Math Meth Oper Res (2010) 72:63–74
DOI 10.1007/s00186-010-0305-6

Efficient, optimal stochastic-action selection when limited by an action budget

John Blatz · Donniell E. Fishkind · Carey E. Priebe

Received: 15 November 2009 / Accepted: 27 March 2010 / Published online: 7 May 2010

Abstract The problem that we consider here is a basic operations research problem, but it also a special case of the Stochastic Shortest Path with Recourse Problem and the Canadian Travellers Problem in the probabilistic path planning literature, and it is also a special case of maximizing a submodular set function subject to a matroid constraint. Specifically, suppose an agent has a task and suppose that there is a set of actions, any of which the agent might perform, with respective probabilities of the actions successfully accomplishing the task and respective rewards for the agent if the actions are successful; the agent is to select a sequence of some of these actions that will be performed sequentially, until the task is accomplished or the selected actions are exhausted, but there is a budget on the number of actions that can be performed. We provide an efficient algorithm that chooses a sequence of actions that, under the budget, maximize the agent's expected reward. An example illustrates how, when conditioning on partial selection of actions, there can be changes to the order of the remaining actions' adjusted utilities. However, we prove and exploit a nesting result involving solutions.

Keywords Submodular function · Matroid constraint · Canadian travellers problem

1 The problem

Let $\mathcal{A}$ be a set of *actions*, let $p : \mathcal{A} \to (0, 1)$ be a function assigning the *probability* p_α to each action $\alpha \in \mathcal{A}$, and let $r : \mathcal{A} \to \mathbb{R}_{>0}$ be an injective function assigning the

J. Blatz · D. E. Fishkind (✉) · C. E. Priebe
Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218, USA
e-mail: fishkind@ams.jhu.edu

reward r_α to each action $\alpha \in \mathcal{A}$. Define the objective function f in the following way: To each $A \subseteq \mathcal{A}$ and injective *precedence function* $\Gamma : \mathcal{A} \to \mathbb{R}$, f assigns the value

$$f(A, \Gamma) := \sum_{\alpha \in A} \left(\prod_{\beta \in A : \Gamma(\beta) > \Gamma(\alpha)} (1 - p_\beta) \right) p_\alpha r_\alpha. \tag{1}$$

For each integer $k : 0 \le k \le |\mathcal{A}|$, we are interested in efficiently solving the problem $P(\mathcal{A}, r, p, k)$:

$$\begin{array}{lll} P(\mathcal{A}, r, p, k): & \text{maximize} & f(A, \Gamma) \\ & \text{such that} & A \subseteq \mathcal{A}, |A| = k \\ & & \Gamma : \mathcal{A} \to \mathbb{R} \text{ is injective.} \end{array}$$

This problem may be interpreted as follows: An agent has a particular task, and each $\alpha \in \mathcal{A}$ corresponds to an actual action that the agent may perform at most once in an attempt to accomplish the task. If the agent performs α then there is an (independent) probability p_α that α successfully accomplishes the task, and if the task is accomplished by action α then there is a reward of r_α for the agent. If the agent will perform the subset of actions $A \subseteq \mathcal{A}$ sequentially and in order of decreasing values of Γ until an action is successful (at which point the mission is terminated) or until no more actions are available (in which case the reward is 0 and the mission is terminated) then the objective function from Eq. 1 is the expected reward for the agent. Thus, if the agent is required to do the task in this manner (sequentially performing actions) except that the agent is restricted to performing no more than k actions[1], then the problem $P(\mathcal{A}, r, p, k)$ is precisely to find which actions A (and in what order Γ) will maximize the agent's expected reward.

Our main result is Theorem 11, which establishes the correctness of the BEST k-SUBSET($\mathcal{A}, r, p$) algorithm. This algorithm solves $P(\mathcal{A}, r, p, k)$ for each of $k = 0, 1, 2, \ldots, |\mathcal{A}|$, collectively in $\mathcal{O}(|\mathcal{A}|^2)$ operations. Using Proposition 10, which establishes the existence of nested solutions for $P(\mathcal{A}, r, p, k)$ as $k = 0, 1, 2, \ldots, |\mathcal{A}|$, BEST k-SUBSET($\mathcal{A}, r, p$) is a greedy algorithm where, iteratively for $k = 0, 1, 2, \ldots |\mathcal{A}| - 1$, the algorithm appends to a solution of $P(\mathcal{A}, r, p, k)$ an action of maximum utility, yielding a solution to $P(\mathcal{A}, r, p, k+1)$. However, it is not a completely greedy algorithm, in the sense that it would not be correct to a priori (when $k = 0$) sort the actions by utility and let that order govern their order of entry into the solution (as k iterates); on the contrary, we show in the second observation on Example 4 that the order of the actions by utility can change as we condition on different sets of actions being included in the agent's selection.

First, in Sect. 1.1, we discuss how our problem of interest is a special case of the Stochastic Shortest Path with Recourse Problem and the Canadian Travellers

[1] "No more than k actions" is equivalent here to "exactly k actions" since, considering the form of f in Eq. 1, we have that adding more actions to be performed after any given sequence of actions can only raise expected reward. Thus the constraint "$A \subseteq \mathcal{A}, |A| \le k$" can substitute for "$A \subseteq \mathcal{A}, |A| = k$" in the problem formulation, and later in (2).

Problem and, in Sect. 1.2, we discuss how our problem of interest is a special case of maximizing a submodular set function subject to a matroid constraint; the reader may skip these two subsections and proceed to the results in Sect. 2 without loss of continuity.

