the information about the lowest level for each attribute (positivity on $\mathbf{w}$ required for the optimization problem (5)), we incorporated this information also for the proposed method as also done in (Evgeniou et al 2004). So for each individual we added constraints that capture the knowledge of the *actual* lowest level for each attribute for the true utility function of the individual. All methods used initially only these positivity constraints to do a first estimate of the utility vectors.

We compare the methods using the Root Mean Square Error (RMSE) of the estimated partworths. Both estimated and true partworths were always normalized for comparability. In particular we followed (Toubia et al 2004): each attribute is made such that the sum of the levels is 0, and the utility vector is then normalized such as the sum of the absolute values for each attribute is 1.

We also tested whether the advantage/disadvantage of the proposed conjoint analysis approach is due to the questionnaire design or the estimation part – or both. We note, however, that it is not clear how one can separate the two issues, as one would expect that the better the estimation method, the better also the questionnaire design within the general framework we developed. We performed the following tests: we used the final questionnaire to estimate the utility vectors $\mathbf{w}$ using an estimation method independent of how the questionnaire was designed. As discussed above, we can also use the SVM-type estimation method of (Cui and Curry 2004; Evgeniou et al 2004) for the final questionnaire. We therefore tested the following:

- Questionnaires used:
  1. A random questionnaire;
  2. An orthogonal questionnaire;
  3. The questionnaire designed adaptively using the method of Toubia et al (2004) (noted with POLY);
  4. The questionnaire designed adaptively using our method (noted with RN)
- Estimation methods used:
  1. The estimation method of Toubia et al (2004);
  2. The RN estimation method;

Mag	Het	Design	POLY estimation	RN estimation	SVM estimation
H	L	Random	0.71	0.54	0.59
		Orthogonal	0.74	0.78	0.69
		POLY	0.47	**0.43**	0.46
		RN	0.47	**0.43**	0.46
L	L	Random	1.39	0.86	0.96
		Orthogonal	0.83	0.89	0.83
		POLY	0.97	0.85	0.96
		RN	0.88	**0.81**	0.84
H	H	Random	0.67	0.53	0.54
		Orthogonal	0.78	0.76	0.66
		POLY	0.41	0.41	**0.40**
		RN	0.41	**0.39**	**0.39**
L	H	Random	1.19	0.77	0.88
		Orthogonal	0.76	0.84	0.77
		POLY	0.83	0.74	0.81
		RN	0.78	**0.72**	0.74

Table 1: Comparison of different questionnaires and estimation methods. Bold indicates best or not significantly different than best at $p < 0.05$ for each (magnitude $\times$ heterogeneity) experiment.

3. The conjoint estimation method of Cui and Curry (2004) and Evgeniou et al (2004) (noted with SVM).

3.1 Experimental Results

We summarize all the results in Table 3.1. Each row is a different questionnaire design, and each column is a different estimation method. From the table we make the following key observations:

1. In all cases the RN estimation method coupled with the RN-based questionnaire design was always the best or not significantly different from the best among all

combinations (estimation × design).

2. The RN-based questionnaire is never worse than the POLY questionnaire for all estimation methods. It is even better than the POLY questionnaire when there is high response error (low magnitude).

3. Both the RN and the SVM estimations lead to significantly better performance than the estimation method of Toubia et al (2004), except for the orthogonal design. As noted by Evgeniou et al (2004), the SVM-type estimation method (like the RN-type one used here) performs relatively poorly when an orthogonal design is used. Moreover, the advantage of the RN and SVM estimations is larger when there is high response error (low magnitude).

4. Overall the improvement of the "complete" (questionnaire design plus estimation) proposed RN-based conjoint analysis method relative to POLY is much larger when the response error is higher (low magnitude).

