European Journal of Operational Research 226 (2013) 115–121

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

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

Innovative Applications of O.R.

Optimal resource allocation in survey designs

Melania Calinescu [a,*], Sandjai Bhulai [a], Barry Schouten [b]

[a] VU University Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
[b] Statistics Netherlands, Henri Faasdreef 312, 2492 JP The Hague, The Netherlands

A R T I C L E I N F O

Article history:
Received 3 April 2012
Accepted 29 October 2012
Available online 14 November 2012

Keywords:
Resource allocation
Nonlinear optimization
Dynamic programming
Markov decision theory
Survey design
Scheduling

A B S T R A C T

Resource allocation is a relatively new research area in survey designs and has not been fully addressed in the literature. Recently, the declining participation rates and increasing survey costs have steered research interests towards resource planning. Survey organizations across the world are considering the development of new mathematical models in order to improve the quality of survey results while taking into account optimal resource planning. In this paper, we address the problem of resource allocation in survey designs and we discuss its impact on the quality of the survey results. We propose a novel method in which the optimal allocation of survey resources is determined such that the quality of survey results, i.e., the survey response rate, is maximized. We demonstrate the effectiveness of our method by extensive numerical experiments.

1. Introduction

Surveys are used all around the world to measure socio-economic status and well-being of people, to test theories, or to make investment decisions, driven by the impossibility of observing the entire population of interest (see [6]). No matter what the framework of a survey is, its success relies on the active participation of the sampled households and businesses. Nonresponse occurs when members of a sample cannot or will not participate in the survey. The impact of nonresponse appears in the inability of computing a full-sample estimator of the population mean. Thus, a bundle of practical issues is created, including bias in point estimates, bias in estimators of precision, and inflation of the variance of point estimators. The error caused by nonresponse is one of the several sources of error in surveys and it has attracted a great deal of interest among researchers across the world (see [6]). An apparent solution to the problem is to increase the frequency of attempts to gather information from reluctant sample members. Under these circumstances, the costs of conducting surveys increase significantly, which leads to new problems, such as budget overruns. Therefore, a constant scientific challenge to the survey community concerns developing new survey designs to accommodate the presence of both nonresponse and high costs.

Modeling the bundle of processes behind a survey and understanding the numerous interactions between these processes have been a constant obstacle for researchers in their attempts to design quality but cost-effective surveys. As a consequence, only few processes have been investigated from a cost perspective, e.g., call scheduling in [9]. More advanced studies investigate the relationship between costs, quality and few survey features (e.g., the interview mode, the schedule of calls). For example, in [7,11], the main idea is to identify a set of design features that potentially influence the survey costs and errors in the estimates and to monitor them throughout the survey run. This information helps in subsequent phases to alter the design features such that a desired balance between costs and errors is achieved.

When person or household characteristics (e.g., social and financial) are employed to adjust the design features to a given set of characteristics (i.e., different design features can be applied to sample units with different characteristics) the resulting survey design is termed *adaptive* (introduced in [16,14]). Adaptive designs render realistic survey models and can be used to capture the interactions between survey features, sample unit characteristics, survey costs and quality.

In the present paper, adaptive survey designs are analyzed from the perspective of resource allocation problems. To our knowledge, this is the first paper that addresses designing surveys from a resource allocation perspective. Given a budget, a set of household characteristics, and a list of survey features that influence costs and quality, we model the allocation of survey resources such that quality is maximized while costs meet the budget constraint. Our interest in the problem is motivated by the increased difficulty (e.g., higher costs, and higher nonresponse) survey organizations are faced with in order to obtain high-quality survey estimates. Statistics Netherlands is among the first organizations to consider

* Corresponding author. Address: VU University Amsterdam, Faculty of Sciences, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands. Tel.: +31 205987687.
E-mail address: melania.calinescu@vu.nl (M. Calinescu).

0377-2217/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2012.10.046

redesigning their surveys such that planning of resources is taken into account.

Resource allocation problems can be found in a wide range of applications. In [12], the author investigates applications where the resource allocation can be modeled as a continuous convex nonlinear problem. Algorithms to solve such problems are also surveyed and they most often involve finding the optimal value of the Lagrange multiplier for the explicit constraint (mainly through some type of line search). There are also numerous applications that require a relaxation of the condition on strict convexity and differentiability of the cost constraints, which increases significantly the complexity of the problem. The auction algorithm presented in [1] finds a near optimal solution of this problem in finite time.

In its integer or mixed-integer formulation, the resource allocation problem has an NP-complete worst case complexity (see [8]). Therefore, only few such applications have been addressed in the literature, e.g., optimal sample allocation in stratified sampling (see [10,3]), manufacturing capacity planning (see [4]). The proposed algorithms take advantage of the convexity in the objective function and/or constraints. Applications where the objective function and/or the constraints include separable nonconvex functions are often encountered (e.g., due to economies of scale). In this case, additional difficulties in solving the problem are posed by the presence of several local optima. In [5], an approach is suggested to solve such problems, namely solve a convex lower bounding problem (e.g., the convex envelope) at every node of the branch-and-bound search tree. Using the branch-and-bound framework developed by [2], the optimal solution is reached in a finite number of iterations. However, no implementation results or optimality gap assessments are reported.

The resource allocation problem for survey designs has specific features that lead to a formulation as a nonconvex integer nonlinear problem, which prohibits the application of many algorithms that are found in the literature. A possible approach could be to implement solutions of convex approximations of the problem, however, this may result in major errors in survey estimates. We present an algorithm that solves the problem to optimality using Markov decision theory. The algorithm reaches the optimal solution in a finite number of iterations. The numerical experiments discussed here displayed short computational times on an Intel Xeon L5520 processor.