1.1 Our problem of interest is a special case of The Stochastic Shortest Path with Recourse Problem and the Canadian Travellers Problem

The problem we consider here is a special case of the Stochastic Shortest Path with Recourse Problem (SSPR) of Andreatta and Romeo (1988) and the Canadian Travellers Problem (CTP) of Papadimitriou and Yannakakis (1991). Although there are many variants of these problems [see the surveys in Bar-Noy and Schieber (1991) and Provan (2003)], the SSPR and CTP problem formulation most relevant to us is as follows: An instance consists of a graph, a length function assigning nonnegative numbers to all of the graph's edges, a probability function assigning probabilities to some of the edges (these are called the "probabilistic" edges, and the rest are "deterministic"), and two designated vertices s and t. An agent is to walk from s to t in this graph, and the agent can make use of any deterministic edge, but each probabilistic edge is traversable with probability specified by the probability function (independently of the traversability of the other edges). Furthermore, the agent can only learn the traversability status of an edge when the agent (is engaged in the walk and) is at the edge's endpoints, and what is wanted in SSPR and CTP is to find a policy minimizing the expected length of the agent's walk. This problem is $\#P$-hard (Polychronopoulos and Tsitsiklis 1996). (Indeed, Provan remarks in (2003) that "all known no-reset versions of the problem are NP-hard").

Consider the following special case of SSPR and CTP. Suppose there are only two vertices, say s and t, and suppose that the edges $e_0, e_1, e_2, \ldots, e_n$ each have s as one endpoint and t as the other endpoint. For $i = 1, 2, \ldots, n$, say the length of e_i is l_i, and the probability that e_i is traversable is p_i. There is a deterministic edge (or else the expected s, t traversal length would be infinite), so lets say e_0 is deterministic, and (without loss of generality) its length l_0 is greater than the lengths of the other edges (which are probabilistic). Further suppose that disambiguation of an edge is done as it is traversed (so that disambiguation commits the agent to the use of the edge if it is traversable), and that there is a limit of k disambiguations allowed (at which point the agent must traverse e_0 to arrive at t). This problem is now precisely $P(\mathcal{A}, r, p, k)$ where $\mathcal{A}$ consists of $e_1, e_2, \ldots, e_n$, the respective probabilities p are $p_1, p_2, \ldots, p_n$, and the respective rewards r are $(l_0 - l_1)$, $(l_0 - l_2)$, ..., $(l_0 - l_n)$; this is by subtracting l_0 from all of the edges' lengths (which doesn't affect the optimization problem, but makes the length of the deterministic edge 0) and then multiplying by -1 to make the lengths positive and change the problem from a minimization problem into the maximization problem $P(\mathcal{A}, r, p, k)$. In this manner, we can view $P(\mathcal{A}, r, p, k)$ as a special case of SSPR and CTP.

1.2 Our problem of interest may be seen as a special case of maximizing a submodular set function subject to matroid constraint

In Proposition 2 we will show that, for every $A \subseteq \mathcal{A}$ and $\Gamma : \mathcal{A} \to \mathbb{R}$ injective, it holds that $f(A, r) \geq f(A, \Gamma)$, so we then redefine $f(A) := f(A, r)$ and remove Γ as a variable from $P(\mathcal{A}, r, p, k)$. We will prove in Proposition 13 that f is submodular (as a function on the power set of $\mathcal{A}$). Thus $P(\mathcal{A}, r, p, k)$ is a maximization problem of the submodular function f subject to the k-uniform-matroid constraint $\{A \in \mathcal{A} : |A| \leq k\}$ (note that f is an increasing function, meaning that $A \subseteq B \Rightarrow f(A) \leq f(B)$).

There exists strongly polynomial time algorithms for minimizing arbitrary submodular functions over a whole power set (Iwata et al. 2001; Schrijver 2000), however maximizing arbitrary submodular functions over a whole power set is NP-hard (since the maximum cut problem is NP-hard). Indeed, maximizing submodular functions subject to a k-uniform-matroid constraint is NP-hard, but there exists a $\left(1 - \frac{1}{e}\right)$-approximation algorithm for maximizing increasing submodular functions subject to a k-uniform-matroid constraint (Nemhauser et al. 1978) or even any matroid constraint (Vondrak 2008). While it is difficult in general to exactly maximize an arbitrary submodular function subject to a k-uniform-matroid constraint, our algorithm is efficient due to Proposition 10, which establishes that there exists $A^{(0)} \subset A^{(1)} \subset A^{(2)} \subset \cdots \subset A^{(|\mathcal{A}|)} = \mathcal{A}$ such that, for all k, $A^{(k)}$ is a solution of $P(\mathcal{A}, r, p, k)$. This is a special feature of our particular submodular function f and, indeed, it is easy to come up with examples of submodular functions which, when maximized subject to a k-uniform-matroid constraint, the solutions for different values of k would not be nested.[2]

2 The results

Our main result is Theorem 11, which will be established after we prove a number of structural results and technical results.