5. Using an SVM estimation method with the RN questionnaire design does not improve performance further. RN also outperforms SVM in most of the other cases. Performance similarity between RN and SVM estimation methods has been also observed recently by Rifkin et al (2003) in a different context. This indicates – but not conclusively, as other loss functions used by other researchers in the future may lead to different results – that the important part of the proposed approach may be the complexity control Φ and the trade off λ in (3) and not the particular loss

function V used. To the best of our knowledge this is still an open question, and it is beyond the scope of this work to study this issue.

Overall, one of the key finding of the experiments is that the proposed approach handles noise better than POLY. We believe that this is because, as we mentioned in the previous section and discuss in the appendix, the POLY method does not handle response mistakes well. In particular, once a wrong answer is given to some question, from that point on the method of Toubia et al (2004) will always try to find an estimate that is in agreement with the wrong answer given (in the vocabulary of Toubia et al (2004), the method will be searching for a solution in the wrong half-space, that is, the actual utility vector will be "left" on the other side of the hyperplane defined by the question answered wrong). However, with the proposed approach such mistakes are less influential, as also observed by (Evgeniou et al 2004).

4 Hybrid Aggregate and Individual-Specific Questionnaire Customization

Having formulated the problem of designing adaptive individual questionnaires within the general framework in section 2, one can clearly develop more methods by choosing different functional forms for (3). We now discuss how to also consider developing questionnaires *adaptively for each individual but simultaneously for a number of individuals.* One can therefore follow a hybrid approach between aggregate customization (Arora and

Huber 2001; Huber and Zwerina 1996; Sandor and Wedel 2001) and the individual specific customization methods we studied here.

We now show that this can be done through solving an optimization problem of the form (3) for handling heterogeneity, hence within our general framework. In particular, our RN based method (method (4) and equations (9) and (10)) can be extended to handle heterogeneity as proposed by Evgeniou and Pontil (2004).

Let T be the number of individuals and let $\mathbf{w}_n^t$ be the utility function of individual t after n questions. For each question we do (as for the single-individual RN method developed in section 2.4):

- Estimate the T utility functions through the minimization of:

$$\mathbf{min}_{\mathbf{w}_n^t} \sum_{i=1\ldots n} \sum_{t=1\ldots T} (1-\mathbf{w}_n^t \cdot (\mathbf{x}_{i1}^t - \mathbf{x}_{i2}^t))^2 + \lambda(\gamma \frac{1}{T} \sum_{t=1\ldots T} \|\mathbf{w}_n^t\|^2 + (1-\gamma) \frac{1}{T^2} \sum_{i,j=1\ldots T} \|\mathbf{w}_n^i - \mathbf{w}_n^j\|^2) \tag{11}$$

  where $(\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t)$ is the i^{th} question for individual t. Notice that this is again in the form of (3) and it is also convex and twice differentiable. This approach to handling heterogeneity is first developed in (Evgeniou and Pontil 2004) where various properties and extensions are studied. Parameter γ is between 0 and 1 and controls "how much we want the utilities of the individuals to be similar to each other". Notice, for example, that setting $\gamma = 1$, hence removing the term $\sum_{i,j=1\ldots T} \|\mathbf{w}_n^i - \mathbf{w}_n^j\|^2$, takes us to our individual specific RN method as the $\mathbf{w}_n^t$'s are decoupled in (11), while decreasing γ "forces" the individual $\mathbf{w}_n^t$'s to be close to each other. Parameter γ can be chosen using, for example, only a sub-population of the individuals – playing the role of a validation set. In the experiment below we only use 20 individuals (out of the 500) to select parameter γ. We note that this approach has been tested by Evgeniou and Pontil (2004) where it was also shown experimentally using the same simulation setup as well as real data that it often outperforms other heterogeneity methods such as Hierarchical Bayes (Allenby et al 1998; Allenby and Rossi 1999; DeSarbo et al 1997; Lenk et al 1996; Arora et al 1998).