The remainder of the paper is structured as follows. Section 2 discusses the mathematical model and Section 3 discusses the algorithm to derive optimal adaptive survey design policies. Section 4 presents a range of practical problems that can be handled through this model and solution method. Numerical examples of these situations are given in Section 5. Section 6 concludes the results of the paper and gives directions for future research.

2. Problem formulation

Consider a survey sample consisting of N units that can be clustered into homogeneous groups based on characteristics, such as age, gender, and ethnicity (information that can be extracted from external sources of data). Let $\mathcal{G} = \{1, \ldots, G\}$ be the set of homogeneous groups with size N_g for group $g \in \mathcal{G}$ in the survey sample. The survey fieldwork is divided into time slots, denoted by the set $\mathcal{T} = \{1, \ldots, T\}$, at which units in a group can be approached for a survey. The survey itself can be conducted using certain interview modes, such as a face-to-face, phone, web/paper survey; the set of different modes is denoted by $\mathcal{M} = \{1, \ldots, M\}$. At each time slot $t \in \mathcal{T}$ one can decide to approach units in group $g \in \mathcal{G}$ for a survey using mode $m \in \mathcal{M}$. In doing so, successful participation in the survey depends on first establishing contact, and then be responsive by answering the questionnaire. From historical data group-dependent contact probabilities $p_g(t,m)$ and participation probabilities $r_g(t,m)$ can be estimated, which we consider as given quantities in our problem. Note that from historical data it can also be observed that certain time slots (e.g., morning, evening) have an influence on the availability of the unit and the willingness to respond. Therefore, to employ most of the available information, the contact and response probabilities are modeled at the level of time slots for each group as well rather than the mode only.

Denote by $x_g(t,m)$ a binary 0–1 decision variable that denotes if units in group g are approached for a survey at time t using mode m. Note that at time t only one mode can be employed to approach a group, yielding the constraint $\sum_{m\in\mathcal{M}} x_g(t,m) \leqslant 1$. When a successful contact is established and the unit agrees to participate, the survey ends with success; this happens with probability $p_g(t,m)$ $r_g(t,m)$. Note that we assume independence between participation and contact. However, if the unit refuses participation after successful contact, the unit is not considered for a future survey approach; this happens with probability $p_g(t,m)(1 - r_g(t,m))$. Only in the case that the unit is not contacted successfully, the unit can be considered for a future survey approach (see Fig. 1); this happens with probability $1 - p_g(t,m)$. Thus, if the unit is approached again at time t' using mode m', then the probability of a successful approach is $(1 - p_g(t,m))p_g(t',m')r_g(t',m')$, and the probability of a contact failure is $(1 - p_g(t,m))(1 - p_g(t',m'))$. In general, the probability that a contact fails up to time t' is denoted by $f_g(t')$ given by

$$f_g(t') = \prod_{t=1}^{t'} \prod_{m\in\mathcal{M}} [x_g(t,m)(1 - p_g(t,m)) + (1 - x_g(t,m))]$$
$$= \prod_{t=1}^{t'} \prod_{m\in\mathcal{M}} [1 - x_g(t,m)p_g(t,m)].$$

Note that this is a highly non-linear expression in the decision variables, which can be recursively computed by

$$f_g(t') = \prod_{m\in\mathcal{M}} [x_g(t',m)(1 - p_g(t',m)) + (1 - x_g(t',m))]f_g(t' - 1)$$
$$= \prod_{m\in\mathcal{M}} [1 - x_g(t',m)p_g(t',m)]f_g(t' - 1), \quad (1)$$

using the fact that $f_g(0) = 1$. Using this definition, the response rate for group g can then be computed by

$$\sum_{t\in\mathcal{T}} \sum_{m\in\mathcal{M}} f_g(t-1)\ x_g(t,m)p_g(t,m)r_g(t,m).$$

The clustering of the N units usually results in groups that are not of the same size or importance. Therefore, the response rates for the groups are usually weighted by a factor w_g (e.g., $w_g = N_g/N$ is taken in practice). Hence, the objective of the decision maker becomes to maximize

$$\sum_{g\in\mathcal{G}} \sum_{t\in\mathcal{T}} \sum_{m\in\mathcal{M}} w_g f_g(t-1)x_g(t,m)p_g(t,m)r_g(t,m), \quad (2)$$

by setting the decision variables $x_g(t,m)$ optimally. The decision variables are subject to constraints, though, due to scarcity in resources. In practice, due to resource management constraints, the number of times that a group can be approached by mode m is limited to $\bar{k}_g(m)$ times, leading to the constraint $\sum_{t\in\mathcal{T}} x_g(t,m) \leqslant \bar{k}_g(m)$. By combining the objectives with all the constraints, we can draft our optimization problem as a binary programming problem in the following manner.