Proposition 1 *Suppose that (A', Γ') is a solution to the problem $P(\mathcal{A}, r, p, k')$ for a particular integer $k' : 0 < k' \leq |\mathcal{A}|$. Then for each integer $k'' : 0 < k'' \leq |\mathcal{A}|$ there exists a solution (A'', Γ'') to the problem $P(\mathcal{A}, r, p, k'')$ such that $A' \cap A'' \neq \emptyset$.*

Proof First, we establish that $A' \cap \arg\max_{\beta \in \mathcal{A}} p_\beta r_\beta \neq \emptyset$; by way of contradiction, suppose not. Select any $\alpha \in \arg\max_{\beta \in \mathcal{A}} p_\beta r_\beta$, let γ be $\arg\min_{\beta \in A'} \Gamma'(\beta)$, define $\hat{A} := (A' \backslash \{\gamma\}) \cup \{\alpha\}$, and define $\hat{\Gamma}$ to be Γ' except that $\hat{\Gamma}(\alpha) := \min_{\beta \in \mathcal{A}} \Gamma'(\beta) - 1$. Note that $f(\hat{A}, \hat{\Gamma})$ and $f(A', \Gamma')$ differ for only one summand in Eq. 1, hence $f(\hat{A}, \hat{\Gamma}) - f(A', \Gamma') = \left(\prod_{\beta \in A : \beta \neq \gamma} (1 - p_\beta)\right)(p_\alpha r_\alpha - p_\gamma r_\gamma) > 0$, thus $f(\hat{A}, \hat{\gamma}) > f(A', \Gamma')$ contradicts the optimality of (A', Γ'). Hence there exists $\alpha' \in A' \cap \arg\max_{\beta \in \mathcal{A}} p_\beta r_\beta$.

[2] For example, let $B^{(0)}, B^{(1)}, B^{(2)}, \ldots, B^{(|\mathcal{A}|)} \subseteq \mathcal{A}$ be **any** sets such that $|B^{(k)}| = k$ for all k. Consider any strictly increasing, strictly concave real function $h : [0, |\mathcal{A}|] \to \mathbb{R}$, and define the set function $g : 2^{\mathcal{A}} \to \mathbb{R}$ by, for all $A \subseteq \mathcal{A}$, $g(A) := h(|A|) + \epsilon \cdot \mathbf{1}_{A = B^{(|A|)}}$, where $\mathbf{1}$ is the indicator function and ϵ is "small enough." Then, for all k, we have that $B^{(k)}$ maximizes g over $\{A \in \mathcal{A} : |A| \leq k\}$, without having stipulated nesting among $B^{(0)}, B^{(1)}, B^{(2)}, \ldots, B^{(|\mathcal{A}|)}$.

Next, let $(\tilde{A}, \tilde{\Gamma})$ be any solution to the problem $P(\mathcal{A}, r, p, k'')$; if $\alpha' \in \tilde{A}$ then we are done. Otherwise, say that δ is $\arg\min_{\beta \in \tilde{A}} \tilde{\Gamma}(\beta)$, define $A'' := (\tilde{A} \backslash \{\delta\}) \cup \{\alpha'\}$, and define Γ'' to be $\tilde{\Gamma}$ except that $\Gamma''(\alpha') := \min_{\beta \in \mathcal{A}} \tilde{\Gamma}(\beta) - 1$. Now, $f(A'', \Gamma'')$ and $f(\tilde{A}, \tilde{\Gamma})$ differ for only one summand in Eq. 1, hence $f(A'', \Gamma'') - f(\tilde{A}, \tilde{\Gamma}) = \left(\prod_{\beta \in \tilde{A}: \beta \neq \delta} (1 - p_\beta)\right) (p_{\alpha'} r_{\alpha'} - p_\delta r_\delta) \geq 0$, and thus $f(A'', \Gamma'') \geq f(\tilde{A}, \tilde{\Gamma})$, which means that (A'', Γ'') is also a solution to $P(\mathcal{A}, r, p, k'')$, and note that $\alpha' \in A' \cap A''$. □

Proposition 2 *For every $A \subseteq \mathcal{A}$ and $\Gamma : \mathcal{A} \to \mathbb{R}$ injective, it holds that $f(A, r) \geq f(A, \Gamma)$.*

Proof For any $A \subseteq \mathcal{A}$ and $\Gamma : \mathcal{A} \to \mathbb{R}$ injective, suppose that $\alpha, \beta \in A$ satisfy $\Gamma(\alpha) > \Gamma(\beta)$ such that there is no $\gamma \in A$ satisfying $\Gamma(\alpha) > \Gamma(\gamma) > \Gamma(\beta)$. Let $\Gamma' : \mathcal{A} \to \mathbb{R}$ be defined to agree with Γ, except that $\Gamma'(\alpha) := \Gamma(\beta)$ and $\Gamma'(\beta) := \Gamma(\alpha)$. Note that $f(A, \Gamma')$ and $f(A, \Gamma)$ differ for only two summands in Eq. 1 and, simplifying the difference, $f(A, \Gamma') - f(A, \Gamma) = \left(\prod_{\delta \in A: \Gamma(\delta) > \Gamma(\alpha)} (1 - p_\delta)\right) p_\alpha p_\beta (r_\beta - r_\alpha)$ thus, if $r_\alpha < r_\beta$, then $f(A, \Gamma') \geq f(A, \Gamma)$. By this observation, a bubble sort of A (initially ordered by decreasing values of Γ) on the key r (to achieve the order of decreasing values of r) generates monotonically nondecreasing values of f as the sort iterates, hence $f(A, r) \geq f(A, \Gamma)$. □

Example 3 Consider the set of actions $\mathcal{A} = \{\alpha, \beta\}$ with $r_\alpha = 20$, $p_\alpha = .1$, $r_\beta = 10$, $p_\beta = .5$; if $k = 1$ then the optimal set of actions A is $\{\beta\}$ with expected reward 5 (as opposed to $\{\alpha\}$ with expected reward 2), but if $k = 2$ then the optimal A is $\{\alpha, \beta\}$ and α is attempted before β (expected reward 6.5), as opposed to β attempted before α (expected reward 6). In particular, note that α has a higher reward than β and comes first when the two actions are allowed, yet α is omitted in favor of β if only one action is allowed.