It is easy to see that solving (11), like for the individual specific RN-method (9), requires the inversion of a single matrix of size equal to $Tn \times Tn$. This can be very fast for a small number of questions n and for a small number of people T. Clearly a practical limitation of this approach is that when T is large (typically in practice) we cannot solve (11) in real-time – which defeats our purpose of having a computationally efficient adaptive questionnaire design method. We therefore may have, for example, to either "group" the individuals into small groups and solve (11) for each group, or just solve (11) for all T individuals only at the end of the questionnaires – or both, which is what we did in the experiment below. We believe that exploring variations and studying how to speed up the proposed hybrid questionnaire and estimation approach when we use (11) is an important direction for future work.

- We then take the Hessian of (11) w.r.t. each $\mathbf{w}_n^t$ to design the next question for individual t. It is easy to see that the Hessian of (11) is actually the same as the Hessian in (10) using only the data matrix X_n from the t^{th} individual (there is no non-linear interaction between the $\mathbf{w}_n^t$'s, hence after taking the second derivative the $\mathbf{w}_n^t$'s do not interact with each other). Hence, the next question for individual t is again designed using the eigenvectors of (10) but with replacing the estimate $\mathbf{w}_n^t$ with the solution of (11) insead of the solution of (4).

An extreme scenario in using the proposed hybrid approach is that one requires that all individuals respond simultaneously to their questions (say questions 1-16), where each individual responds to his *individual-specific question at each step.* After each question we estimate all utilities and design all next questions for all individuals simultaneously using the hybrid approach above. The other extreme, since it may not be practical to require that all individuals respond to their questions the same time, is that we do the estimation at an aggregate level using (11) only at the end of the questionnaires designed using our individual specific method. For simplicity we tested here the latter approach by simply estimating all utility vectors simultaneously once all individuals finish responding their individual specific questionnaires. As the estimation of (11) for all 500 individuals was not possible computationally (we used a Matlab implementation that was running out of memory), we solved (11) using only 20 people at a time. As shown by Evgeniou and Pontil (2004), accuracy can further increase if we can combine more people in methods like (11). Moreover, we used 20 of the 500 people to choose parameter γ in (11). Parameter λ was, as before, set to $\frac{1}{n}$, which is $\frac{1}{16}$ in this case. We show the RMSE errors in Table 4. We

Mag	Het	POLY	RN	Hetero Estimation
H	L	0.47	0.43	**0.36**
L	L	0.97	0.81	**0.52**
H	H	0.41	0.39	**0.35**
L	H	0.83	0.72	**0.52**

Table 2: The effects of combining individual specific and aggregate customization: the performance for the RN-based questionnaire coupled with the heterogeneity estimation method (11) is shown in the third column. The first column is the POLY conjoint analysis method, and the second is the proposed individual specific RN based conjoint analysis method. Bold indicates best or not significantly different than best at $p < 0.05$.

notice that there is a very large improvement of the performance, particularly in the high noise (low magnitude) case.

Clearly one can develop variations of this setup by having, for example, individuals answering individually-specific questionnaires independently, but at each question iteration (where each individual has answered n_{ind} questions, with n_{ind} possibly being 0 for some people) we do the estimations simultaneously using (11). The possibilities are endless, and depend on the limitations in practice to coordinate the questionnaires among many individuals as well as the computational efficiency of solving (11). We leave this direction for future work which we believe is very promising, as the experiments above indicate.

5 Conclusions

We presented a general framework for designing questionnaires adaptively at the individual level. After an individual answers one question the next question is computed very quickly – on the fly – based on the individual's responses until that point.

A number of conjoint analysis methods can be developed within the proposed framework. For example the methods of Toubia et al (2003; 2004) can be seen within this framework as special cases. We presented here a method that is in the spirit of the conjoint estimation methods of Cui and Curry (2004) and Evgeniou et al (2004). We compared the proposed method experimentally with random and orthogonal design based conjoint analysis and the method of Toubia et al (2004) using a standard simulation setup. The results show that the proposed method significantly outperforms the other methods *both* in terms of the quality of the questionnaire designed, *and* the estimation of the utility functions, particularly when there is high response error. Furthermore, combining the methods proposed here with the approach for handling heterogeneity of Evgeniou and Pontil (2004) further improves performance, which indicates that future work on the hybrid approach we presented here is very promising.

