Efficiency of the coordinate-exchange algorithm in constructing exact optimal discrete choice experiments

T. Tian & M. Yang

JOURNAL OF STATISTICAL THEORY AND PRACTICE
2017, VOL. 11, NO. 2, 254–268
http://dx.doi.org/10.1080/15598608.2016.1203842

Efficiency of the coordinate-exchange algorithm in constructing exact optimal discrete choice experiments

T. Tian[a] and M. Yang[a,b]

[a]Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, Illinois, USA; [b]College of Mathematics, Chongqing University of Science and Technology, Chongqing, China

ABSTRACT
The use of discrete choice experiments (DCEs) for modeling real marketplace choices, in both fundamental and applied research, has gained much attention recently. To improve the quality of designing DCEs, most researchers have drawn on optimal design theory. Because of the nonlinearity of the probabilistic choice models, to construct a proper choice design, one needs the help of efficient search algorithms, among which the coordinate-exchange algorithm (CEA) has shown itself to work very well under the widely used multinomial logit discrete choice model. However, due to the discrete nature of the choice design, there are no computationally feasible ways to verify that the resulting design is indeed optimal or efficient. In this article, an approach of evaluating the performance of the CEA for Bayesian optimal designs is proposed. This approach gives a lower bound of the efficiency of the resulting design under the continuous/approximate optimal design framework where well-established mathematical tools and theories can be modified and utilized. Empirical studies show that the CEA is highly efficient for deriving homogeneous optimal designs.

ARTICLE HISTORY
Received 9 August 2015
Accepted 16 June 2016

KEYWORDS
Bayesian optimal design; D-optimality; multinomial logit model; A-optimality; V-optimality

AMS SUBJECT CLASSIFICATION
Primary 62K05; Secondary 62J12, 62P25

1. Introduction

Discrete choice experiments (DCEs), or conjoint choice experiments, have become very popular among researchers and practitioners in studying consumers' preferences since they were first discussed by Green (1974) in his innovative work on the design of choice experiments. Typically, in a DCE, respondents are presented with alternatives/profiles grouped in choice sets where the alternatives/profiles are characterized by a combination of attribute levels of the product or service of interest. Then the choices made by respondents are analyzed to help estimate the appeal of each attribute and further predict the behaviors of consumers when facing options in a real marketplace. Therefore, in order to accurately emulate market decisions and forecast market demands, the issue of designing efficient DCEs is of great concern. For origins of DCEs in conjoint analysis, see Li et al. (2013).

In the problem of designing efficient DCEs, two components need to be carefully examined: the choice of parameter setup in the probabilistic model to which the data are fitted, and the algorithm that helps finding the optimal discrete choice designs under certain optimality criteria.

CONTACT M. Yang myang2@uic.edu Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, MSCS, 851 South Morgan, Chicago, IL 60607, USA.

The challenge of the first issue lies in the fact that the probabilistic choice models are nonlinear, which implies that the efficiency of the design depends on the unknown parameters. Thus, the difficulty in designing an experiment under such a situation is that while one is looking for the best design with the aim of estimating the unknown parameters, one has to know the parameters in the first place to identify the best design. To tackle this circular problem, three approaches have been proposed. The first one, the utility-neutral approach, sets the parameters in a nonlinear model to zeros and thereby transforms it to a linear model. Such locally optimal linear choice designs are then widely used due to their simplicity and well-understood theory (see Anderson and Wiley 1992; Lazari and Anderson 1994; Bunch, Louviere, and Anderson 1996; Grossmann, Holling, and Schwabe 2002; etc.). However, by using a linear model to fit the discrete choice data, one implicitly assumes that the utilities associated with each profile are equal, which suggests the consumers have a nondistinct preference to all the choices of different attribute levels. Since this clearly does not match realistic market situations, the second approach advocates the use of locally optimal designs, which assigns the parameters a set of predetermined nonzero prior values (see Huber and Zwerina 1996; Carlsson and Martinsson 2003; Zwerina, Huber, and Kuhfeld 2005; etc.). Finally, in order to bring the prior knowledge as well as the uncertainty associated with the model parameters into the experiments, Bayesian structures are incorporated into the design of DCEs. This methodology provides us with a sound solution when dealing with the nonlinear choice design setup where it circumvents the trouble of specification of the parameter values by introducing a prior distribution into the formulation of the design problem (see Sandor and Wedel 2001; 2002; Kessels, Goos, and Vandebroek 2006; Gotwalt, Jones, and Steinberg 2009; etc.). In this article, we focus on locally and Bayesian optimal design structures, both of which have been found to outperform the utility-neutral choice designs (Huber and Zwerina 1996; Sandor and Wedel 2001; Kessels et al. 2011).

As to choosing the base statistical choice model, the focus has mainly been on the multinomial logit (MNL) model (McFadden, 1974). As pointed out by Kessels et al. (2011), the basic form of the MNL model provides the foundation for many other complex nonlinear choice models such as the mixed logit model, the latent class model, and the scale heterogeneity model. The construction of Bayesian optimal designs for this family of models require the study of the MNL model as a tool to evaluate their design efficiency. Therefore, we also adopt the MNL model in this article.

The second issue concerns the way of selecting and combining profiles such that maximum information can be elicited from the choice sets they form. In terms of optimal design theory, the choice sets with the largest information matrix "value" are considered as the optimal discrete choice design. Note that "value" of the information matrix here has different meanings under different optimality criteria. For example, D-optimality aims at maximizing (minimizing) the determinant of the information matrix (inverse); A-optimality focuses on minimizing the trace of the inverse of the information matrix, that is, the sum of the variances of the parameter estimates; and G- and V-optimalities look for linear transformations, in this case, the predicted choice probabilities. Note that D- and A-optimalities are more suited for parameter estimation since they consider the variances and covariances of the estimators, whereas G- and V-optimalities are appropriate for predicting customers' behavior. A comparison of these four criteria in DCEs was given by Kessels, Goos, and Vandebroek (2006). They showed that the D-optimality criterion is nearly as good as the latter two with respect to prediction quality but with a much higher computational effectiveness.