For each $A \subseteq \mathcal{A}$, define $f(A) := f(A, r) = \sum_{\alpha \in A} \left(\prod_{\beta \in A: r_\beta > r_\alpha} (1 - p_\beta)\right) p_\alpha r_\alpha$. By Proposition 2 we can equivalently formulate $P(\mathcal{A}, r, p, k)$ as

$$P(\mathcal{A}, r, p, k): \text{ maximize } f(A) \quad \text{such that } A \subseteq \mathcal{A}, |A| = k. \tag{2}$$

For any $A \subseteq \mathcal{A}$ and $\alpha \in \overline{A}$ (where $\overline{A}$ denotes the complement of A), define the *adjusted reward of α given A* to be $r_\alpha^A := \frac{f(A \cup \{\alpha\}) - f(A)}{p_\alpha}$; notice that $r_\alpha^\emptyset = r_\alpha$ for all $\alpha \in \mathcal{A}$. For any $A \subseteq \mathcal{A}$ and $\alpha \in \overline{A}$, define the *utility of α given A* to be the quantity $f(A \cup \{\alpha\}) - f(A)$ or, equivalently, $p_\alpha r_\alpha^A$. Consider the following example:

Example 4 Let $\mathcal{A}$ consist of the actions α, β, γ with respective rewards 3, 2, 1 and respective probabilities .25, .50, and .90. Then the values of f for all subsets of $\mathcal{A}$ are in the following table:

subset $\{\cdot\}$	$\emptyset$	$\{\alpha\}$	$\{\beta\}$	$\{\gamma\}$	$\{\beta, \gamma\}$	$\{\alpha, \gamma\}$	$\{\alpha, \beta\}$	$\{\alpha, \beta, \gamma\}$
$f(\{\cdot\})$	0	.75	1	.90	1.45	1.425	1.5	1.8375

and the values of all adjusted rewards and utilities are in the following table:

$\mathcal{A}$	$p_\cdot$	$r_\cdot$	$p_\cdot r_\cdot$	$r_\cdot^{\{\alpha\}}$	$p_\cdot r_\cdot^{\{\alpha\}}$	$r_\cdot^{\{\beta\}}$	$p_\cdot r_\cdot^{\{\beta\}}$	$r_\cdot^{\{\gamma\}}$
α	.25	3	.75			2	.50	2.1
β	.50	2	1	1.5	.750			1.1
γ	.90	1	.90	.75	.675	.5	.45	
$\mathcal{A}$	$p_\cdot r_\cdot^{\{\gamma\}}$	$r_\cdot^{\{\beta,\gamma\}}$	$p_\cdot r_\cdot^{\{\beta,\gamma\}}$	$r_\cdot^{\{\alpha,\gamma\}}$	$p_\cdot r_\cdot^{\{\alpha,\gamma\}}$	$r_\cdot^{\{\alpha,\beta\}}$	$p_\cdot r_\cdot^{\{\alpha,\beta\}}$	
α	.525	1.55	.3875					
β	.550			.825	.4125			
γ						.375	.3375	

For $k = 0, 1, 2, 3$*, solutions for* $P(\mathcal{A}, r, p, k)$ *are respectively* $\emptyset, \{\beta\}, \{\alpha, \beta\}, \{\alpha, \beta, \gamma\}$.

There are several important observations to make in Example 4:

(1) Here α had the highest reward, yet α wasn't in the solution when $k = 1$. Although we showed in Proposition 2 that once actions are selected they are best attempted in order of their rewards, yet in Theorem 11 and its accompanying algorithm we will see that it is utility which governs the selection of **which** actions to attempt—and from this example we see that the order of actions by utility is not in general the same as the order of actions by reward or adjusted reward.

(2) Here the utility of α given $\emptyset$ was less than the utility of γ given $\emptyset$ (indeed, $.75 < .90$), yet the utility of α given $\{\beta\}$ was greater than the utility of γ given $\{\beta\}$ (indeed, $.50 > .45$). Thus, in general, the ordering of actions by utility-given-A is not preserved when A is changed, even if the new A is a superset or a subset of the original A.

(3) In contrast to Point 2, here the ordering of the actions by adjusted-reward-given-A was always the same regardless of A; α's adjusted reward was always greater than β's and γ's, and β's adjusted reward was always greater than γ's. We will later prove in Proposition 7 that this is always the case; i.e. in general it holds that the ordering of actions by adjusted-reward-given-A is always the same as the ordering of the actions by reward.

(4) Here the solutions for $k = 0, 1, 2, 3$ were nested. In fact, it will be shown in Proposition 10 that in general such nested solutions exist. Indeed, this will be the key to the efficiency of the algorithm accompanying Theorem 11.

(5) Here the function f is strictly submodular on the powers set of $\mathcal{A}$, i.e. for all $A, B \subseteq \mathcal{A}$ it holds that $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$, with strict inequality if neither $A \subseteq B$ nor $B \subseteq A$. In Proposition 13 we will prove that in general f is strictly submodular.