We finally note that we believe that the proposed framework can be used to further develop theories and methods for conjoint analysis. We note that the proposed framework does not aim by any means to replace existing methods for conjoint analysis, but instead to contribute to the field new tools and methods that can complement existing ones.

References

[1] Allenby, Greg M., Neeraj Arora, and James L. Ginter (1998) "On the Heterogeneity of Demand," Journal of Marketing Research, 35, (August) 384–89.

[2] Allenby, Greg M. and Peter E. Rossi (1999) "Marketing Models of Consumer Heterogeneity", Journal of Econometrics, 89, March/April, p. 57 – 78.

[3] Andrews, Rick, Asim Ansari, and Imran Currim (2002) "Hierarchical Bayes versus finite mixture conjoint analysis models: a comparison of fit, prediction, and partworth recovery", Journal of Marketing Research, 39, p. 87-98, February 2002.

[4] Arora, Neeraj and Joel Huber (2001) "Improving parameter estimates and model prediction by aggregate customization in choice experiments", Journal of Consumer Research, Vol. 28, September 2001.

[5] Arora, Neeraj, Greg Allenby, and James Ginter (1998) "A Hierarchical Bayes Model of Primary and Secondary Demand", Marketing Science, 17,1, p. 29–44.

[6] Ben-Akiva, Moshe and Steven R. Lerman (1985), "Discrete Choice Analysis: Theory and Application to Travel Demand", MIT Press, Cambridge, MA.

[7] Carmone, Frank and Arun Jain (1978), "Robustness of Conjoint Analysis: Some Monte Carlo Results" *Journal of Marketing Research*, 15, p. 300-303.

[8] Cortes, Corinna and Vladimir Vapnik (1995), 'Support Vector Networks'. *Machine Learning* 20, 1–25.

[9] Cui, Dapeng and David Curry (forthcoming), "Predicting Consumer Choice Using Support Vector Machines with Benchmark Comparisons to Multinomial Logit", *Marketing Science*, (forthcoming – Manuscript 3003).

[10] DeSarbo, Wayne and Asim Ansari (1997) "Representing Heterogeneity in Consumer Response Models" Marketing Letters, 8:3, p. 335-348.

[11] Evgeniou, T., M. Pontil, and T. Poggio (2000), "Regularization Networks and Support Vector Machines" *Advances in Computational Mathematics* 13 (2000), p. 1–50.

[12] Evgeniou, Theodoros, Constantinos Boussios, and Giorgos Zacharia (2004) "Generalized Robust Conjoint Estimation", *Marketing Science*, (forthcoming – Manuscript 3081).

[13] Evgeniou, Theodoros, and Massimiliano Pontil (2004), "Regularized Multi-task Learning" In proceeding of the 10^{th} conference on '*Knowledge Discovery and Data Mining*, Seattle, WA, August 2004.

[14] Huber, Joel, and Klaus Zwerina (1996), "The Importance of Utility Balance in Efficient Choice Designs" *Journal of Marketing Research*, 33, p. 307-317.

[15] Lenk, Peter J., Wayne S. DeSarbo, Paul E. Green, and Martin R. Young (1996) "Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs," *Marketing Science*, 15, 173–91.

[16] Rifkin, Ryan, Gene Yeo, and Tomaso Poggio (2003) "Regularized Least Squares Classification," *Advances in Learning Theory: Methods, Model and Applications*, NATO Science Series III: Computer and Systems Sciences, Vol. 190, IOS Press, Amsterdam 2003. Edited by Suykens, Horvath, Basu, Micchelli, and Vandewalle.

[17] Sandor, Zsolt, and Michel Wedel (2001), “Designing Conjoint Choice Experiments Using Managers’ Prior Beliefs” *Journal of Marketing Research*, 38, 4, p. 430-444.