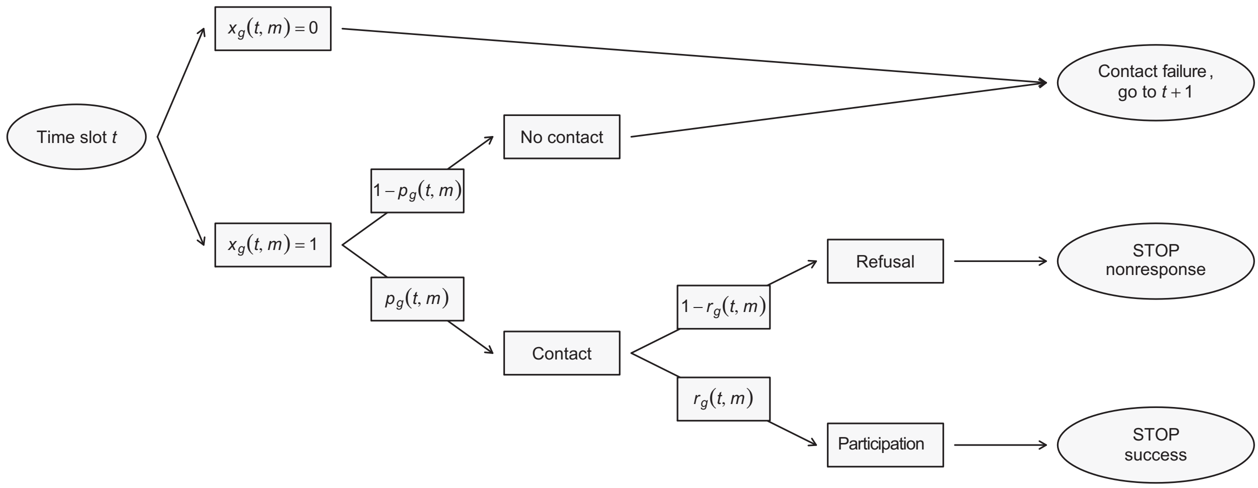

Fig. 1. Sequence of events for a given attempt.

$$
\begin{aligned}
\max \; & \sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}}\sum_{m\in\mathcal{M}} w_g f_g(t-1) x_g(t,m) p_g(t,m) r_g(t,m) \\
\text{s.t.} \; & \sum_{t\in\mathcal{T}} x_g(t,m) \leqslant \bar{k}_g(m), \quad \forall g\in\mathcal{G},\ \forall m\in\mathcal{M}, \\
& \sum_{m\in M} x_g(t,m) \leqslant 1, \quad \forall g\in\mathcal{G}, \quad \forall t\in\mathcal{T}, \\
& f_g(t) = \prod_{m\in\mathcal{M}} [x_g(t,m)(1-p_g(t,m)) + 1 - x_g(t,m)] f_g(t-1), \\
& \quad \forall g\in\mathcal{G}, t\in\mathcal{T} \\
& f_g(0) = 1, \quad \forall g\in\mathcal{G}, \\
& x_g(t,m)\in\{0,1\}, \quad \forall g\in\mathcal{G}, \quad \forall t\in\mathcal{T}, \quad \forall m\in\mathcal{M}.
\end{aligned} \tag{3}
$$

Problem (3) represents the adaptive survey design problem in which survey features with significant influence on the quality of the survey are balanced. In our model, the features are the *interview mode*, the *number of allowed attempts*, and *number of time slots*. The solution of the problem is, however, not trivial. The objective function is a nonconvex nonlinear function, and the constraints do not form a convex polytope either. As a consequence, our problem is non-tractable from a mathematical programming perspective, even for small-sized problems (e.g., one group and four time slots). In the next section, we develop an algorithm that is able to derive optimal solutions by aggregating information in the adaptive survey design problem.

3. Adaptive survey design policies

In this section, we reformulate the adaptive survey design problem such that the problem becomes numerically tractable. In order to do this, note that at any time t, it is sufficient to know the probability of contact failure up to time t, $f_g(t-1)$, instead of the complete configuration $x_g(t',m)$ for $t'<t$ for all g. Denote by $\vec{f}=(f_1,\ldots,f_G)$ the vector storing the probability of contact failure up to time t. Hence, given $\vec{f}(T-1)$, the decision at time T is obvious when one also keeps track of the number of times that mode m has been used for each group g. Since the decision at time T is completely determined, one can then calculate the optimal decisions at time $T-1$, and continue working back towards the first time epoch (see Fig. 2). By keeping track of the time, the contact failure probability, and the utilization of the different modes, the problem becomes completely Markovian and the problem can be cast as a Markov decision problem.

Let the state space of the Markov decision problem be denoted by $\mathcal{S}=\mathcal{T}\times[0,1]^G\times\{0,1,\ldots\}^{G\cdot M}$, where $s=(t,\vec{f},K)\in\mathcal{S}$ has components t, denoting the time at which the process resides, $\vec{f}$ the probability of contact failure up to time t, and $K=(k_g(m))_{g\in\mathcal{G},m\in\mathcal{M}}$ denoting that mode m can still be used $k_g(m)$ times for group g. The action space $\mathcal{A}_s$ is given by

$$
\mathcal{A}_s = \{(a_g(m))_{g\in\mathcal{G},m\in\mathcal{M}} | a_g(m)\in\{0,1\}, a_g(m)\leqslant k_g(m)\}, \sum_{m\in\mathcal{M}} a_g(m) \leqslant 1\},
$$

where $a_g(m)$ denotes the available action for group g using mode m. More specifically, given the state space s the process is in, choosing an action translates to choosing whether to approach ($a_g(m)=1$) or not ($a_g(m)=0$) provided that there are attempts left. If the number of attempts has been exhausted, the only allowed action is not approach. The transition probability p is given by $p(s,a,s')=1$ if $t'=t+1$, $f'_g=\prod_{m\in\mathcal{M}}[1-a_g(m)p_g(t,m)]f_g$, and $k'_g(m)=k_g(m)-a_g(m)$, and zero for all other s'. The rewards r are given by

$$
r(s,a) = \sum_{g\in\mathcal{G}}\sum_{m\in\mathcal{M}} w_g a_g(m) f_g p_g(t,m) r_g(t,m).
$$

The tuple $(\mathcal{S},\mathcal{A},p,r)$ completely defines the Markov decision problem (see also [13]).