Since the MNL model is nonlinear, it is difficult to theoretically construct Bayesian optimal designs. Therefore, one strongly requires the help of an efficient computational search algorithm. Various algorithms have been proposed in the field of finding optimal designs of DCEs, such as the modified Federov algorithm (see Cook and Nachtsheim 1980; Kuhfeld, Garratt, and Tobias 1994; Kuhfeld and Tobias 2005; Kessels, Goos, and Vandebroek 2006; etc.), the RS (relabeling and swapping) algorithm (see Huber and Zwerina 1996), the RSC (relabeling, swapping, and cycling) algorithm (see Sandor and Wedel 2001; 2002; etc.), the quadrature scheme (see Gotwalt, Jones, and Steinberg 2009; Yu, Goos, and Vandebroek 2009; etc.), and the coordinate-exchange algorithm (see Meyer and Nachtsheim 1995; Kessels et al. 2009; Liu and Arora 2011; etc.). Most of the recent work in DCEs is based on the coordinate-exchange algorithm (from here on, referred to as the CEA) due to its substantially improved computation speed. For example, compared with the modified Federov algorithm, where the design from the last step is iteratively improved by exchanging its profiles with profiles from the candidate sets, the CEA changes the previous design on an attribute-to-attribute basis and therefore is candidate-set-free. This property becomes more crucial as the number of attributes or attribute levels increases since it enables the CEA to run in polynomial time while the modified Federov algorithm runs in exponential time (Kessels et al. 2009).

However, although the CEA is widely applied and has shown itself to work well through many simulation studies, there still remains a question as to whether or not the design generated by the algorithm is indeed globally optimal. Because of the discrete nature of the choice design under search, there is simply no mathematical verification to guarantee that the true efficiency of the resulting design is actually 1, or at least close to 1. From the process of the algorithm, all we can be certain about is that the updated design is better than the previous one, and therefore the resulting design is indeed improved compared to the one we start with. In order to guarantee its optimality, one needs to exhaust all possible choice designs in the searching process, which is not practically feasible due to the size of the design space even for a simple DCE setup. For example, if a product with three attributes among which two have three levels and one has two is to be studied, and the choice design is required to be constructed with 8 choice sets each with 3 profiles in it, we will have a total number of $\binom{\binom{3\times3\times2}{3}}{8} > 4\times10^{18}$ possible choice designs, which makes it impossible to exhaustively calculate them all in practice.

To somehow remedy this problem of resulting in a local optimum, multiple tries with different starting designs are used (Kessels, Goos, and Vandebroek 2006; Yu, Goos, and Vandebroek 2009; Liu and Arora 2011). Although design efficiency can be improved through this methodology, computational effort is also increased at the same time. Moreover, the question of exactly how many tries are sufficient for landing an efficient design still remains unknown to this day.

The purpose of this article is thus to provide a justification of the efficiency achieved by the CEA when applied to the optimal discrete choice design problem. More specifically, we evaluate the CEA by giving a lower bound of the efficiency of the resulting design based on the continuous/approximate optimal design framework. Unlike the discrete/exact design space, where an exhaustive search is computationally prohibitive, well-established mathematical theories and tools are ready to be modified and utilized to obtain a fast yet accurate identification of optimal design in the context of an approximate design structure. The Fedorov–Wynn algorithm (FWA; Wynn 1970; Fedorov 1972) and the multiplicative algorithm (MA; Silvey, Titterington, and Torsney 1978) were among the very first algorithms for constructing approximate optimal designs and proved convergence. Since the FWA and the MA both work in a first-order optimization manner (only updates the support points or the design weights), in terms of convergence speed, they are asymptotically slow. As a result, various modifications of these two algorithms have been proposed, among which the Cocktail algorithm (CA; Yu 2011), has shown to be able to increase the convergence speed considerably compared with most of the previous methods. However, these aforementioned algorithms focus mainly on constructing D-optimal designs for the full parameter vector. Some of their adaptations, though they can be extended to be applied under other optimal criteria, are usually not fully developed and may take much longer time to implement. Therefore, we base our work here on a general yet efficient algorithm called the optimal weight exchange algorithm (OWEA; Yang, Biedermann, and Tang 2013). The OWEA is applicable to a large class of optimality criteria, and outperforms the current state-of-the-art algorithms in the sense of convergence speed. More specifically, we extend the OWEA to fit the problem structure in DCEs, and then adopt the celebrated general equivalence theorem (GET) introduced by Kiefer (1974) to verify our conclusion. Later on we will give a detailed description of how exact and approximate choice designs are related to each other and how this relationship helps us justify the efficiency of the CEA. The OWEA has been used in DCE problems before. Liu and Tang (2015) derived optimal/efficient heterogeneous designs under a mixed logit model. They investigated D-optimality and their setup is more or less related to locally optimal designs. In this article, under the standard logit model, we focus on investigating the performance of the CEA for homogeneous choice designs where all respondents receive the same design, which is a common assumption in deriving optimal/efficient choice designs. Utilizing the latest development of the OWEA (Biedermann and Yang 2016), we focus on deriving Bayesian optimal designs instead of locally optimal designs. In addition, we also consider more optimality criteria, including the D-, A-, and V-optimality. As far as we know, this is the first attempt to evaluate the performance of the CEA of deriving Bayesian optimal designs under homogeneous choice design context.

The remainder of the article is organized as follows. Brief reviews of the MNL model and the CEA are presented in section 2. Then we elaborate the rationale of our proposed methodology in section 3. A modified version of the GET and the OWEA in the context of choice design are introduced in section 4. Section 5. focuses on the empirical study and its results. We conclude the article with a summary and discussion in section 6.