[18] Scholkopf, Bernhard, Chris Burges, and Vladimir Vapnik (1996) “Incorporating Invariances in Support Vector Learning Machines”, Artificial Neural Networks — ICANN’96, Springer Lecture Notes in Computer Science, Vol. 1112, Berlin.

[19] Srinivasan, V. (1998), “A Strict Paired Comparison Linear Programming Approach to Nonmetric Conjoint Analysis” Operations Research: Methods, Models and Applications, Jay E. Aronson and Stanley Zionts (eds), Westport, CT: Quorum Books, p. 97-111.

[20] Srinivasan, V. and Allan D. Shocker (1973), “Linear Programming Techniques for Multidimensional Analysis of Preferences” *Psychometrica*, 38,3, p. 337-369.

[21] Tikhonov, A. N., and V. Y. Arsenin (1977), *Solutions of Ill-posed Problems.* W. H. Winston, Washington, D.C.

[22] Toubia, Olivier, Duncan I. Simester, John R. Hauser, and Ely Dahan (2003), “Fast Polyhedral Adaptive Conjoint Estimation”, *Marketing Science*, 22 (3).

[23] Toubia, Olivier, John R. Hauser, and Duncan I. Simester (2004), “Polyhedral methods for adaptive choice-based conjoint analysis”, *Journal of Marketing Research* (forthcoming).

[24] Vapnik, Vladimir (1998), *Statistical Learning Theory.* New York: Wiley.

A Analysis of the Estimation Method of Toubia et al (2004)

It turns out that the estimation method of Toubia et al (2003; 2004) can be seen within the general framework developed here. Let us first consider the case where the space of feasible solutions (version space) is not empty, that is, there exists utility vectors $\mathbf{w}$ that fulfill all constraints $\mathbf{w} \cdot \mathbf{z}_i \geq 0$ for $i = 1, \ldots, n$, where we use the notation:

$$\mathbf{z}_i = \mathbf{x}_{i1} - \mathbf{x}_{i2}.$$

The utility estimation method proposed by Toubia et al (2003) consists in finding the analytic center of the polyhedron delimited by the constraints on the simplex

$$\left\{ \mathbf{w} \geq 0 : \sum_{i=1}^{p} w_i = 100 \right\},$$

that is, to solve the following problem:

$$\mathbf{min}_{\mathbf{w}} \quad -\sum_{i=1}^{n'} \ln\left(\mathbf{w} \cdot \mathbf{z}_i\right) - \sum_{i=1}^{p} \ln\left(w_i\right) \tag{12}$$

Subject to

$$\mathbf{w} \cdot \mathbf{1}_p = 100,$$

for the n' questions that define the version space. Observe that the barrier function $\ln(\mathbf{w} \cdot \mathbf{z}_i)$ is $-\infty$ if $\mathbf{w} \cdot \mathbf{z}_i \leq 0$, which forces the solution of (12) to be positive and satisfy all constraints $\mathbf{w} \cdot \mathbf{z}_i > 0$ for $i = 1, \ldots, n$.

Forming the Lagrangian for the convex problem (12), we see that there is a number ν (the Lagrange multiplier) such that the solution of (12) solves the following unconstrained problem:

$$\mathbf{min}_{\mathbf{w}} \; -\sum_{i=1}^{n'} \ln(\mathbf{w} \cdot \mathbf{z}_i) - \sum_{i=1}^{p} \ln(w_i) + \nu \mathbf{w} \cdot \mathbf{1}_p. \tag{13}$$

This shows that the estimation method of Toubia et al (2003), in the case when the space of feasible solutions (consistent with the responses to the questions) is nonempty, can indeed be written in the general form of (3) with the error function:

$$V(\mathbf{w}, \{(\mathbf{x}_{i1}, \mathbf{x}_{i2})\}) = -\sum_{i=1}^{n'} \ln(\mathbf{w} \cdot (\mathbf{x}_{i1} - \mathbf{x}_{i2})),$$

and the complexity cost:

$$\lambda \Phi(\mathbf{w}) = -\sum_{i=1}^{p} \ln(w_i) + \nu \mathbf{w} \cdot \mathbf{1}_p = \sum_{i=1}^{p} \{\nu w_i - \ln(w_i)\}.$$