The Markov decision problem can be solved by dynamic programming (or backward recursion) by formulating the recursion equations for state $s=(t,\vec{f},K)$ given by

$$
\begin{aligned}
V(s) = & \max_{a\in\mathcal{A}_s}\left[r(s,a)+\sum_{s'\in\mathcal{S}} p(s,a,s')V(s')\right] = \max_{a\in\mathcal{A}_s}\left[\sum_{g\in\mathcal{G}}\sum_{m\in\mathcal{M}} w_g a_g(m) f_g p_g(t,m) r_g(t,m)\right. \\
& \left. + V\left(t+1, \left(\prod_{m\in\mathcal{M}}[1-a_g(m)p_g(t,m)]f_g\right)_{g\in\mathcal{G}}, (k_g(m)-a_g(m))_{g\in\mathcal{G},m\in\mathcal{M}}\right)\right],
\end{aligned} \tag{4}
$$

and by setting $V(s)=0$ for all $s=(T+1,\vec{f},K)$. Note that the algorithm only needs T iterations, and in each iteration only $2^{G\cdot M}$ actions need to be considered. Hence, for values of realistic size, the algorithm is computationally feasible and is guaranteed to converge to the optimal solution. The weighted response rate is then given by $V(s)$ with $t=1$, $f_g=1$, and $K=(\bar{k}_g(m))_{g\in\mathcal{G},m\in\mathcal{M}}$.

4. Features of the model

In the previous section, we formulated the adaptive survey design problem in which the focus was on the quality of the survey

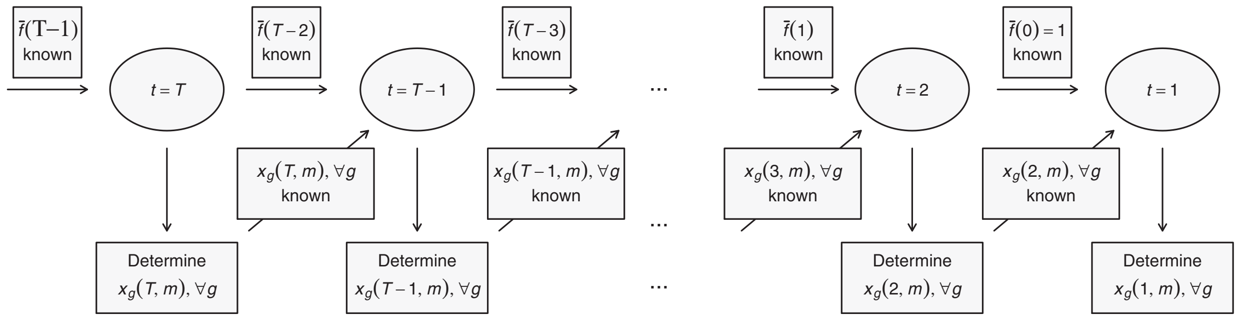

Fig. 2. Sequence of decisions.

results modeled by maximizing the weighted response rates. However, the model formulation is sufficiently flexible to include other features as well, such as budgetary constraints or capacity restrictions. In this section, we discuss how these features can be integrated within our existing framework.

First, we consider a constraint on the budget. Every time a sample unit is approached for a survey, costs are incurred for the effort. These costs mainly depend on the interview mode and also on the outcome of each approach. Denote by $b^s(m)$ the costs that are incurred by using mode m with a successful outcome. For the costs that are incurred by mode m that results in a failure, we distinguish two types of costs: $b^{f_c}(m)$ when the failure occurs due to failure of contact, and $b^{f_r}(m)$ when the failure occurs due to failure to participate. Let B be the total budget that is available for the survey. An approach at time t using mode m bears the following costs

$$p_g(t,m)[r_g(t,m)b^s(m) + (1-r_g(t,m))b^{f_r}(m)] + (1-p_g(t,m))b^{f_c}(m).$$

In general, the costs $b_g(t,m)$ at time t using mode m depend on the contact failures before time t. These costs can be written as follows

$$b_g(t,m) = x_g(t,m)f_g(t-1)[p_g(t,m)[r_g(t,m)b^s(m) + (1 - r_g(t,m))b^{f_r}(m)] + (1-p_g(t,m))b^{f_c}(m)],$$

with $f_g(t)$ given by (1). Hence, using this definition, the budgetary constraint that needs to be added to problem (3) is given by

$$\sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}}\sum_{m\in\mathcal{M}} N_g b_g(t,m) \leqslant B.$$

A capacity constraint can be addressed in a manner analogous to the constraint on the budget. Let C be the available capacity, measured by the number of interviewer hours available to survey the sample. Similar to the cost structure, the required capacity depends on the interview mode and the outcome of each approach. Denote by $c^s(m)$, $c^{f_c}(m)$, and $c^{f_r}(m)$ the capacity utilized when the approach is successful, or has failed due to contact failure, or failed due to participation failure, respectively. Following the same steps as above, the capacity constraint to be added to problem (3) is given by

$$\sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}}\sum_{m\in\mathcal{M}} N_g c_g(t,m) \leqslant C,$$

with $c_g(t,m)$ defined as

$$c_g(t,m) = x_g(t,m)f_g(t-1)[p_g(t,m)[r_g(t,m)c^s(m) + (1 - r_g(t,m))c^{f_r}(m)] + (1-p_g(t,m))c^{f_c}(m)].$$