2. Review of the multinomial logit model and the coordinate-exchange algorithm

2.1. *Review of the multinomial logit model*

Suppose the product or service of interest has K attributes, each with l_k levels, $k = 1, \ldots, K$. Then we have a total of $L = \prod_{k=1}^{K} l_k$ distinct profiles that constitute the profile candidate set. Suppose the choice design is required to be constructed with S choice sets

$\{C_s|s = 1, ..., S\}$, each with J profiles; then we have a total of $Q = \binom{L}{J}$ possible choice sets and the size of the design space is $M = \binom{Q}{S}$. Let x_{jq} be a $\tilde{K} \times 1$ vector denoting the attribute levels of profile j in the qth choice set. Then these vectors make up the design region $\mathcal{X} = \{\{x_{1q}, ..., x_{Jq}\}|q = 1, ..., Q\}$. The utility a respondent attaches to that profile is thus modeled as

$$u_{js} = x'_{js}\beta + \varepsilon_{js} \tag{1}$$

where β is a $\tilde{K} \times 1$ vector of parameter representing the main effects of the attribute levels on the utility and ε is the extreme value error term. Note that according to the effects-type coding method (Kessels, Goos, and Vandebroek 2006), here $\tilde{K} = \sum_{k=1}^{K} l_k - K$ is the total number of main effect levels subject to the "zero-sum" constraint that makes all parameters estimable.

The MNL probability that a respondent chooses profile j in choice set s amounts to

$$p_{js} = f(\mathrm{E}(u_{js})) = f(x'_{js}\beta) = \frac{\exp(x'_{js}\beta)}{\sum\limits_{i=1}^{J} \exp(x'_{is}\beta)}. \tag{2}$$

The information matrix for one respondent under discrete choice design $\xi = \{C_s|s = 1, ..., S\}$ is then obtained as

$$\mathrm{I}(\xi, \beta) = \frac{1}{S}\sum_{s=1}^{S} \mathrm{I}(X_s, \beta) = \frac{1}{S}\sum_{s=1}^{S} X'_s(P_s - p_s p'_s)X_s \tag{3}$$

where $\mathrm{I}(X_s, \beta)$ is the information matrix for a single choice set C_s in design ξ with $X_s = (x_{1s}, \ldots, x_{Js})'$, $p_s = (p_{is}, \ldots, p_{Js})'$, and $P_s = \mathrm{diag}(p_{is}, \ldots, p_{Js})$.

As mentioned before, we discuss both locally and Bayesian optimal designs. The former setup requires a specification of the parameter values β_0 through knowledge and information obtained from experiences or prestudies. The latter assumes $\pi(\beta)$ as the prior distribution for β, which allows for more uncertainty about the model parameters. The four objective functions of the Bayesian optimality criteria are listed here:

$$\begin{aligned}
D(\xi) &= \int_{\mathrm{R}^{\tilde{K}}} \log(\det(\mathrm{I}^{-1}(\xi, \beta)))\pi(\beta)\mathrm{d}\beta, \\
A(\xi) &= \int_{\mathrm{R}^{\tilde{K}}} \mathrm{tr}(\mathrm{I}^{-1}(\xi, \beta))\pi(\beta)\mathrm{d}\beta, \\
G(\xi) &= \int_{\mathrm{R}^{\tilde{K}}} \max_{x_{jq}\in\mathcal{X}} \mathrm{Var}(\hat{p}_{jq}(x_{jq}, \beta))\pi(\beta)\mathrm{d}\beta = \int_{\mathrm{R}^{\tilde{K}}} \max_{x_{jq}\in\mathcal{X}} c'(x_{jq}, \beta)\mathrm{I}^{-1}(\xi, \beta)c(x_{jq}, \beta)\pi(\beta)\mathrm{d}\beta, \\
V(\xi) &= \int_{\mathrm{R}^{\tilde{K}}}\int_{\mathcal{X}} \mathrm{Var}(\hat{p}_{jq}(x_{jq}, \beta))\pi(\beta)\mathrm{d}x_{jq}\mathrm{d}\beta = \int_{\mathrm{R}^{\tilde{K}}} \sum_{x_{jq}\in\mathcal{X}} c'(x_{jq}, \beta)\mathrm{I}^{-1}(\xi, \beta)c(x_{jq}, \beta)\pi(\beta)\mathrm{d}\beta,
\end{aligned} \tag{4}$$

where ξ stands for the choice design we are evaluating, and $\hat{p}_{jq}(x_{jq}, \beta)$ represents the predicted choice probability for profile x_{jq} with corresponding $\tilde{K} \times 1$ vector,

$$c(x_{jq},\beta)=\frac{\partial p_{jq}(x_{jq},\beta)}{\partial\beta}=p_{jq}\left(x_{jq}-\sum_{i=1}^{J}p_{iq}x_{iq}\right) \tag{5}$$

denoting the first-order truncated Taylor series expansion under model (2). Here and in the following context, for locally optimal designs, we simply remove the outside integral and replace every β with its preestimated value β_0.

2.2. *Review of the coordinate-exchange algorithm*

The CEA can be viewed as a special case of the profile-exchange algorithm. More specifically, for one profile, instead of updating it on a profile basis, which requires evaluating the objective function $\prod_{k=1}^{K} l_k$ times, the CEA changes each attribute in a profile one by one where we only need to evaluate the objective function $\sum_{k=1}^{K} l_k$ times, which is a much smaller number than $\prod_{k=1}^{K} l_k$ and thereby shortens the computation time by a great amount.

For each run of the algorithm, a starting design with J profiles in each of the S choice sets is constructed by randomly choosing from the design space that contains M candidate choice designs. For every profile, each attribute level in that profile is then exchanged with all possible levels of that attribute. A level is updated when the new design results in a smaller criterion value. A complete cycle of iteration terminates when the algorithm has found the best exchange for all attributes of all profiles of the design. After that, the algorithm goes back to the first attribute of the first profile and continues another round until no substitution is made in a whole cycle or some other stopping criterion is met. To avoid local optima, T tries of the algorithm are repeated, each with a different starting design, and the optimal design is chosen from the T resulting designs.