In case that the space of $\mathbf{w}$'s consistent with all the responses is empty, clearly (13) has no solution. In that case, (Toubia et al 2004) propose the following approach. First compute:

$$\delta^\star = \max_{\mathbf{w}} \min_{i} \{\mathbf{w} \cdot \mathbf{z}_i\},$$

and then replace (13) by the following problem:

$$\textbf{min}_{\ \mathbf{w}} \quad -\textstyle\sum_{i=1}^{n'} \ln\left(\mathbf{w} \cdot \mathbf{z}_i + \delta^\star\right) - \sum_{i=1}^{p} \ln\left(w_i\right) \tag{14}$$

Subject to

$$\mathbf{w} \cdot \mathbf{1}_p = 100,$$

This can also be reformulated in the general form (3). Indeed, consider the following problem, for a fixed constant C:

$$\textbf{min}_{\ \mathbf{w},\delta} \quad -\textstyle\sum_{i=1}^{n'} \ln\left(\mathbf{w} \cdot \mathbf{z}_i + \delta\right) - \sum_{i=1}^{p} \ln\left(w_i\right) + C\delta\theta\left(\delta\right) \tag{15}$$

Subject to

$$\mathbf{w} \cdot \mathbf{1}_p = 100,$$

Then it is easy to see that when C becomes large (i.e., tends to $+\infty$), the δ solution of (15) converges to $\delta^\star$, and the $\mathbf{w}$ solution of (15) converges to the solution of (14). Notice that when $C \to \infty$ effectively we *do not find the right trade off* between fitting the data and controlling the complexity $\Phi(\mathbf{w})$ of the estimated utility function – as discussed in

(Vapnik 1998) this trade off is very important. We believe that this may be a reason why in the experiments the method of Toubia et al (2004) is performing relatively poorly when there is response error, as also observed by Evgeniou et al (2004).

We conclude by noticing that if we consider δ as an additional dimension to the unknown vector $\mathbf{w}$, we observe that after taking the Lagrangian reformulation of (15) one can find a Lagrange multiplier ν such that the solution of (15) minimizes the general estimation method (3) with the error function:

$$V\left(\left(\mathbf{w},\delta\right),\left\{\left(\mathbf{x}_{i1},\mathbf{x}_{i2}\right)\right\}\right) = -\sum_{i=1}^{n'} \ln\left(\mathbf{w}\cdot\left(\mathbf{x}_{i1}-\mathbf{x}_{i2}\right)+\delta\right),$$

and the complexity cost:

$$\lambda\Phi\left(\mathbf{w}\right) = \sum_{i=1}^{p}\left\{\nu w_i - \ln\left(w_i\right)\right\} + C\delta\theta\left(\delta\right).$$

B RN Based Questionnaire Design Method

Regularization networks (RN) are defined as:

$$R_{RN}\left(\mathbf{w}\right) = \sum_{i=1}^{n}\left(\mathbf{w}\cdot\left(\mathbf{x}_{i1}-\mathbf{x}_{i2}\right)-1\right)^2 + \lambda\mathbf{w}^\top\mathbf{w}.$$

Simple linear algebra shows that R_{RN} and its derivatives can be written in matrix form as follows:

$$\begin{aligned} R_{RN}(\mathbf{w}) &= \mathbf{w}^\top \left(X^\top X + \lambda \mathbf{I}_p\right) \mathbf{w} - \mathbf{w}^\top X^\top \mathbf{1}_n - \mathbf{1}_n^\top X \mathbf{w} + \mathbf{1}_n^\top \mathbf{1}_n, \\ \nabla R_{RN}(\mathbf{w}) &= 2\left(X^\top X + \lambda \mathbf{I}_p\right) \mathbf{w} - 2X^\top \mathbf{I}_n, \\ \nabla^2 R_{RN}(\mathbf{w}) &= 2\left(X^\top X + \lambda \mathbf{I}_p\right), \end{aligned} \tag{16}$$

where X is the matrix with rows $X_i = (\mathbf{x}_{i1} - \mathbf{x}_{i2})$ (the data matrix).