We now develop a couple of useful identities for adjusted reward (and hence for utility):

Proposition 5 *For any* $A \subseteq \mathcal{A}$ *and* $\alpha \in \overline{A}$,

$$r_\alpha^A = \left(\prod_{\beta \in A : r_\beta > r_\alpha} (1 - p_\beta) \right) \left(r_\alpha - f(\{\gamma \in A : r_\gamma < r_\alpha\}) \right).$$

Proof Breaking Eq. 1 into two summations, we get

$$
\begin{aligned}
f(A) &= \sum_{\gamma \in A} \left(\prod_{\beta \in A : r_\beta > r_\gamma} (1 - p_\beta) \right) p_\gamma r_\gamma \\
&= \sum_{\gamma \in A : r_\gamma > r_\alpha} \left(\prod_{\beta \in A : r_\beta > r_\gamma} (1 - p_\beta) \right) p_\gamma r_\gamma + \sum_{\gamma \in A : r_\gamma < r_\alpha} \left(\prod_{\beta \in A : r_\beta > r_\gamma} (1 - p_\beta) \right) p_\gamma r_\gamma \\
&= \sum_{\gamma \in A : r_\gamma > r_\alpha} \left(\prod_{\beta \in A : r_\beta > r_\gamma} (1 - p_\beta) \right) p_\gamma r_\gamma \\
&\quad + \left(\prod_{\beta \in A : r_\beta > r_\alpha} (1 - p_\beta) \right) \sum_{\gamma \in A : r_\gamma < r_\alpha} \left(\prod_{\beta \in A : r_\alpha > r_\beta > r_\gamma} (1 - p_\beta) \right) p_\gamma r_\gamma .
\end{aligned}
$$

Similarly, $f(A \cup \{\alpha\})$ is

$$
\begin{aligned}
&\sum_{\gamma \in A : r_\gamma > r_\alpha} \left(\prod_{\beta \in A : r_\beta > r_\gamma} (1 - p_\beta) \right) p_\gamma r_\gamma \\
&+ \left(\prod_{\beta \in A : r_\beta > r_\alpha} (1 - p_\beta) \right) \left[p_\alpha r_\alpha + (1 - p_\alpha) \sum_{\gamma \in A : r_\gamma < r_\alpha} \left(\prod_{\beta \in A : r_\alpha > r_\beta > r_\gamma} (1 - p_\beta) \right) p_\gamma r_\gamma \right] .
\end{aligned}
$$

Taking the difference, we obtain that

$$
\begin{aligned}
f(A \cup \{\alpha\}) - f(A) &= \left(\prod_{\beta \in A : r_\beta > r_\alpha} (1 - p_\beta) \right) \Bigg[p_\alpha r_\alpha \\
&\quad - p_\alpha \sum_{\gamma \in A : r_\gamma < r_\alpha} \left(\prod_{\beta \in A : r_\alpha > r_\beta > r_\gamma} (1 - p_\beta) \right) p_\gamma r_\gamma \Bigg] \\
&= \left(\prod_{\beta \in A : r_\beta > r_\alpha} (1 - p_\beta) \right) p_\alpha \left[r_\alpha - f(\{\gamma \in A : r_\gamma < r_\alpha\}) \right],
\end{aligned}
$$

then the result follows from the definition of r_α^A. □

Proposition 6 *For any $A \subseteq \mathcal{A}$ and $\alpha, \beta \in \overline{A}$ distinct,*

$$
r_\alpha^{A \cup \{\beta\}} = \begin{cases} r_\alpha^A - p_\beta r_\beta^A & \textit{if } r_\alpha > r_\beta \\ r_\alpha^A - p_\beta r_\alpha^A & \textit{if } r_\alpha < r_\beta \end{cases} .
$$

Proof First, suppose that $r_\alpha < r_\beta$. Observe the expression for r_α^A in the statement of Proposition 5; the only distinction between r_α^A and $r_\alpha^{A\cup\{\beta\}}$ would be one additional factor $(1-p_\beta)$ in the product in the displayed equation, so $r_\alpha^{A\cup\{\beta\}} = (1-p_\beta)r_\alpha^A$, as desired. Next, if $r_\alpha > r_\beta$ then $r_\beta^{A\cup\{\alpha\}} = (1-p_\alpha)r_\beta^A$, as just noted. By this fact, and using the definition of marginal reward,

$$\begin{aligned}
p_\alpha r_\alpha^{A\cup\{\beta\}} &= f(A\cup\{\alpha,\beta\}) - f(A\cup\{\beta\}) \\
&= [f(A\cup\{\alpha,\beta\}) - f(A\cup\{\alpha\})] + [f(A\cup\{\alpha\}) - f(A)] \\
&\quad -[f(A\cup\{\beta\}) - f(A)] \\
&= p_\beta r_\beta^{A\cup\{\alpha\}} + p_\alpha r_\alpha^A - p_\beta r_\beta^A \\
&= p_\beta(1-p_\alpha)r_\beta^A + p_\alpha r_\alpha^A - p_\beta r_\beta^A \\
&= p_\alpha(r_\alpha^A - p_\beta r_\beta^A).
\end{aligned}$$

Dividing by p_α then yields $r_\alpha^{A\cup\{\beta\}} = r_\alpha^A - p_\beta r_\beta^A$, as desired. □

Proposition 7 *For each* $A \subseteq \mathcal{A}$ *and* $\alpha, \beta \in \overline{A}$, *it holds that* $r_\beta^A > 0$, *and we have* $r_\alpha > r_\beta \Rightarrow r_\alpha^A > r_\beta^A$.