3. Rationale of the approach

The proposed approach makes use of the structure under the continuous/approximate design framework. Unlike an exact discrete choice design, where each choice set either appears once or does not show up at all, an approximate design is based on the weights associated with each of the Q choice sets in the experiment, thus making it independent of the total number of respondents the study allows. In particular, an approximate design in the DCEs context can be denoted as $\tilde{\xi}=\{(C_q,w_q)|q=1,\dots,Q\}$, where C_q is the qth choice set and w_q is its corresponding weight under the constraints that $w_q\in[0,1]$ and $\sum_{q=1}^{Q}=1$. Recall that while an approximate design cannot always be transformed into an exact design, an exact design always has its corresponding approximate form. For example, an exact choice design $\xi^0=\{C_s^0|s=1,\dots,S\}$ is equivalent to its approximate form $\tilde{\xi}^0=\{(C_q^0,w_q)|q=1,\dots,Q\}$ with $w_1=\cdots=w_S=1/S$, and $w_{S+1}=\cdots=w_Q=0$.

The information matrix for an approximate choice design $\tilde{\xi}=\{(C_q,w_q)|q=1,\dots,Q\}$ is defined as $\mathrm{I}(\tilde{\xi},\beta)=\sum_{q=1}^{Q}w_q\mathrm{I}(X_q,\beta)$ where $\mathrm{I}(X_q,\beta)$, as defined under Eq. (3), is the information matrix for a single choice set C_q and $X_q=(x_{1q},\dots,x_{Jq})'$ is its corresponding design vector.

Here are the notations for different designs obtained through different methodologies:

- ξ_S^{CEA}: the resulting exact design (with S choice sets) obtained via the CEA for which the efficiency is of interest in this article.
- ξ_S^*: the exact optimal design that we have no knowledge about and under most situations is impossible to find.
- ξ^*: the approximate optimal design obtained via the OWEA.

With the preceding notation, the definition of A- and V-efficiency of a design ξ is given as follows (Yang, Biedermann, and Tang 2013):

$$\mathrm{Eff}(\xi_S^{CEA}) = \frac{\mathrm{Obj}(\xi_S^*)}{\mathrm{Obj}(\xi_S^{CEA})} \geq \frac{\mathrm{Obj}(\xi^*)}{\mathrm{Obj}(\xi_S^{CEA})} \tag{6}$$

where $\mathrm{Obj}(\cdot)$ is one of the objective criterion functions defined in section 2. As mentioned earlier, a discrete choice design can also be viewed as a special form of approximate design. Thus, the last inequality in Eq. (6) stems from the fact that since ξ^* is the optimal design with the smallest criterion value in the approximate design space that also includes ξ_S^*, certainly we have $\mathrm{Obj}(\xi_S^*) \geq \mathrm{Obj}(\xi^*)$.

Notice that the last ratio in Eq. (6) provides us with a lower bound for the efficiency of design ξ_S^{CEA}, $\frac{\mathrm{Obj}(\xi^*)}{\mathrm{Obj}(\xi_S^{CEA})}$, whose value can be calculated, and thus enables us to circumvent finding ξ_S^*, which is computationally infeasible to obtain as discussed in section 1. Therefore, if the lower bound, $\frac{\mathrm{Obj}(\xi^*)}{\mathrm{Obj}(\xi_S^{CEA})}$, achieves a relatively large value, say 95%, then we know for sure that the efficiency of the resulting design from the CEA, ξ_S^{CEA}, is at least 95%, which implies that the CEA is indeed highly efficient.

On the other hand, the D-efficiency is defined in a slightly different way (Yang, Biedermann, and Tang 2013):

$$\mathrm{Eff}(\xi_S^{CEA}) = \exp\big((\mathrm{Obj}(\xi_S^*) - \mathrm{Obj}(\xi_S^{CEA}))/\tilde{K}\big) \geq \exp\big((\mathrm{Obj}(\xi^*) - \mathrm{Obj}(\xi_S^{CEA}))/\tilde{K}\big), \tag{7}$$

where $\tilde{K}$ is the number of parameters. Notice the log transformation in the D-objective function defined in Eq. (4); the equalities and inequality in Eq. (7) can be established through arguments similar to those for Eq. (6). The meaning of the lower bound for D-efficiency is as the same as that of A- and V-efficiency.

4. Modified version of the general equivalence theorem and the optimal weight exchange algorithm

4.1. *About the general equivalence theorem*

The general equivalence theorem (GET) provides the necessary and sufficient condition to verify whether a continuous design is in fact globally optimal or not. Now we illustrate the modified version of the GET in the context of choice design problem.

Suppose $\xi = \{(C_q, w_q) | q = 1, \ldots, Q\}$ is a general choice design, exact or approximate, where C_q, $q = 1,\ldots,Q$, are all the possible choice sets. We calculate the objective criterion function $\mathrm{Obj}(\xi)$, and compute the directional derivative of this function separately in the direction of each I(Xq), that is, the information matrix for the single choice set C_q. Then a positive derivative, say $d(\xi,C_s)$, implies that the corresponding choice set C_s contains information that is carried inadequately in ξ, and thus its weight w_s should be increased to improve the performance of the current design. Furthermore, if none of the Q derivatives are larger than zero, the proposed design ξ is globally optimal. As we can see, the GET depends on the first-order derivative of the optimality criterion, and the G-optimality criterion, which is calculated through the maximization of $\{c'(x,\beta)\mathrm{I}^{-1}(\xi,\beta)c(x,\beta)\}$, may not be differentiable at some of the design points. Thus, the theorem could not be applied to the G -optimality criterion for now. Therefore, we restrict our attention to the other three criteria in this article. With the help of matrix algebra, we can express the derivatives under different criteria in explicit forms:

$$d(\xi,C_q)=\int_{\mathrm{R}^{\tilde{K}}}\mathrm{tr}\{\mathrm{I}^{-1}(\xi,\beta)\cdot R\cdot\mathrm{I}^{-1}(\xi,\beta)\cdot[\mathrm{I}(X_q,\beta)-\mathrm{I}(\xi,\beta)]\}\pi(\beta)\mathrm{d}\beta \tag{8}$$

with

$$R=\begin{cases}\mathrm{I}(\xi,\beta) & \text{for } D\text{-optimality}\\ \mathrm{I} & \text{for } A\text{-optimality}\\ \sum\limits_{\mathrm{x}\in\mathcal{X}}\mathrm{c}(\mathrm{x},\beta)\mathrm{c}'(\mathrm{x},\beta) & \text{for } V\text{-optimality}\end{cases}$$

where I is the identity matrix; $c(x,\beta)$, as defined in Eq. (5), is the derivative of the predicted probability; and $\mathcal{X}=\{\{x_{1q},\ldots,x_{Jq}\}|q=1,\ldots,Q\}$ is the design region.