This shows that this loss function leads to the following estimate of the utility function after n questions:

$$\mathbf{w}_n = \left(X_n^\top X_n + \lambda \mathbf{I}_p\right)^{-1} X_n^\top \mathbf{1}_n,$$

and that the directions selected to form the next question is the second smallest eigenvector of the matrix:

$$\left(I_p - \frac{\mathbf{w}_n \mathbf{w}_n^\top}{\mathbf{w}_n^\top \mathbf{w}_n}\right)\left(X_n^\top X_n + \lambda \mathbf{I}_p\right) = X_n^\top X_n + \lambda \mathbf{I}_p - \frac{\left(X_n^\top X_n + \lambda \mathbf{I}_p\right)^{-1} X_n^\top \mathbf{1}_n \mathbf{1}_n^\top X_n}{\mathbf{1}_n^\top X_n \left(X_n^\top X_n + \lambda \mathbf{I}_p\right)^{-2} X_n^\top \mathbf{1}_n}.$$

where X_n is the matrix after n questions (with n rows), $\mathbf{w}_n$ is the estimate after n questions.

C Designing Binary Products

To transform the real-valued products we design using the proposed method into binary ones, we followed exactly the approach of (Toubia et al 2004) which we briefly review here. In particular, to create 4 binary products for a question as done in the experiments, we start from the 2 designed difference vectors, $\mathbf{v}_1$ and $\mathbf{v}_2$ – which are the second and third smallest eigenvectors of the Hessian matrix (8) scaled so that they have square norm 1 – and our estimate $\mathbf{w}$, and we find a quadrilateral with center w and four corners $\mathbf{w} + \alpha_1 \mathbf{v}_1$, $\mathbf{w} - \beta_1 \mathbf{v}_1$, $\mathbf{w} + \alpha_2 \mathbf{v}_2$, $\mathbf{w} - \beta_2 \mathbf{v}_2$. The α's and $\beta's$ are chosen as the maximum positive real numbers for which the corners are *consistent* with the data points, i.e. $\alpha_1 = \max\{\alpha : \mathbf{z}_i \cdot (\mathbf{w} + \alpha_1 \mathbf{v}_1) \geq 0, \forall i\}$, where the $\mathbf{z}_i$'s are the data points obtained from each of the previous answers. We exclude data points which are misclassified by our estimate, that is, where $\mathbf{w} \cdot \mathbf{z}_i < 0$.

Having obtained $\{\mathbf{c}_1, \ldots, \mathbf{c}_4\}$, we then find a binary vector for each $\mathbf{c}_i$. To this purpose we use, as (Toubia et al 2004) a knapsack problem formulation. We pick a random budget constraint M, and we look for a binary vector $\mathbf{b}_i$ which will correspond to our $\mathbf{c}_i$. The goal is to maximize $\mathbf{b}_i \cdot \mathbf{c}_i$ subject to the constraint that $\mathbf{w} \cdot \mathbf{b}_i \leq M$. This will ensure that each $\mathbf{b}_i$ is as close to $\mathbf{c}_i$, while requiring that each $\mathbf{b}_i$ has $\mathbf{w}$-utility close to M – hence they all have the same utility according to our estimate $\mathbf{w}$. Using this procedure, we obtain a binary $\mathbf{b}_i$ for every i. We check to see if our $\mathbf{b}_i$'s are distinct. If they are, we return these vectors as our four products for the next question. If not, we choose a different M and try again. If we have tried more than k different values of M (where in our simulations k was chosen to be 10, as in Toubia et al (2004)) we stop and we simply use the nondistinct set of $\mathbf{b}_i$'s as our question set – the method in that case is converging as there are no more binary products left to create "informative" questions.