Proof The result is shown by induction on $|A|$; the statement is trivially true if $A = \emptyset$, so suppose that the statement is true for all A of a particular cardinality, and consider any particular A with cardinality one larger, and let $\gamma \in A$ be selected arbitrarily.

For all $\beta \in \overline{A}$ there are two possibilities:
If $r_\gamma > r_\beta$ then by the induction hypothesis $r_\beta^{A\setminus\{\gamma\}} > 0$ and, consequently, by Proposition 6 we have $r_\beta^A = (1-p_\gamma)r_\beta^{A\setminus\{\gamma\}} > 0$, as desired.
If $r_\beta > r_\gamma$ then by the induction hypothesis $r_\beta^{A\setminus\{\gamma\}} > r_\gamma^{A\setminus\{\gamma\}} > 0$ and, consequently, by Proposition 6 $r_\beta^A = r_\beta^{A\setminus\{\gamma\}} - p_\gamma r_\gamma^{A\setminus\{\gamma\}} > r_\beta^{A\setminus\{\gamma\}} - p_\gamma r_\beta^{A\setminus\{\gamma\}} = (1-r_\gamma)r_\beta^{A\setminus\{\gamma\}} > 0$, as desired. So, in all cases $r_\beta^A > 0$.

Next, for all $\alpha, \beta \in \overline{A}$ such that $r_\alpha > r_\beta$ there are three possibilities:
If $r_\alpha > r_\beta > r_\gamma$ then by the induction hypothesis $r_\alpha^{A\setminus\{\gamma\}} > r_\beta^{A\setminus\{\gamma\}}$ and, consequently, by Proposition 6 we have $r_\alpha^A = r_\alpha^{A\setminus\{\gamma\}} - p_\gamma r_\gamma^{A\setminus\{\gamma\}} > r_\beta^{A\setminus\{\gamma\}} - p_\gamma r_\gamma^{A\setminus\{\gamma\}} = r_\beta^A$, as desired.
If $r_\gamma > r_\alpha > r_\beta$ then by the induction hypothesis $r_\alpha^{A\setminus\{\gamma\}} > r_\beta^{A\setminus\{\gamma\}}$ and, consequently, by Proposition 6 we have $r_\alpha^A = (1-p_\gamma)r_\alpha^{A\setminus\{\gamma\}} > (1-p_\gamma)r_\beta^{A\setminus\{\gamma\}} = r_\beta^A$, as desired.
Finally, if $r_\alpha > r_\gamma > r_\beta$ then by the induction hypothesis $r_\alpha^{A\setminus\{\gamma\}} > r_\gamma^{A\setminus\{\gamma\}} > r_\beta^{A\setminus\{\gamma\}}$, which implies that $r_\alpha^{A\setminus\{\gamma\}} - r_\beta^{A\setminus\{\gamma\}} > p_\gamma(r_\gamma^{A\setminus\{\gamma\}} - r_\beta^{A\setminus\{\gamma\}})$, i.e. $r_\alpha^A = r_\alpha^{A\setminus\{\gamma\}} - p_\gamma r_\gamma^{A\setminus\{\gamma\}} > r_\beta^{A\setminus\{\gamma\}} - p_\gamma r_\beta^{A\setminus\{\gamma\}} = r_\beta^A$ by Proposition 6. So in all cases $r_\alpha^A > r_\beta^A$, and the result follows by induction. □

The next result, Proposition 8, provides a way to describe the utility of a set of actions A given another set of actions B.

Proposition 8 *For any $B \subseteq \mathcal{A}$ and $A \subseteq \overline{B}$ it holds that*

$$f(B \cup A) - f(B) = \sum_{\alpha \in A} \left(\prod_{\beta \in A : r_\beta > r_\alpha} (1 - p_\beta) \right) p_\alpha r_\alpha^B.$$

Proof Suppose that $D \subseteq \overline{B}$ and $\alpha \in \overline{B}$ such that for all $\gamma \in D$, $r_\gamma > r_\alpha$. Then by applying Proposition 6 recursively to $r_\alpha^{\{\cdot\}}$, starting out with $\{\cdot\} = B$, then adding to $\{\cdot\}$ the elements of D one by one, we obtain that

$$r_\alpha^{B \cup D} = \left(\prod_{\beta \in D} (1 - p_\beta) \right) r_\alpha^B. \tag{3}$$

Next, we write $f(B \cup A) - f(B)$ as a telescoping sum

$$\begin{aligned} f(B \cup A) - f(B) &= \sum_{\alpha \in A} \big(f(B \cup \{\beta \in A : r_\beta \geq r_\alpha\}) - f(B \cup \{\beta \in A : r_\beta > r_\alpha\}) \big) \\ &= \sum_{\alpha \in A} p_\alpha r_\alpha^{B \cup \{\beta \in A : r_\beta > r_\alpha\}} \qquad \text{(definition of utility)} \\ &= \sum_{\alpha \in A} \left(\prod_{\beta \in A : r_\beta > r_\alpha} (1 - p_\beta) \right) p_\alpha r_\alpha^B, \qquad \text{(by Eq. 3)} \end{aligned}$$

as desired. □

Now, observe the form of f in Eq. 1, and note the common structure with the expression for $f(B \cup A) - f(B)$ in the statement of Proposition 8; together with Proposition 7, this immediately established the following Proposition 9:

Proposition 9 *Let $B \subseteq \mathcal{A}$ be fixed. The problem $P(\mathcal{A}, r, p, k)$—with an added constraint "$A \supseteq B$" imposed in Displayed Line (2)—is equivalent to the problem $P(\overline{B}, r^B, p, k - |B|)$. Specifically, the objective function of any feasible A in the former problem is exactly $f(B)$ greater than the objective function of $A \backslash B$ in the latter problem; this association (i.e. A to $A \backslash B$) is a one-to-one correspondence between the respective problems' feasible regions and, of course, $f(B)$ is fixed since B is fixed, hence the equivalence of the two problems.*

Proposition 10 *For each integer $k : 0 \leq k < |\mathcal{A}|$ and solution $B \subseteq \mathcal{A}$ for the problem $P(\mathcal{A}, r, p, k)$, there exists a solution $B' \subseteq \mathcal{A}$ for the problem $P(\mathcal{A}, r, p, k + 1)$ such that $B \subseteq B'$.*

Proof Let $B' \subseteq \mathcal{A}$ be a solution for $P(\mathcal{A}, r, p, k + 1)$ such that $|B \cap B'|$ is as large as possible; we claim that $B \subseteq B'$. By way of contradiction, suppose $B \backslash B' \neq \emptyset$. Note that $P(\mathcal{A}, r, p, k)$ and $P(\mathcal{A}, r, p, k + 1)$ are unaffected by imposing an additional constraint in Displayed Line (2) that $A \supseteq B \cap B'$. Thus, by Proposition 9, we have that $B\backslash B'$ is a solution to $P(\overline{B \cap B'}, r^{B \cap B'}, p, k - |B \cap B'|)$. Thus, by Proposition 1, there exists a solution D to $P(\overline{B \cap B'}, r^{B \cap B'}, p, k - |B \cap B'| + 1)$ such that $D \cap (B\backslash B') \neq \emptyset$. By Proposition 9, $(B \cap B') \cup D$ is a solution of $P(\mathcal{A}, r, p, k + 1)$, and $|[(B \cap B') \cup D] \cap B| > |B' \cap B|$ contradicts the maximality of $|B \cap B'|$. □

As stated in Theorem 11, the following algorithm BEST k-SUBSET($\mathcal{A}$, r, p) solves $P(\mathcal{A}, r, p, k)$ for each of $k = 0, 1, 2, \ldots |\mathcal{A}|$.

algorithm BEST k-SUBSET($\mathcal{A}$, r, p)
  $A := \emptyset$
  for $i = 1 : |\mathcal{A}|$, **do**
    select any $\alpha \in \arg\max_{\beta \notin A} p_\beta r_\beta$
    put α in A
    Enterorder[i] := α
    foreach $\gamma \in \overline{A}$, **do**
      if $r_\gamma > r_\alpha$ **then** $r_\gamma := r_\gamma - p_\alpha r_\alpha$
      else $r_\gamma := r_\gamma - p_\alpha r_\gamma$
    endfor
  endfor
end
return Enterorder

Theorem 11 *If* Enterorder *is the output of the algorithm* BEST k-SUBSET($\mathcal{A}$, r, p) *then, for each* $k = 0, 1, 2, \ldots, |\mathcal{A}|$, *a solution to* $P(\mathcal{A}, r, p, k)$ *is* $\{\text{Enterorder}(j) : j \leq k\}$.

Proof Induction on k; when $k = 0$ it is trivial that $\emptyset = \{\text{Enterorder}(j) : j \leq 0\}$ is a solution to $P(\mathcal{A}, r, p, 0)$. Suppose, for any particular k, that $\{\text{Enterorder}(j) : j \leq k\}$ is a solution to $P(\mathcal{A}, r, p, k)$. By Proposition 10, there exists $A \subseteq \mathcal{A}$ such that $|A| = k+1$, and $A \supseteq \{\text{Enterorder}(j) : j \leq k\}$, and A is a solution to $P(\mathcal{A}, r, p, k+1)$. Hence, by the definition of utility, a solution to $P(\mathcal{A}, r, p, k+1)$ may be obtained by appending to $\{\text{Enterorder}\,(j) : j \leq k\}$ the action with greatest utility-given-$\{\text{Enterorder}\,(j) : j \leq k\}$. Indeed, the algorithm does this, and updates (by overwrite) the adjusted rewards in the manner dictated by Proposition 6. □

Several important observations should be made:

(1) The number of operations performed by BEST k-SUBSET($\mathcal{A}$, r, p) is $\mathcal{O}(|\mathcal{A}|^2)$, even though the size of $P(\mathcal{A}, r, p, k)$'s feasible region is exponential in $|\mathcal{A}|$ when k is approximately $\frac{1}{2}|\mathcal{A}|$.
(2) Suppose that for each $k = 0, 1, 2, \ldots, |\mathcal{A}|$ we were to arbitrarily select an arbitrary $A^{(k)}$ from the feasible region of $P(\mathcal{A}, r, p, k)$; the computational expense just to evaluate $f(A^{(k)})$ for all k is already on the order of $|\mathcal{A}|^2$ operations, which is (on the order of) the number of operations for BEST k-SUBSET($\mathcal{A}$, r, p) to find optimal solutions for $P(\mathcal{A}, r, p, k)$ for all k.
(3) The BEST k-SUBSET($\mathcal{A}$, r, p) algorithm can also compute the optimal objective function value of $P(\mathcal{A}, r, p, k)$ for each $k = 0, 1, 2, \ldots, |\mathcal{A}|$ without

increasing the order of the running time. In the algorithm, just add the initialization $fval[0] := 0$ and then, after the "**foreach** $\gamma \in \overline{A}$, **do**" statement, add the statement "$fval[i] := fval[i-1] + p_\alpha r_\alpha$," which computes the optimal objective function value of $P(\mathcal{A}, r, p, i)$.