Remember that for an exact choice design, in order to verify its optimality, one needs to check all possible designs in the design space with size equal to $M=\binom{Q}{S}$, whereas for an approximate choice design, its optimality can be examined via the GET by only going through all Q possible choice sets, which saves computational time and effort. For the case given in section 1, we can see that $Q=\binom{3\times3\times2}{3}=816<<M=\binom{816}{8}>4\times10^{18}$.

4.2. *About the optimal weight exchange algorithm*

The importance of a general and efficient algorithm to search for optimal designs in scientific studies is inarguable. Here in our simulation studies, we adopt the optimal weight exchange algorithm (OWEA), which was proposed by Yang, Biedermann, and Tang (2013). The OWEA has been shown to be applicable to a wide class of optimality problems: any set of differentiable functions of the parameters of interest; all Φ_p-optimality criteria with p being an integer; and locally or multistage optimal designs. The OWEA works by iteratively updating the choice sets and their corresponding weights until convergence to a globally optimal continuous design is achieved through verification of the GET. During this process, to optimize the weights, instead of relying completely on numeric computation, the OWEA adopts Newton's method, a second-order optimization method that features a quadratic convergence rate, which will increase the speed of convergence to a great extent. For various optimality targets, through applications to many commonly studied nonlinear models, in the sense of convergence speed, the OWEA has been found to be consistently outperforming the existing algorithms, for example, the Cocktail algorithm (Yu 2011).

In the work of Yang, Biedermann, and Tang (2013), the OWEA was implemented under locally optimal design framework. Here in this article, since we consider Bayesian optimal designs, the OWEA may not be directly applied. Recently, Biedermann and Yang (2016) extended the OWEA under the Bayesian optimal design structure. They proved the convergence of the extended OWEA and showed that the extended OWEA enjoys fast computational speed under various situations. The extended OWEA is similar to the original one. The main difference is that one needs to change the derivatives of the objective functions to their integration forms with respect to model parameter β. All the other logic and procedure of the algorithm are similar to the original OWEA. Details can be seen from Biedermann and Yang (2016).

A step-by-step procedure for the implementation of the OWEA in the context of choice design is described as follows:

1. Start by randomly choosing r choice sets from Q candidate sets and assign equal weight of $1/r$ to each one of the selected choice set. This serves as the starting design ξ_0. Note that the value of r does not matter in theory since different resulting continuous optimal designs from different starting designs should all be equivalent in the sense that they all have the same objective criterion values. But we need to mention that the increase of r will cause an increase of difficulty in computing the Hessian matrix embedded in Newton's method, thus slowing down the computation speed. Here we recommend $r \approx S$, the required number of choice set in the problem setup.
2. With ξ^{t-1} being the initial design, update the existent weights in ξ^{t-1} to optimal weights for ξ^t using Newton's method. Elimination of zero, one, or multiple choice sets from ξ^{t-1} may occur during Newton's iterations.
3. For ξ^t, find $C^* = \arg\max_{\{C_q, q=1,\ldots,Q\}}\{d(\xi^t, C_q)\}$ and check whether the value of $d(\xi^t, C^*)$ is less than ϵ, a prespecified threshold. If so, ξ^t is the desired design.
4. Otherwise, update ξ^t by adding C^* into the design with weight zero. This serves as the new initial design to repeat steps 2 and 3.

For design ξ^t with m choice sets, let its initial weight be w_0^t, an $(m-1)\times 1$ vector denoting the weights of the first m−1 choice sets in ξ^t. Note that the dimension is m−1 instead of m since the last one can be obtained by subtracting the summation of the first m−1 weights from one. After j−1 iterations, a detailed procedure for updating w_j^t, utilizing Newton's method, is presented as follows:

(i) Start with $\alpha = 1$,

$$w_j^t = w_{j-1}^t - \alpha\left(\frac{\partial^2 g(\mathrm{I}(\xi_{j-1}^t))}{\partial w \partial w'}|_{w=w_{j-1}^t}\right)^{-1}\frac{\partial g(\mathrm{I}(\xi_{j-1}^t))}{\partial w}|_{w=w_{j-1}^t} \tag{9}$$

with

$$g(\mathrm{I}(\xi_{j-1}^t)) = \begin{cases} \int_{\mathrm{R}^{\tilde{K}}} \log(\det(\mathrm{I}^{-1}(\xi_{j-1}^t), \beta)))\pi(\beta)\mathrm{d}\beta, & \text{for } D\text{−optimality} \\ \int_{\mathrm{R}^{\tilde{K}}} \mathrm{tr}(R\cdot \mathrm{I}^{-1}(\xi_{j-1}^t))\pi(\beta)\mathrm{d}\beta & \text{for } A \text{ −and } V\text{−optimalities} \end{cases}$$

where matrix R for A- and V-optimalities is defined respectively in Eq. (8). Note that here the $g(\cdot)$ function is just a modified version of the criterion/objective function given in Eq. (4).

(ii) Check whether there are nonpositive components in w_j^t. If so, go to step (iv); otherwise, go to step (iii).

(iii) Check whether $||\frac{\partial g(\mathrm{I}(\xi_{j-1}^t))}{\partial w}|_{w=w_j^t}||$, where $||\cdot||$ denotes the Euclidean norm, is less than ϵ', another predetermined cutoff. If so, w_j^t is the desired weight. Otherwise, start the next iteration.

(iv) Reduce α by half, and repeat (i) and (ii) until α reaches ϵ'', also a prespecified limit. Then remove the choice set with the smallest weight; go back to $\alpha = 1$ and repeat the preceding procedure with the new set of choice sets as well as their weights.

Note that here the three thresholds, ϵ, ϵ', and ϵ'', are set to 10^{-6} in the empirical study in the next section.