Note that if the budgetary constraint and the capacity limitation are added to the model, then the maximum number of attempts $\bar{k}_g(m)$ becomes obsolete. Hence, the binary programming problem now becomes

$$\begin{aligned}
\max\ & \sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}}\sum_{m\in\mathcal{M}} w_g f_g(t-1)x_g(t,m)p_g(t,m)r_g(t,m)\\
\text{s.t.}\ & \sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}}\sum_{m\in\mathcal{M}} N_g b_g(t,m) \leqslant B,\\
& \sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}}\sum_{m\in\mathcal{M}} N_g c_g(t,m) \leqslant C,\\
& \sum_{m\in M} x_g(t,m) \leqslant 1, \quad \forall g\in\mathcal{G}, \quad \forall t\in\mathcal{T},\\
& f_g(t) = \prod_{m\in\mathcal{M}}[x_g(t,m)(1-p_g(t,m)) + 1 - x_g(t,m)]f_g(t-1),\\
& \quad \forall g\in\mathcal{G}, t\in\mathcal{T},\\
& f_g(0) = 1, \quad \forall g\in\mathcal{G},\\
& b_g(t,m) = x_g(t,m)f_g(t-1)[p_g(t,m)[r_g(t,m)b^s(m) + (1-r_g(t,m))b^{f_r}(m)]\\
& \quad + (1-p_g(t,m))b^{f_c}(m)], \quad \forall g\in\mathcal{G}, \quad \forall t\in\mathcal{T},\ \forall m\in\mathcal{M},\\
& c_g(t,m) = x_g(t,m)f_g(t-1)[p_g(t,m)[r_g(t,m)c^s(m) + (1-r_g(t,m))c^{f_r}(m)]\\
& \quad + (1-p_g(t,m))c^{f_c}(m)], \quad \forall g\in\mathcal{G},\ \forall t\in\mathcal{T},\ \forall m\in\mathcal{M},\\
& x_g(t,m)\in\{0,1\}, \quad \forall g\in\mathcal{G}, \quad \forall t\in\mathcal{T}, \quad \forall m\in\mathcal{M}.
\end{aligned} \tag{5}$$

Note that in this formulation, we have chosen to model the budgetary constraint and the capacity restriction as a global constraint over all the groups. However, it is quite easy to divide the budget B into budgets B_g for each group g, and then have a constraint per group. A similar remark holds for the capacity restriction as well.

In order to incorporate the budgetary constraint and the capacity restriction in the Markov decision problem, we need to add the state variables b and c for both the budget and the capacity, respectively. In each state $s = (t, \vec{f}, K, b, c)$, these variables denote the budget and the capacity that are left for the rest of the survey. At time t, the budget and the capacity after taking an action $a_g(m)$ are decreased by $\sum_{g\in\mathcal{G}}\sum_{m\in\mathcal{M}} a_g(m)b_g(t,m)$ and $\sum_{g\in\mathcal{G}}\sum_{m\in\mathcal{M}} a_g(m)c_g(t,m)$, respectively. This can only be done as long as the budget and the capacity remain non-negative. This requirement is added to the action set. Hence, the dynamic programming backward recursion equations become

$$\begin{aligned}
V(s) = \max_{a\in\mathcal{A}_s}\Bigg[&\sum_{g\in\mathcal{G}}\sum_{m\in\mathcal{M}} w_g a_g(m) f_g p_g(t,m) r_g(t,m)\\
&+ V\Bigg(t+1, \left(\prod_{m\in\mathcal{M}}[1-a_g(m)p_g(t,m)]f_g\right)_{g\in\mathcal{G}}, (k_g(m)-a_g(m))_{g\in\mathcal{G},m\in\mathcal{M}}, b\\
&- \sum_{g\in\mathcal{G}}\sum_{m\in\mathcal{M}} a_g(m)b'_g(t,m), c - \sum_{g\in\mathcal{G}}\sum_{m\in\mathcal{M}} a_g(m)c'_g(t,m)\Bigg)\Bigg],
\end{aligned} \tag{6}$$

with

$$\mathcal{A}_s = \Big\{ a_g(m) | a_g(m) \in \{0,1\}, a_g(m) \leqslant k_g(m), \sum_{m\in\mathcal{M}} a_g(m) \leqslant 1,$$
$$b - \sum_{g\in\mathcal{G}} \sum_{m\in\mathcal{M}} a_g(m) b'_g(t,m) \geqslant 0, \text{ and } c - \sum_{g\in\mathcal{G}} \sum_{m\in\mathcal{M}} a_g(m) c'_g(t,m) \geqslant 0 \Big\},$$

in which we defined $b'_g(t,m)$ and $c'_g(t,m)$ to be

$$b'_g(t,m) = f_g(t-1)p_g(t,m)\left[r_g(t,m)b^s(m) + (1-r_g(t,m))b^{f_r}(m)\right] + (1-p_g(t,m))b^{f_c}(m)$$

and

$$c'_g(t,m) = f_g(t-1)p_g(t,m)[r_g(t,m)c^s(m) + (1-r_g(t,m))c^{f_r}(m)] + (1 - p_g(t,m))c^{f_c}(m).$$

5. Numerical examples

The previous sections dealt with the theoretical models to solve the problem of resource allocation within adaptive survey designs. In this section, we give two numerical examples to illustrate our methodology.

Our first example shows that the solution of the basic unconstrained model is indeed optimal, although, counterintuitive. Consider a survey sample in which all units belong to the same group g. The set of available interview modes is $\mathcal{M} = \{\text{Face-to-face}, \text{phone}\}$. The survey fieldwork is divided in $T = 6$ time slots. Table 1 gives the contact and participation probabilities $p_g(t,m)$ and $r_g(t,m)$ as estimated from previous such surveys and the maximum number of attempts $\bar{k}_g(m)$.

Note that there is a clear preference for contact at time slots t_3 and t_6 for both interview modes. For participation, on the other hand, the probabilities indicate more than 50% chance for positive participation except for an attempt by face-to-face at t_3 and by phone at t_5. Therefore, it is not obvious what time slots should be chosen in order to maximize the total response. Hence, the optimal solution is hard to derive from intuition. Using the algorithm from Section 3, we obtain the solution depicted in Table 2.