(4) The BEST k-SUBSET$(\mathcal{A}, r, p)$ is a greedy algorithm, but it is not a completely greedy algorithm, in the sense that it would not be correct to a-priori (when $k = 0$) sort the actions by utility and let that order govern their order of entry into the solution (as k iterates); on the contrary, we show in the second observation on Example 4 that the order of the actions by utility can change as we condition on different sets of actions being included in the agent's selection.

We now shift our attention to proving the submodularity of f, which we mentioned in Sect. 1.2. We begin with a lemma, and show submodularity of f afterwards in Proposition 13.

Proposition 12 *Suppose that* $D \subsetneq D' \subseteq \mathcal{A}$ *and suppose that* $\alpha \in \overline{D'}$. *Then* $r_\alpha^D > r_\alpha^{D'}$.

Proof If $|D'\backslash D| = 1$, say this element in $D'\backslash D$ is β, then by Proposition 6 and by the positivity of adjusted rewards shown in Proposition 7 we have that $r_\alpha^{D'} \leq \max\{r_\alpha^D - p_\beta r_\alpha^D, r_\alpha^D - p_\beta r_\beta^D\} < r_\alpha^D$, as desired. If $|D'\backslash D| > 1$ then adding the elements of $D'\backslash D$ into $\{\cdot\} = D$ one at a time yields strictly decreasing $r_\alpha^{\{\cdot\}}$, which establishes Proposition 12. □

Proposition 13 *Given any* $\mathcal{A}, r, p$, *we have that* f *is strictly submodular. I.e. for all* $A, B \subseteq \mathcal{A}$ *we have* $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$, *with strict inequality when neither* $A \subseteq B$ *nor* $B \subseteq A$.

Proof If $A \subseteq B$ or $B \subseteq A$ then the result is trivial. Otherwise $|A\backslash B| \geq 1$ and $|B\backslash A| \geq 1$; say that the elements of $B\backslash A$ are (arbitrarily) $\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(|B\backslash A|)}$. For each of $i = 1, 2, \ldots, |B\backslash A|$ we have by Proposition 12 that $r_{\beta^{(i)}}^{(A\cap B)\cup\{\beta^{(j)}:j<i\}} > r_{\beta^{(i)}}^{A\cup\{\beta^{(j)}:j<i\}}$ which, by the definition of adjusted reward, precisely says that $\frac{f((A\cap B)\cup\{\beta^{(j)}:j\leq i\}) - f((A\cap B)\cup\{\beta^{(j)}:j<i\})}{p_{\beta^{(i)}}} > \frac{f(A\cup\{\beta^{(j)}:j\leq i\}) - f(A\cup\{\beta^{(j)}:j<i\})}{p_{\beta^{(i)}}}$, i.e. we have that $f\left((A\cap B)\cup\{\beta^{(j)} : j \leq i\}\right) - f\left((A\cap B)\cup\{\beta^{(j)} : j < i\}\right) > f\left(A\cup\{\beta^{(j)} : j \leq i\}\right) - f\left(A\cup\{\beta^{(j)} : j < i\}\right)$. Summing the latter strict inequalities over $i = 1, 2, \ldots, |B\backslash A|$ yields the strict inequality (after the cancellation of telescoping terms) $f((A\cap B)\cup(B\backslash A)) - f(A\cap B) > f(A\cup(B\backslash A)) - f(A)$, which is $f(A) + f(B) > f(A\cup B) + f(A\cap B)$, as desired. □

Acknowledgments We are grateful to Kevin Byrnes for helpful conversations. John Blatz, Carey E. Priebe were supported in part by US Office of Naval Research, grant number N000140610013. Donniell E. Fishkind was supported in part by US Office of Naval Research, grant number N000140610013, and also supported in part by US Department of Defense, grant number H9823007C0365.

References

Andreatta G, Romeo L (1988) Stochastic shortest paths with recourse. Networks 18:193–204

Bar-Noy A, Schieber B (1991) The canadian traveller problem. In: SODA '91: Proceedings of the second annual ACM-SIAM symposium on discrete algorithms, pp 261–270

Iwata S, Fleischer L, Fujishige S (2001) A combinatorial, strongly polynomial-time algorithm for minimizing submodular functions. J ACM 48:761–777

Nemhauser GL, Wolsey LA, Fisher ML (1978) An analysis of approximations for maximizing submodular set functions–I. Math Program 14:265–294

Papadimitriou CH, Yannakakis M (1991) Shortest paths without a map. Theor Comput Sci 84:127–150

Polychronopoulos GH, Tsitsiklis JN (1996) Stochastic shortest path problems with recourse. Networks 27:133–143

Provan JS (2003) A polynomial-time algorithm to find shortest paths with recourse. Networks 42:115–125

Schrijver A (2000) A combinatorial algorithm minimizing submodular functions in strongly polynomial time. J Comb Theory B 80:346–355

Vondrak J (2008) Optimal approximation for the submodular welfare problem in the value oracle model. Symposium on theory of computing