5. Empirical study

We illustrate the proposed approach using several sets of design examples. The first two come from Kessels, Goos, and Vandebroek (2006), where two similar attribute structures with two attributes at three levels and one at two levels are discussed. Thus, we have a total of $L = 3 \times 3 \times 2 = 18$ distinct profiles, which constituted the profile candidate set. Designs from the first class consist of 12 choice sets each of size 2, while designs from the second class consist of 8 choice sets each of size 3. We demonstrate these two designs using $3 \times 3 \times 2/2/12$ and $3 \times 3 \times 2/3/8$, respectively.

Using the effects-type coding method mentioned in section 2.1, here the dimension of the parameter β in the MNL model is $\tilde{K} = 3 + 3 + 2 - 3 = 5$. For locally optimal designs, β was set to be $(-1, 0, -1, 0, -1)'$ so that the attribute levels were equally spaced between −1 and 1. For Bayesian optimal designs that involve integration with respect to the model parameter β, numeric approximation is necessary since the computations cannot be accomplished analytically. Several approaches have been proposed regarding this matter: Sloan and Joe (1994) discussed lattice point methods, which are related to quadrature estimates; Sandor and Wedel (2002) used samples based on orthogonal arrays and later constructed quasi-Monte Carlo samples (Sandor and Wedel 2005). Since our goal here is to justify the efficiency of the CEA in general, which incorporates the idea of deriving an efficient design under any prior setup, we do not compare these various ways of integration approximation, but directly adopt the "small designed sample of parameter vectors" method given in Kessels et al. (2009, Table 3) that has been shown to have better performance in terms of saving computational efforts. Twenty representative points were selected using a minimum potential design and were assigned with equal weights, that is, 1/20. Therefore, instead of focusing on the problem of parameter setup in the CEA staring procedure, we put more of our attention on the establishment of the high efficiency of its resulting design.

Setting the initial design size to be 10 in the OWEA and the number of tries to be 200 in the CEA, we got the results shown in Table 1 regarding the lower bound of the CEA efficiency calculated by Eq. (6).

Table 1. Lower bound of the CEA efficiency under different design setups and criteria.

	Locally optimal design		Bayesian optimal design	
	$3 \times 3 \times 2/2/12$	$3 \times 3 \times 2/3/8$	$3 \times 3 \times 2/2/12$	$3 \times 3 \times 2/3/8$
D-optimal	99.65%	99.80%	98.71%	98.82%
A-optimal	99.04%	99.11%	96.83%	97.15%
V-optimal	98.69%	97.52%	96.75%	96.85%

We can see that the lower bounds of the efficiency were above 95% under all scenarios, and some nearly reached 100%. In order to illustrate more clearly, we present in Table 2 the comparison of the two resulting discrete choice designs from the OWEA and the CEA under setup $3 \times 3 \times 2/3/8$, and the optimal criterion was chosen to be D-optimality.

For each design, the first column is the index for the design set; the second column contains the actual attribute levels of all the profiles that are contained in each choice set. We should mention that the actual optimal approximate design with 22 choice sets from the OWEA (not shown here) was converted to an exact choice design by picking out the first eight choice sets with the largest weights. Note that the largest weight for a choice set in the optimal approximate design that was not used was around 0.043; the total weight of the choice sets that were not used was around 0.289.

We can see that out of all the eight choice sets, four are exactly the same; each of the remaining choice sets has at least one common profile with the others. This direct comparison shows, from another angle, that the resulting design from the CEA closely resembles the optimal design, which might be expected given the high efficiency of the algorithm.

Approximate optimal design does not depend on S, the number of choice sets in the exact design. On the other hand, designs derived through CEA do depend on S, which brings up the question of how the value of S affects the lower bound of the efficiency. We

Table 2. Comparison of the two resulting discrete choice designs.

OWEA		CEA	
1	{1 1 2}	1	{1 3 2}
	{2 2 1}		{2 2 1}
	{3 3 1}		{3 1 1}
2	{1 1 1}	2	{1 1 1}
	{2 2 2}		{2 2 2}
	{3 3 1}		{3 3 1}
3	{1 2 2}	3	{1 2 2}
	{2 3 1}		{2 1 2}
	{3 1 1}		{3 3 1}
4	{1 3 2}	4	{1 3 2}
	{2 1 2}		{2 1 2}
	{3 2 1}		{3 2 1}
5	{1 3 2}	5	{1 3 2}
	{2 1 1}		{2 1 1}
	{3 2 2}		{3 2 2}
6	{1 2 1}	6	{1 1 2}
	{2 3 2}		{2 3 2}
	{3 1 2}		{3 1 2}
7	{1 2 1}	7	{1 3 1}
	{2 1 2}		{2 2 2}
	{3 3 1}		{3 3 1}
8	{1 1 1}	8	{1 1 1}
	{2 2 1}		{2 2 1}
	{3 3 2}		{3 3 2}

Table 3. Lower bound of the CEA efficiency under design structure $3 \times 3 \times 2/3/16$.

	Locally optimal design	Bayesian optimal design
D-optimal	99.92%	99.63%
A-optimal	99.56%	99.35%
V-optimal	99.00%	99.49%

thus alter the second class setup by doubling the number of choice sets in the study, which gives us a design structure of $3 \times 3 \times 2/3/16$. The efficiency table is shown as Table 3.