Let us analyze this solution. It looks surprising that for the first time slot mode F2F is chosen and not Ph, although the immediate reward is higher for Ph. However, considering the formula given in (1) for the group average response, we see that the lower the contact probability for the first time slot, the higher the future reward. Also, the participation probability $r_g(t_1,\text{F2F})$ is higher than $r_g(t_1,\text{Ph})$. The situation changes when $r_g(t_1,\text{F2F}) < r_g(t_1,\text{Ph})$. For example, take $r_g(t_1,\text{F2F}) = 0.7$. As expected, the new optimal solution (see Table 3) uses *phone* as first approach interview mode.

The structure of the solution given in Table 2 is motivated by the choice of $\bar{k}_g(m)$. From t_3 onward $k_g(\text{F2F}) = 0$, therefore *phone* is the only interview mode left available. Thus, the choice for time slots t_3, t_4, and t_6 is logical. However, taking action 0 at t_5 again looks counterintuitive. Since there are enough attempts left for mode Ph and there are no budget or capacity constraints, it feels natural to choose for an attempt to approach. The explanation lies in the value of the objective function that is higher in this case (0.753 compared to 0.752 if the unit is approached).

Table 1
Input data for group g.

Mode	Probability	t_1	t_2	t_3	t_4	t_5	t_6	$\bar{k}_g(m)$
Face-to-face	$p_g(t,m)$	0.3	0.4	0.8	0.2	0.3	0.7	2
	$r_g(t,m)$	0.9	0.7	0.3	0.8	0.8	0.6	
Phone	$p_g(t,m)$	0.4	0.5	0.9	0.4	0.4	0.8	4
	$r_g(t,m)$	0.8	0.5	0.7	0.6	0.4	0.6	

Table 2
Optimal solution–original setting.

Time slot	t_1	t_2	t_3	t_4	t_5	t_6	Response rate
Mode	F2F	F2F	Ph	Ph	0	Ph	0.753

Table 3
Optimal solution–different participation probability at t_1.

Time slot	t_1	t_2	t_3	t_4	t_5	t_6	Response rate
Mode	Ph	F2F	Ph	Ph	F2F	Ph	0.736

Table 4
Optimal solution–more attempts available.

Time slot	t_1	t_2	t_3	t_4	t_5	t_6	Response rate
Mode	F2F	F2F	Ph	Ph	F2F	Ph	0.755

Table 5
Input data for group g_2.

Mode	Probability	t_1	t_2	t_3	t_4	t_5	t_6	$k_{g_2}(m)$
Face-to-face	$p_{g_2}(t,m)$	0.8	0.6	0.4	0.6	0.4	0.2	1
	$r_{g_2}(t,m)$	0.9	0.7	0.6	0.8	0.5	0.3	
Phone	$p_{g_2}(t,m)$	0.7	0.6	0.5	0.6	0.5	0.4	2
	$r_{g_2}(t,m)$	0.8	0.6	0.5	0.6	0.4	0.2	

Table 6
Optimal solution for group g_2.

Time slot	t_1	t_2	t_3	t_4	t_5	t_6	Response rate
Mode	F2F	Ph	0	Ph	0	0	0.821

The optimal solution in Table 2 does not employ all attempts available for mode Ph. Therefore, we cannot obtain a different solution if we increase the number of attempts for this mode. On the other hand, if we increase the number of attempts to 3 for F2F, then the average response improves (see Table 4). The structure of the optimal solution does not change much from the original setting. The only difference appears at t_5 where this time there are enough attempts for mode F2F, and selecting this mode leads to higher response.

Our second example depicts the optimization mechanism for two groups in the presence of budgetary and capacity constraints. Consider again the setting from the previous example, where Table 1 has the input data for group g_1. Table 5 below gives the corresponding input data for group g_2.

Approaching group g_2 for the survey follows a more intuitive behavior, e.g., high participation probabilities correspond to high contact probabilities. In the case of single group optimization, the optimal solution for group g_2 (see Table 6) starts with the choice of face-to-face as interview mode at t_1, since this results in the highest immediate reward. The same argument governs the entire structure of the solution.

Now consider a sample of $N = 2000$ units that can be clustered in two groups given age, i.e., young and old. The proportion of the two groups in the survey sample is $w = (0.62, 0.38)$. A total budget $B = 4000$ monetary units is available to survey the sample units using two modes, i.e., face-to-face and phone. For simplicity

Table 7
Optimal solution for groups young and old.

Time slot		t_1	t_2	t_3	t_4	t_5	t_6	Response rate
Group	Young	F2F	F2F	Ph	F2F	F2F	Ph	0.757
	Old	F2F	F2F	F2F	F2F	F2F	Ph	0.851

we assume that one attempt costs one monetary unit regardless of the employed survey mode. Tables 1 and 5 give the estimates for contact and cooperation probabilities for the two groups, where g_1 denotes the young group and g_2 the old group, respectively. For the sake of simplicity we assume that capacity is unlimited. The overall response rate in this case is 0.793 and the optimal solution for the two groups is given in Table 7.

The costs incurred with this solution amount to 2841 units for g_1 and 1033 units for g_2. The remaining budget could be an indication that the group response rates have attained their maximum, given the input probabilities. An easy approach to confirm such a hypothesis is to optimize for $B > 4000$. The solution does not change which leads to the conclusion that $B = 3873$ units is sufficient to collect maximum response from the two groups. Evidently, dropping the constraint on the number of attempts has created a larger feasible region. This in turn leads to a higher response rate, 0.793 compared to 0.779 obtained if weighting the group response rates from Tables 2 and 6 with the corresponding values in w. The increase of 3.8% in the response rate could be explained by the relatively high budget. Fig. 3 depicts the evolution of the response rate for various levels of budget.

