

Supplementary materials for “Combinatorial Semi-Bandits with Knapsacks”

This supplement is structured as follows. Section 6 provides full proof of the main result. Section 7 gives the details of the simulations. Finally, we provide two “appendices” for the sake of making this paper more self-contained: we derive Theorem 2.1 in Appendix A, and list definitions and special cases of matroids in Appendix B.

All references to Sections 1-5 refer to the main paper; references to subsequent sections refer to this supplement. All citations refer to the bibliography in the main paper.

6 Proof of the main result

This section presents a detailed and self-contained proof of the main result: Theorem 3.1. We actually prove a slightly stronger statement involving high-probability regret rather than expected regret:

Theorem 6.1 (main result). *Consider the SemiBwK problem with a linearizable action set \mathcal{F} that admits a negatively correlated RRS. Then algorithm SemiBwK-RRS with this RRS achieves*

$$\text{Regret} \leq O(\log(ndT/\delta)) \sqrt{n} \left(\text{OPT} / \sqrt{B} + \sqrt{T + \text{OPT}} \right) \quad (6.1)$$

with probability at least $1 - \delta$. Here T is the time horizon, n is the number of atoms, B is the budget, and $\delta > 0$ is a given parameter. Parameter α in the confidence radius is set to $\alpha = c_{\text{conf}} \log(ndT/\delta)$, for a large enough absolute constant $c_{\text{conf}} > 0$. Parameter ϵ in the algorithm is set to $\epsilon = \sqrt{\frac{\alpha n}{B}} + \frac{\alpha n}{B} + \frac{\sqrt{\alpha n T}}{B}$. The result holds as long as $B > 3(\alpha n + \sqrt{\alpha n T})$.

6.1 Linear programs

We argue that LP_{ALG} provides a good benchmark that we can use instead of OPT. Fix round t and let $\text{OPT}_{\text{ALG}, t}$ denote the optimal value for LP_{ALG} in this round. Then:

Lemma 6.2. $\text{OPT}_{\text{ALG}, t} \geq \frac{1}{T}(1 - \epsilon) \text{OPT}$ with probability at least $1 - \delta$.

We will prove this by constructing a series of LP’s, starting with a generic linear relaxation for BwK and ending with LP_{ALG} . We show that along the series the optimal value does not decrease with high probability.

The first LP, adapted from Badanidiyuru et al. [2013a], has one decision variable for each action, and applies generically to any BwK problem.

$$\begin{aligned} & \text{maximize} && \sum_{S \in \mathcal{F}} \mu(S) x(S) \\ & \text{subject to} && \sum_{S \in \mathcal{F}} C(S, j) x(S) \leq B/T \quad j = 1, \dots, d \\ & && 0 \leq \sum_{S \in \mathcal{F}} x(S) \leq 1. \end{aligned} \quad (\text{LP}_{\text{BwK}})$$

Let $\text{OPT}_{\text{BwK}}(B)$ denote the optimal value of this LP with a given budget B . Then:

Claim 6.3. $\text{OPT}_{\text{BwK}}(B_\epsilon) \geq (1 - \epsilon) \text{OPT}_{\text{BwK}}(B) \geq \frac{1}{T}(1 - \epsilon) \text{OPT}$.

Proof. The second inequality in Claim 6.3 follows from [Lemma 3.1 in Badanidiyuru et al., 2013a]. We will prove the first inequality as follows. Let \mathbf{x}^* denote an optimal solution to $\text{LP}_{\text{BwK}}(B)$. Consider $(1 - \epsilon)\mathbf{x}^*$; this is feasible to $\text{LP}_{\text{BwK}}(B_\epsilon)$, since for every S , $(1 - \epsilon)x^*(S) \leq 1$ and $\sum_{S \subseteq \mathcal{A}: S \in \mathcal{F}} C(S, j)(1 - \epsilon)x^*(S) \leq B_\epsilon/T$. Hence, this is a feasible solution. Now, consider the objective function. Let \mathbf{y} denote an optimal solution to $\text{LP}_{\text{BwK}}(B_\epsilon)$. We have that

$$\text{OPT}_{\text{BwK}}(B_\epsilon) = \sum_{S \subseteq \mathcal{A}: S \in \mathcal{F}} \mu(S) y^*(S) \geq \sum_{S \subseteq \mathcal{A}: S \in \mathcal{F}} \mu(S) (1 - \epsilon) x^*(S) = (1 - \epsilon) \text{OPT}_{\text{BwK}}(B). \quad \square$$