Comparing the two columns under structure $3 \times 3 \times 2/3/8$ in Table 1, the lower bounds of efficiency increased, especially in the Bayesian case. This could be explained by the rationale of our approach. Figure 1 shows the comparison of these two structures. First, the whole squares represent the same continuous design space they share, which is solely determined by the first four arguments in the setup, that is, $3 \times 3 \times 2/3$. Then the discrete design space consists of all the intersection points inside the square. This is because every exact design is equivalent to its corresponding approximate form, and thus the discrete design space is a subspace of the approximate one. Since the design space of structure $3 \times 3 \times 2/3/8$ has fewer candidate designs than structure $3 \times 3 \times 2/3/16$, that is, $\binom{\binom{18}{3}}{8} < \binom{\binom{18}{3}}{16}$, the square of the former contains fewer points than the latter. That is to say, the CEA will have a coarser grid in the search of optimal discrete choice design. Adopting notations from section 2.1, suppose we have:

- $\tilde{\xi}_0^* = \{(C_q, w_q) | q = 1, \ldots, Q\}$ with $w_q \in [0, 1]$ and $\sum_{q=1}^{Q} = 1$: the optimal continuous design under setup $3 \times 3 \times 2/3$ that can be obtained through implementing the OWEA (solid point in Figure 1).
- $\tilde{\xi}^1 = \{(C_1^1, 1/8), \ldots, (C_8^1, 1/8)\}$: the resulting design from the CEA under setup $3 \times 3 \times 2/3/8$ (hollow point in Figure 1a).
- $\tilde{\xi}^2 = \{(C_1^2, 1/16), \ldots, (C_{16}^2, 1/16)\}$: the resulting design from the CEA under setup $3 \times 3 \times 2/3/16$ (hollow point in Figure 1b).

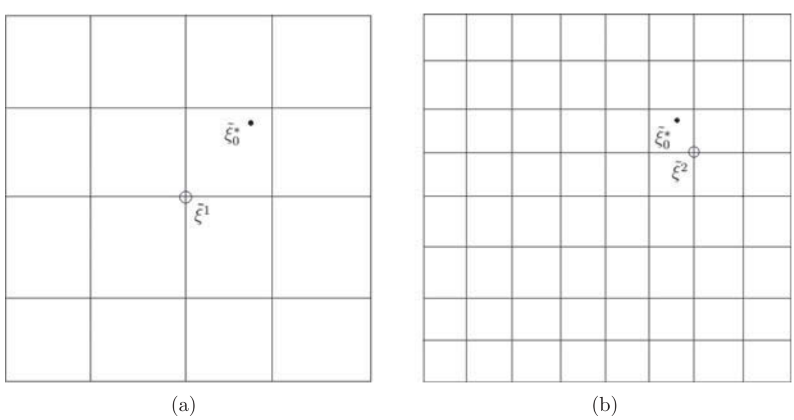

Figure 1. Graphic comparison: (a) Under setup $3 \times 3 \times 2/3/8$. (b) Under setup $3 \times 3 \times 2/3/16$.

Table 4. Lower bound of the CEA efficiency under extended design setups.

	$3 \times 3 \times 2 \times 2 \times 2/3/18$		$4 \times 3 \times 3 \times 2/3/15$	
	Locally	Bayesian	Locally	Bayesian
D-optimal	97.73%	98.97%	97.85%	99.04%
A-optimal	94.90%	98.11%	96.70%	97.57%
V-optimal	96.78%	98.86%	96.32%	98.22%

It can be easily seen that when the candidate discrete designs are more crowded together, the resulting design from the CEA would have a larger chance to come nearer to the optimal continuous design. As illustrated in Figure 1, point $\tilde{\xi}^2$ is closer to point $\tilde{\xi}_0^*$ than $\tilde{\xi}^1$, which explains the higher efficiency achieved by the CEA when we increase the value of S.

We now expand our simulation study to more complicated setups. We consider the following two designs: $3 \times 3 \times 2 \times 2 \times 2/3/15$ and $4 \times 3 \times 2 \times 2/3/12$. Compared with one of our previous examples, $3 \times 3 \times 2/3$, the number of parameters increases from 5 to 7 and 8, respectively, and the size of the design space increases from $\binom{3\times3\times2}{3} = 816$ to $\binom{3\times3\times2\times2\times2}{3} = \binom{4\times3\times3\times2}{3} = 59640$. Similarly, we explored both locally optimal designs with parameter priors being set to $(-1, 0, -1, 0, -1, -1, -1)$ and $(-1, 0, 1, -1, 0, -1, 0, -1)$, respectively, and Bayesian optimal designs with parameter priors (not shown here) being set to 30 equally weighted representative points selected using minimum potential designs.

Setting the initial design size to be 10 in the OWEA and the number of tries to be 100 in the CEA, we present the results in Table 4. Other than the only entry that fell slightly below 95% (94.90%), all of the others were still beyond 95%; since the real CEA efficiency ought to be greater than the lower bound we simulated, we can again conclude that the expanded simulation examples demonstrated that the CEA is truly highly efficient under various design setups.

All previous examples are under the scenario that the resulting approximate optimal designs are not also exact; we next consider a scenario where the resulting approximate and exact optimal design coincide with each other. Adopting conclusions given in Graßhoff et al. (2004), we examine the designs generated by the CEA under experiment setups $3 \times 3 \times 3 \times 3/2/9$ and $2 \times 2 \times 3/2/24$, where orthogonal arrays OA(9, 3^4, 2) and OA(24, 2^2×3, 2) are theoretically D-optimal for these two setups, respectively.

We run the CEA as well as the OWEA under these two setups, and the designs we get can indeed be converted to the corresponding OAs (details about the conversion between an exact design and its OA form can be found in Graßhoff et al. [2004, section 6]). That is to say, the resulting designs from the CEA and the OWEA agree with each other in the sense that their efficiencies both reach to 100%. In other words, the CEA does possess the ability to pick out the true optimal exact design under certain scenarios.

6. Summary and discussion

Most of the extant research on DCEs uses the CEA to help derive a discrete choice design that may later be used in market decision making or demand predicting, which renders the efficiency of the underlying algorithm crucial to the modeling of real marketplace behaviors. Though the CEA has been shown to be computationally economical, especially under the design setup where the number of attributes and attribute levels are large, the efficiency of its resulting design still remains a question. In other words, due to the discrete nature of the DCEs, in order to verify its optimality, one has to check its critical value against all the other possible choice designs in the design space, which, to this day, is computationally unfeasible to achieve. Therefore, in this article, we provide another approach to establish the efficiency of the CEA by utilizing the relationship between exact designs and approximate designs.