Let us take a look at the changes in the optimal solution (see Table 8) that cause the two steep jumps in the response rate. As expected, group g_2 receives more effort since it yields higher response per attempt than group g_1. This is particularly interesting in the case of $B = 1250$ where the young group is not at all surveyed whereas the old group receives enough monetary units to yield its maximum response rate. From a cost perspective there is no difference between approaching the group at time t_1 or later. The reason for the various situations that group g_1 is not approached at time slot t_1 is the corresponding response probability. For example, for $B = 2250$ at t_3 there is no difference between the two modes in

Table 8
Optimal solution for groups young and old for various values of budget.

B	Group	Time slot						Group response rate	Response rate
		t_1	t_2	t_3	t_4	t_5	t_6		
1000	g_1	0	0	0	0	0	0	0	0.322
	g_2	F2F	F2F	0	F2F	Ph	0	0.849	
1250	g_1	0	0	0	0	0	0	0	0.323
	g_2	F2F	F2F	F2F	F2F	F2F	Ph	0.851	
1750	g_1	0	0	Ph	F2F	F2F	Ph	0.692	0.429
	g_2	0	0	0	0	0	0	0	
2250	g_1	0	0	Ph	0	0	0	0.63	0.701
	g_2	F2F	0	0	F2F	0	0	0.816	
2500	g_1	0	0	Ph	0	F2F	Ph	0.688	0.749
	g_2	F2F	F2F	0	F2F	Ph	0	0.849	
2750	g_1	0	0	Ph	F2F	F2F	Ph	0.692	0.752
	g_2	F2F	F2F	F2F	F2F	F2F	Ph	0.851	
3250	g_1	F2F	0	Ph	0	0	Ph	0.745	0.782
	g_2	F2F	F2F	0	F2F	0	0	0.842	
3500	g_1	F2F	0	Ph	F2F	F2F	Ph	0.754	0.791
	g_2	F2F	F2F	F2F	F2F	F2F	Ph	0.851	
4000	g_1	F2F	F2F	Ph	F2F	F2F	Ph	0.757	0.793
	g_2	F2F	F2F	F2F	F2F	F2F	Ph	0.851	

the cost for an attempt. The yielded response however is higher when using *phone*.

A sensitivity analysis is essential when deciding upon a good value for the necessary budget. Survey designers could decide to increase slightly the available budget if the corresponding increase in the response rate is significant. For example, a budget increase of 11% from 2250 to 2500 leads to an expected 6.9% more response. On the other hand, a similar budget increase from 2500 to 2750 leads to only 0.4% additional expected response.

The algorithm is implemented in C++. Table 9 presents some computational times for the two-group example. All run times are for an Intel Xeon L5520 processor with 4 cores. The run times increase with the increase in the budget. An increase in the budget expands the feasible region with points that yield a response rate at least as high as the previous feasible region. Therefore, additional time is spent on exploring the new points. The significant drop in the runtime for $B = 4000$ can be explained by the fact that, at this point, the sequence of actions that yields maximal group

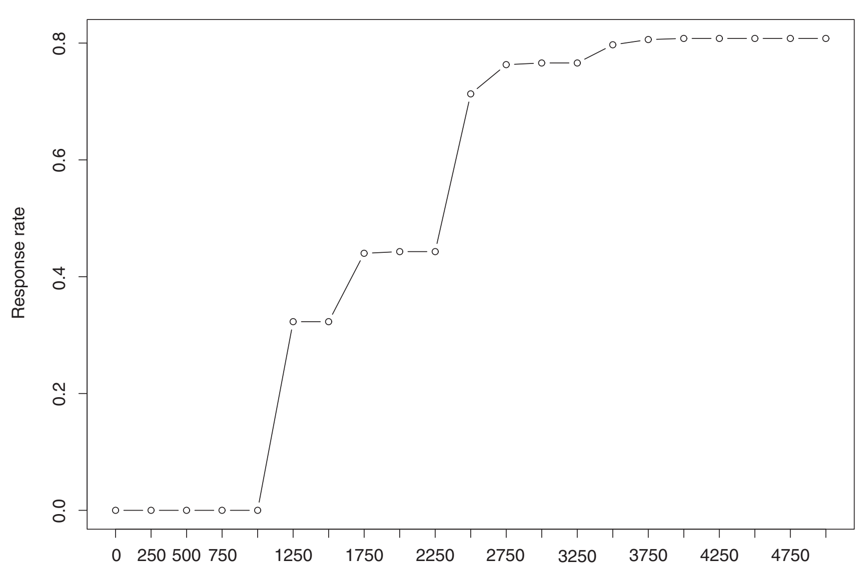

Fig. 3. Response rate evolution for various budget levels.

Table 9
Computational times.

Budget	1000	1250	1750	2250	2500	2750	3250	3500	4000
Runtime (s)	16.2	17	52.7	54.2	55	55.5	55.9	55.8	19

response rates is feasible. The algorithm converges then very quickly to this point.

Other software tools such as Xpress, Maple and R were used in the attempt to solve the resource allocation problem for adaptive survey designs as a mathematical program. However, presence of nonconvexity prohibited convergence to the global optimum. We suppress presentation of computational times for these tools since the optimal solution was only a local optimum.

6. Conclusions

In the current paper we have addressed the important problem of optimal survey designs from a novel perspective of optimal resource allocation. This problem is formulated as a nonconvex integer variable nonlinear resource allocation problem for which currently only approximations are available. We present an algorithm that solves the problem to optimality by exploiting the structure of the survey design problem. The results are relevant to survey organizations which are struggling to obtain high-quality survey results. Decreasing participation to surveys leads to increased efforts to convince sample units to respond. This in turn leads to spending higher budgets. Temporary solutions can be found in replacing expensive designs with cheaper ones. That, however, has a negative influence on the quality of the survey results. Therefore, a new perspective needs to be taken.