Now consider a simpler LP where the decision variables correspond to atoms. As before, \mathcal{P} denotes the polytope induced by action set \mathcal{F} .

$$\begin{aligned} & \text{maximize} && \boldsymbol{\mu} \cdot \mathbf{x} \\ & \text{subject to} && C^\dagger \cdot \mathbf{x} \preceq B_\epsilon/T \quad \mathbf{x} \in \mathcal{P} \quad \mathbf{x} \in [0, 1]^n. \end{aligned} \quad (\text{LP}_{\text{ATOMS}})$$

Here $C = (C(a, j) : a \in A, j \in d)$ is the $n \times d$ matrix of expected consumption, and C^\dagger denotes its transpose. The notation \preceq means that the inequality \leq holds for for each coordinate.

Letting $\text{OPT}_{\text{atoms}}$ denote the optimal value for LP_{ATOMS} , we have:

Claim 6.4. *With probability at least $1 - \delta$ we have, $\text{OPT}_{\text{ALG}, t} \geq \text{OPT}_{\text{atoms}} \geq \text{OPT}_{\text{BwK}}(B_\epsilon)$.*

Proof. We will first prove the second inequality.

Consider the optimal solution vector \mathbf{x} to $\text{LP}_{\text{BwK}}(B_\epsilon)$. Define $S^* := \{S : x(S) \neq 0\}$.

We will now map this to a feasible solution to LP_{ATOMS} and show that the objective value does not decrease. This will then complete the claim. Consider the following solution \mathbf{y} defined as follows.

$$y(a) = \sum_{S \in S^* : a \in S} x(S).$$

We will now show that \mathbf{y} is a feasible solution to the polytope \mathcal{P} . From the definition of \mathbf{y} , we can write it as $\mathbf{y} = \sum_{S \in S^*} x(S) \times \mathbb{I}[S]$. Here, $\mathbb{I}[S]$ is a binary vector, such that it has 1 at position a if and only if atom a is present in set S . Hence, \mathbf{y} is a point in the polytope since it can be written as convex combination of its vertices.

Now, we will show that, \mathbf{y} also satisfies the resource consumption constraint.

$$\begin{aligned} C(\mathbf{j}) \cdot \mathbf{y} &= \sum_{a \in A} C(a, j) \sum_{S \in S^* : a \in S} x(S) \\ &= \sum_{S \in S^*} \sum_{a \in S} C(a, j) x(S) \\ &= \sum_{S \in S^*} C(S, j) x(S) \leq B_\epsilon / T. \end{aligned}$$

The last inequality is because in the optimal solution, the x value corresponding to subset S^* is 1 while rest all are 0. We will now show that \mathbf{y} produces an objective value at least as large as \mathbf{x} .

$$\begin{aligned} \text{OPT}_{\text{atoms}} &= \boldsymbol{\mu} \cdot \mathbf{y}^* \geq \boldsymbol{\mu} \cdot \mathbf{y} = \sum_{a=1}^n \mu(a) \sum_{S \in S^* : a \in S} x(S) \\ &= \sum_{S \in S^*} \sum_{a \in S} \mu(a) x(S) = \sum_{S \in S^*} \mu(S) x(S) \\ &= \text{OPT}_{\text{subsets}}(B_\epsilon) \end{aligned}$$

Now we will prove the first inequality. We will assume the ‘‘clean event’’ that $\boldsymbol{\mu}_t^+ \geq \boldsymbol{\mu}$ and $C_t^- \leq C_t$ for all t . Hence, the inequality holds with probability at least $1 - \delta$.