Since we already have access to the optimal approximate design in a DCE problem via the use of the OWEA, we can convert the resulting design from the CEA to its corresponding approximate form and calculate the relative efficiency in the continuous design space. The approach yields a lower bound of the CEA efficiency. From the empirical study, Table 1 demonstrates that, under all experiment setups and all of the three criteria we care about, the relative efficiency is at least 95%. Moreover, the follow-up exhibition in Table 3 shows that the CEA performance would actually be enhanced when there are more choice sets involved in the study, all of which concludes our effort in proving that under any common DCE setup, the resulting design from the CEA comes very close to the actual optimal one in the discrete design space.

We notice that the performance of the CEA in Liu and Tang (2015) is not as good as demonstrated in this article. This difference is very likely due to the distinction in design setups. Liu and Tang (2015) studied heterogeneous choice designs where different respondents or groups of respondents get different designs, while in this article we consider homogeneous designs where the design is the same for each respondent. All in all, under homogeneous design setup, we recommend the CEA due to its high efficiency and level of practicability.

Acknowledgment

The authors are thankful for the detailed comments and suggestions made by the editors and referees on previous versions of the article, which clearly helped to improve the article.

References

Anderson, D. A., and J. B. Wiley. 1992. Efficient choice set designs for estimating cross-effects models. *Marketing Letters* 3 (4):357–70.

Biedermann, S., and M. Yang. 2016. Numerical construction of parameter robust optimal designs: A new efficient algorithm. In preparation.

Bunch, D. S., J. J. Louviere, and D. A. Anderson. 1996. A comparison of experimental design strategies for multinomial logit models: The case of generic attributes. University of California Davis Graduate School of Management Working Paper, 11–96. Davis, CA.

Carlsson, F., and P. Martinsson. 2003. Design techniques for stated preference methods in health economics. *Health Economics* 12 (4):281–94.

Cook, R. D., and C. J. Nachtsheim. 1980. A comparison of algorithms for constructing exact *D*-optimal designs. *Technometrics* 22 (3):315–24.

Fedorov, V. V. 1972. *Theory of optimal experiments*. New York, NY: Academic Press.

Graßhoff, U., H. Großmann, H. Holling, and R. Schwabe. 2004. Optimal designs for main effects in linear paired comparison models. *Journal of Statistical Planning and Inference* 126 (1):361–76.

Gotwalt, C. M., B. Jones, and D. M. Steinberg. 2009. Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics* 51 (1):88–95.

Green, P. E. 1974. On the design of choice experiments involving multifactor alternatives. *Journal of Consumer Research* 1 (2):61–68.

Grossmann, H., H. Holling, and R. Schwabe. 2002. Advances in optimum experimental design for conjoint analysis and discrete choice models. *Advances in Econometrics* 16:93–118.
Huber, J., and K. Zwerina. 1996. The importance of utility balance in efficient choice designs. *Journal of Marketing Research* 33 (3):307–17.
Kessels, R., P. Goos, and M. Vandebroek. 2006. A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research* 43 (3):409–19.
Kessels, R., B. Jones, P. Goos, and M. Vandebroek. 2009, An efficient algorithm for constructing Bayesian optimal choice designs. *Journal of Business and Economic Statistics* 27 (2):279–91.
Kessels, R., B. Jones, P. Goos, and M. Vandebroek. 2011. The usefulness of Bayesian optimal designs for discrete choice experiments. *Applied Stochastic Models in Business and Industry* 27 (3):173–88.
Kiefer, J. 1974. General equivalence theory for optimum designs (approximate theory). *Annals of Statistics* 2:849–79.
Kuhfeld, W. F., M. Garratt, and R. D. Tobias. 1994. Efficient experimental design with marketing research applications. *Journal of Marketing Research* 31 (4):545–57.
Kuhfeld, W. F., and R.D. Tobias. 2005, Large factorial designs for product engineering and marketing research applications. *Technometrics* 47 (2):132–41.
Lazari, A. G., and D. A. Anderson. 1994. Designs of discrete choice set experiments for estimating both attribute and availability cross effects. *Journal of Marketing Research* 31 (3):375–83.
Li, W., C. J. Nachtsheim, K. Wang, R. Reul, and M. Albrecht. 2013. Conjoint analysis and discrete choice experiments for quality improvement. *Journal of Quality Technology* 45 (1):74–99.
Liu, Q., and N. Arora. 2011. Efficient choice designs for a consider-then-choose model. *Marketing Science* 30 (2):321–38.
Liu, Q., and Y. Tang. 2015. Construction of heterogeneous conjoint choice designs: A new approach. *Marketing Science* 34 (3):346–66.
McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in economics*, ed. P. Zarembka, 105–42. New York, NY: Academic Press.
Meyer, R. K., and C. J. Nachtsheim. 1995. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics* 37 (1):60–69.
Sándor, Z., and M. Wedel. 2001. Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research* 38 (4):430–44.
Sándor, Z., and M. Wedel. 2002. Profile construction in experimental choice designs for mixed logit models. *Marketing Science* 21 (4):455–75.
Sándor, Z., and M. Wedel. 2005. Heterogeneous conjoint choice designs. *Journal of Marketing Research* 42 (2):210–18.
Silvey, S. D., D. M. Titterington, and B. Torsney. 1978. An algorithm for optimal designs on a finite design space. *Communications in Statistics—Theory and Methods* 14 (7):1379–89.
Sloan, I. H., and S. Joe. 1994. *Lattice methods for multiple integration*. New York, NY: Oxford University Press.
Wynn, H. P. 1970. The sequential generation of D-optimal experimental designs. *Annals of Mathematical Statistics* 41 (5):1655–64.
Yang, M., S. Biedermann, and E. Tang. 2013. On optimal designs for nonlinear models: A general and efficient algorithm. *Journal of the American Statistical Association* 108 (504):1411–20.
Yu, J., P. Goos, and M. Vandebroek. 2009. Efficient conjoint choice designs in the presence of respondent heterogeneity. *Marketing Science* 28 (1):122–35.
Yu, Y. 2011. D-optimal designs via a cocktail algorithm. *Statistics and Computing* 21 (4):475–81.
Zwerina, K., J. Huber, and W. F. Kuhfeld. 2005. A general method for constructing efficient choice designs (updated version). SAS Technical Document TS-722E. http://support.sas.com/techsup/technote/mr2010e.pdf.