Learning the behavior patterns, i.e., the impact various survey features have on the willingness to participate into surveys, for respondents and nonrespondents, aims at obtaining more insight. Adaptive survey designs provide the necessary framework to study these patterns. The main components of an adaptive survey design are interview modes, number of time slots, number of allowed attempts, and the survey sample divided into homogeneous groups according to some given criteria (e.g., demographics). Our research investigates how resource planning can be addressed in the context of adaptive survey designs. Moreover, we optimize the resource allocation while taking into account the quality of the survey results.

We start by analyzing a simpler version of the problem, i.e., with no budget or capacity constraints. In this setting optimizing the resource allocation is translated to choosing a sequence of time slots such that the response rate is maximized given the contact and participation probabilities for each group, each time slot, and all available interview modes. The history of past actions that has to be considered at each step when choosing an action is a complex non-linear factor. Section 2 explains why the simplified model is non-scalable and non-tractable even for small problems.

Nevertheless, the optimal solution can be found. We present a method that addresses the non-linearity of the problem. The idea is to use a Markovian decision formulation of the problem, in which the state space is extended such that the contact failure probability is included in the state. Thus, there is no need to store the entire configuration of past actions. Via dynamic programming the new formulation is solved to optimality. We have tested the performance of the method by analyzing various survey settings. Section 5 presents the numerical results, where the optimal solution is not entirely intuitive.

The main advantages of our method are guaranteed optimality and short computational times. Thus, the model can be successfully used as a basis for representation of more complex practical settings. For example, Section 4 deals with the necessary changes in the theoretical framework that adjust our method to accommodate cost and capacity constraints.

Statistics bureaus are directly interested in testing adaptive designs as an alternative to classical survey designs. Simulated results prove the efficiency of our current technique. A direct comparison between the two designs is not yet possible due to some practical aspects not fully covered by our method. For example, estimation of the input probabilities is not discussed here. However, a great deal of attention has to be paid to this phase since the optimization part builds upon this input. Issues such as time-dependency, history-dependency, and repeated shift between interview modes have to be taken into account when estimating the input probabilities.

Flexibility in addressing various objective functions is another aspect of interest. As shown by (4), the method is applicable as long as the objective function has an additive property. However, recent literature on survey methodology argues that aiming for high response rates influences negatively the bias of the estimators (see, e.g., [15,11]). Other quality measures, such as low variation in the group response rates (i.e., representativeness of the respondent sample), have been indicated as more suitable. Such a quality function, however, does not posses the additive property which makes it difficult to approach by the current method.

Future research aims at tackling these issues in order to develop a model that meets practical needs. Intuitively, taking two survey designs with similar settings, an adaptive design is expected to outperform the classical design since more information becomes available from historical data and the design is tailored such that the group response rate is optimized.

References

[1] A.J. Bangla, D.A. Castanon, Auction algorithm for nonlinear resource allocation problems, in: Proceedings of the 49th IEEE Conference on Decision and Control-CDC, 2010, pp. 3920–3925.

[2] H.P. Benson, S.S. Erenguc, R. Horst, A note on adapting methods for continuous global optimization to the discrete case, Annals of Operations Research 25 (1990) 243–252.

[3] K.M. Bretthauer, A. Ross, B. Shetty, Nonlinear integer programming for optimal allocation in stratified sampling, European Journal of Operational Research 116 (3) (1999) 667–680.

[4] K.M. Bretthauer, B. Shetty, The nonlinear resource allocation problem, Operations research 43 (4) (1995) 670–683.

[5] K.M. Bretthauer, B. Shetty, S. Syam, A specially structured nonlinear integer resource allocation problem, Naval Research Logistics 50 (7) (2003) 770–792.

[6] R.M. Groves, D.A. Dillman, J.L. Eltinge, R.J.A. Little, Survey Nonresponse, Wiley Series in Probability and Statistics, 2002.

[7] S.G. Heeringa, R.M. Groves, Responsive design for household surveys: tools for actively controlling survey errors and costs, Journal of the Royal Statistical Society: Series A (Statistics in Society) 169 (3) (2006).

[8] T. Ibaraki, N. Katoh, Resource Allocation Problems: Algorithmic Approaches, MIT Press Series in the Foundations of Computing, 1988.

[9] W.D. Kalsbeek, S.L. Botman, J.T. Massey, P-W. Liu, Cost-efficiency and the number of allowable call attempts in the National Health Interview Survey, Journal of Official Statistics 10 (2) (1994) 133–152.

[10] N. Katoh, T. Ibaraki, H. Mine, A polynomial time algorithm for the resource allocation problem with a convex objective function, Journal of Operations Research Society 30 (5) (1979) 449–455.

[11] C. Mohl, F. Laflamme, Research and responsive design options for survey data collection at statistics Canada, in: Proceedings of the American Statistical Association, Survey Research Methods Section, 2007.

[12] M. Patriksson, A survey on the continuous nonlinear resource allocation problem, European Journal of Operational Research 185 (2008) 1–46.

[13] M.L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, John Wiley & Sons, 1994.

[14] B. Schouten, M. Calinescu, A. Luiten, Optimizing quality of response through adaptive survey designs, Survey Methodology, accepted for publication.

[15] B. Schouten, F. Cobben, J. Bethlehem, Indicators for the representativeness of survey response, Survey Methodology 35 (1) (2009) 101–113.

[16] J.R. Wagner, Adaptive Survey Design to Reduce Nonresponse Bias, PhD Thesis, University of Michigan, 2008.