Consider a time t . Given an optimal solution \mathbf{x}^* to LP_{ATOMS} we will show that this is feasible to $\text{LP}_{\text{ALG}, t}$. Note that, \mathbf{x}^* satisfies the constraint set $\mathbf{x} \in \mathcal{P}$ since that is same for both $\text{LP}_{\text{ALG}, t}$ and LP_{ATOMS} . Now consider the constraint $C_t^-(\mathbf{j}) \cdot \mathbf{x} \leq \frac{B_\epsilon}{T}$. Note that $C_t^-(a, j) \leq C(a, j)$. Hence, we have that $C_t^-(\mathbf{j}) \cdot \mathbf{x}^* \leq C(\mathbf{j}) \cdot \mathbf{x}^* \leq \frac{B_\epsilon}{T}$. The last inequality is because \mathbf{x}^* is a feasible solution to LP_{ATOMS} .

Now consider the objective function. Let \mathbf{y}^* denote the optimal solution to $\text{LP}_{\text{ALG}, t}$.

$$\text{OPT}_{\text{ALG}, t} = \boldsymbol{\mu}_t^+ \cdot \mathbf{y}^* \geq \boldsymbol{\mu}_t^+ \cdot \mathbf{x}^* \geq \boldsymbol{\mu} \cdot \mathbf{y}^* = \text{OPT}_{\text{atoms}}. \quad \square$$

Hence, combining Claim 6.3 and Claim 6.4, we obtain Lemma 6.2.

6.2 Negative correlation and concentration bounds

Our analysis relies on several facts about negative correlation and concentration bounds. First, we argue that property (2.1) in the definition of negative correlation is preserved under a specific linear transformation:

Claim 6.5. Suppose (X_1, X_2, \dots, X_m) is a family of negatively correlated random variables with support $[0, 1]$. Fix numbers $\lambda_1, \lambda_2, \dots, \lambda_m \in [0, 1]$. Consider two families of random variables:

$$\mathcal{F}^+ = \left(\frac{1 + \lambda_i(X_i - \mathbb{E}[X_i])}{2} : i \in [m] \right) \quad \text{and} \quad \mathcal{F}^- = \left(\frac{1 - \lambda_i(X_i - \mathbb{E}[X_i])}{2} : i \in [m] \right).$$

Then both families satisfy property (2.1).

Proof. Let us focus on family \mathcal{F}^+ ; the proof for family \mathcal{F}^- is very similar.

Denote $\mu_i = \mathbb{E}[X_i]$ and $Y_i := (1 + \lambda_i(X_i - \mu_i))/2$ and $z_i := (1 - \lambda_i\mu_i)/2$ for all $i \in [m]$. Note that $Y_i = \lambda_i X_i/2 + z_i$ and $z_i \geq 0, X_i \geq 0$. Fix a subset $S \subseteq [m]$. We have,

$$\begin{aligned} \mathbb{E} \left[\prod_{i \in S} Y_i \right] &= \mathbb{E} \left[\sum_{T \subseteq S} \prod_{i \in T} (\lambda_i X_i/2) \prod_{j \in S \setminus T} z_j \right] && \text{by Binomial Theorem} \\ &= \sum_{T \subseteq S} \mathbb{E} \left[\prod_{i \in T} (\lambda_i X_i/2) \right] \prod_{j \in S \setminus T} z_j \\ &\leq \sum_{T \subseteq S} \prod_{i \in T} (\lambda_i \mu_i/2) \prod_{j \in S \setminus T} z_j && (2.1) \text{ invariant under non-negative scaling, } X_i \text{ neg. correlated} \\ &= \prod_{i \in S} ((1 - \lambda_i \mu_i)/2 + \lambda_i \mu_i/2) && \text{by Binomial Theorem} \\ &= \left(\frac{1}{2}\right)^{|S|} = \prod_{i \in S} \mathbb{E}[Y_i] && \square \end{aligned}$$

Second, we extend Theorem 2.1 to a random process that evolves over time, and only assumes that property (2.3) holds within each round conditional on the history.

Theorem 6.6. Let $\mathcal{Z}_T = \{\zeta_{t,a} : a \in \mathcal{A}, t \in [T]\}$ be a family of random variables taking values in $[0, 1]$. Assume random variables $\{\zeta_{t,a} : a \in \mathcal{A}\}$ satisfy property (2.1) given \mathcal{Z}_{t-1} and have expectation $\frac{1}{2}$ given \mathcal{Z}_{t-1} , for each round t . Let $Z = \frac{1}{nT} \sum_{a \in \mathcal{A}, t \in [T]} \zeta_{t,a}$ be the average. Then for some absolute constant c ,

$$\Pr[Z \geq \frac{1}{2} + \eta] \leq c \cdot e^{-2m\eta^2} \quad (\forall \eta > 0). \quad (6.2)$$

Proof. We prove that family \mathcal{Z}_t satisfies property (2.3), and then invoke Theorem 2.1. Let us restate property (2.3) for the sake of completeness:

$$\mathbb{E} \left[\prod_{(t,a) \in S} \zeta_{t,a} \right] \leq 2^{-|S|} \quad \text{for any subset } S \subseteq \mathcal{Z}_T. \quad (6.3)$$

Fix subset $S \subseteq \mathcal{Z}_T$. Partition S into subsets $S_t = \{\zeta_{t,a} \in S\}$, for each round t . Fix round τ and denote

$$G_\tau = \prod_{t \in [\tau]} H_t, \quad \text{where } H_t = \prod_{a \in S_t} \zeta_{t,a}.$$

We will now prove the following statement by induction on τ :

$$\mathbb{E}[G_\tau] \leq 2^{-k_\tau}, \quad \text{where } k_\tau = \sum_{t \in [\tau]} |S_t|. \quad (6.4)$$

The base case is when $\tau = 1$. Note that G_τ is just the product of elements in set ζ_1 and they are negatively correlated from the premise. Therefore we are done. Now for the inductive case of $\tau \geq 2$,

$$\mathbb{E}[H_\tau | \mathcal{Z}_{\tau-1}] \leq \prod_{a \in S_\tau} \mathbb{E}[\zeta_{\tau,a} | \mathcal{Z}_{\tau-1}] \quad \text{From property (2.1) in the conditional space} \quad (6.5)$$

$$\leq 2^{-|S_\tau|} \quad \text{From assumption in Lemma 6.6} \quad (6.6)$$

Therefore, we have

$$\begin{aligned}
 \mathbb{E}[G_\tau] &= \mathbb{E}[\mathbb{E}[G_{\tau-1}H_\tau|\mathcal{Z}_{\tau-1}]] && \text{Law of iterated expectation} \\
 &= \mathbb{E}[G_{\tau-1}\mathbb{E}[H_\tau|\mathcal{Z}_{\tau-1}]] && \text{Since } G_{\tau-1} \text{ is a fixed value conditional on } \mathcal{Z}_{\tau-1} \\
 &\leq 2^{-|S_\tau|}\mathbb{E}[G_{\tau-1}] && \text{From Eq. 6.6} \\
 &\leq 2^{-k_\tau} && \text{From inductive hypothesis}
 \end{aligned}$$

This completes the proof of Eq. 6.4. We obtain Eq. 6.3 for $\tau = T$. \square

Third, we invoke Eq. 2.6 for rewards and resource consumptions:

Lemma 6.7. *With probability at least $1 - e^{-\Omega(\alpha)}$, we have the following:*

$$\begin{aligned}
 |\hat{\mu}_t(a) - \mu_t(a)| &\leq 2 \text{Rad}(\hat{\mu}_t(a), N_t(a) + 1) \\
 \forall j \in [d] \quad |\hat{C}_t(a, j) - C_t(a, j)| &\leq 2 \text{Rad}(\hat{C}_t(a, j), N_t(a) + 1).
 \end{aligned} \tag{6.7}$$

Fourth, we use a concentration bound from prior work which gets sharper when the expected sum is very small, and does not rely on independent random variables:

Theorem 6.8 (Babaioff et al. [2015]). *Let X_1, X_2, \dots, X_m denote a set of $\{0, 1\}$ random variables. For each t , let α_t denote the multiplier determined by random variables X_1, X_2, \dots, X_{t-1} . Let $M = \sum_{t=1}^m M_t$ where $M_t = \mathbb{E}[X_t|X_1, X_2, \dots, X_{t-1}]$. Then for any $b \geq 1$, we have the following with probability at least $1 - m^{-\Omega(b)}$:*

$$\left| \sum_{t=1}^m \alpha_t (X_t - M_t) \right| \leq b(\sqrt{M \log m} + \log m)$$

6.3 Analysis of the ‘‘clean event’’

Let us set up several events, henceforth called *clean events*, and prove that they hold with high probability. Then the remainder of the analysis can proceed conditional on the intersection of these events. The clean events are similar to the ones in Agrawal and Devanur [2014b], but are somewhat ‘‘stronger’’, essentially because our algorithm has access to per-atom feedback and our analysis can use the negative correlation property of the RRS.

In what follows, it is convenient to consider a version of SemiBwK in which the algorithm does not stop, so that we can argue about what happens w.h.p. if our algorithm runs for the full T rounds. Then we show that our algorithm does indeed run for the full T rounds w.h.p.

Recall that \mathbf{x}_t be the optimal fractional solution obtained by solving the LP in round t . Let $\mathbf{Y}_t \in \{0, 1\}^n$ be the random binary vector obtained by invoking the RRS (so that the chosen action $S_t \in \mathcal{F}$ corresponds to a particular realization of \mathbf{Y}_t , interpreted as a subset). Let $\mathcal{G}_t := \{\mathbf{Y}_{t'} : \forall t' \leq t\}$ denote the family of RRS realizations up to round t .

6.3.1 ‘‘Clean event’’ for rewards

For brevity, for each round t let $\boldsymbol{\mu}_t = (\mu_t(a) : a \in A)$ be the vector of realized rewards, and let $r_t := \mu_t(S_t) = \boldsymbol{\mu}_t \cdot \mathbf{Y}_t$ be the algorithm’s reward at this round.

Lemma 6.9. *Consider SemiBwK without stopping. Then with probability at least $1 - nT e^{-\Omega(\alpha)}$:*

$$\left| \sum_{t \in [T]} r_t - \sum_{t \in [T]} \boldsymbol{\mu}_t^+ \cdot \mathbf{x}_t \right| \leq O\left(\sqrt{\alpha n \sum_{t \in [T]} r_t} + \sqrt{\alpha n T} + \alpha n\right).$$

Proof. We prove the Lemma by proving the following three high-probability inequalities.

With probability at least $1 - nT e^{-\Omega(\alpha)}$: the following holds:

$$\left| \sum_{t \in [T]} r_t - \sum_{t \in [T]} \boldsymbol{\mu} \cdot \mathbf{Y}_t \right| \leq 3nT \text{Rad}\left(\frac{1}{nT} \sum_{t \in [T]} \boldsymbol{\mu}_t^+ \cdot \mathbf{x}_t, nT\right) \tag{6.8}$$

$$\left| \sum_{t \in [T]} \boldsymbol{\mu} \cdot \mathbf{Y}_t - \sum_{t \in [T]} \boldsymbol{\mu}_t^+ \cdot \mathbf{Y}_t \right| \leq 12\sqrt{\alpha n} \left(\sum_{t \in [T]} \boldsymbol{\mu}_t^+ \cdot \mathbf{x}_t \right) + 12\sqrt{\alpha n} + 12\alpha n \tag{6.9}$$

$$\left| \sum_{t \in [T]} \boldsymbol{\mu}_t^+ \cdot \mathbf{Y}_t - \sum_{t \in [T]} \boldsymbol{\mu}_t^+ \cdot \mathbf{x}_t \right| \leq \sqrt{\alpha n T}. \tag{6.10}$$

We will use the properties of RRS to prove Eq. 6.10. Proof of Eq. 6.9 is similar to Agrawal and Devanur [2014b], while proof of Eq. 6.8 follows immediately from the setup of the model. Using the parts (6.8) and (6.10) we can now find an appropriate upper bound on $\sqrt{\sum_{t \in [T]} \boldsymbol{\mu}_t^+ \cdot \mathbf{x}_t}$ and using this upper bound, we prove Lemma 6.9.

Proof of Eq. 6.8. Recall that $r_t = \boldsymbol{\mu}_t \mathbf{Y}_t$. Note that, $\mathbb{E}[\boldsymbol{\mu}_t \mathbf{Y}_t] = \boldsymbol{\mu} \mathbf{Y}_t$ when the expectation is taken over just the independent samples of $\boldsymbol{\mu}$. By Theorem 6.8, with probability $1 - e^{-\Omega(\alpha)}$ we have:

$$\begin{aligned} \left| \sum_{t \leq T} r_t - \sum_{t \leq T} \boldsymbol{\mu} \cdot \mathbf{Y}_t \right| &\leq 3nT \text{Rad} \left(\frac{1}{nT} \sum_{t \leq T} \boldsymbol{\mu} \cdot \mathbf{Y}_t, nT \right) \\ &\leq 3nT \text{Rad} \left(\frac{1}{nT} \sum_{t \leq T} \boldsymbol{\mu}_t^+ \cdot \mathbf{Y}_t, nT \right) \\ &\leq 3nT \text{Rad} \left(\frac{1}{nT} \sum_{t \leq T} \boldsymbol{\mu}_t^+ \cdot \mathbf{x}_t, nT \right). \end{aligned}$$

The last inequality is because \mathbf{Y}_t is a feasible solution to LP_{ALG} .

Proof of Eq. 6.9. For this part, the arguments similar to Agrawal and Devanur [2014b] follow with some minor adaptations. For sake of completeness we describe the full proof. Note that we have,

$$\left| \sum_{t \leq T} \boldsymbol{\mu} \cdot \mathbf{Y}_t - \boldsymbol{\mu}_t^+ \cdot \mathbf{Y}_t \right| \leq \sum_{a=1}^n \left| \sum_{t \leq T} \mu(a) Y_t(a) - \mu_t^+(a) Y_t(a) \right|.$$

Now, using Lemma 6.7 in Appendix, we have that with probability $1 - nT e^{-\Omega(\alpha)}$

$$\left| \sum_{t \leq T} \mu(a) Y_t(a) - \mu_t^+(a) Y_t(a) \right| \leq 12 \sum_{t \leq T} \text{Rad}(\mu(a), N_t(a) + 1).$$

Hence, we have

$$\begin{aligned} \sum_{a=1}^n \left| \sum_{t \leq T} \mu(a) Y_t(a) - \mu_t^+(a) Y_t(a) \right| &= 12 \sum_{a \in \mathcal{A}} \sum_{r=1}^{N_T(a)+1} \text{Rad}(\mu(a), r) \\ &\leq 12 \sum_{a \in \mathcal{A}} (N_T(a) + 1) \text{Rad}(\mu(a), N_T(a) + 1) \\ &\leq 12 \sqrt{\alpha n (\boldsymbol{\mu} \cdot (\mathbf{N}_T + \mathbf{1}))} + 12\alpha n. \end{aligned}$$

The last inequality is from the definition of Rad function and using the Cauchy-Swartz inequality. Note that $\boldsymbol{\mu} \mathbf{N}_T = \sum_{t \leq T} \boldsymbol{\mu} \cdot \mathbf{Y}_t$. Also, since we have with probability $1 - e^{-\Omega(\alpha)}$, $\mu(a) \leq \mu_t^+(a)$, we have,

$$12 \sqrt{\alpha n (\boldsymbol{\mu} \cdot (\mathbf{N}_T + \mathbf{1}))} + 12\alpha n \leq 12 \sqrt{\alpha n \left(\sum_{t \leq T} \boldsymbol{\mu}_t^+ \cdot \mathbf{Y}_t \right)} + 12 \sqrt{\alpha n} + 12\alpha n.$$

Finally note that \mathbf{Y}_t is a feasible solution to the semi-bandit polytope \mathcal{P} . Hence, we have that

$$\boldsymbol{\mu}_t^+ \cdot \mathbf{Y}_t \leq \boldsymbol{\mu}_t^+ \cdot \mathbf{x}_t.$$

Hence,

$$12 \sqrt{\alpha n \left(\sum_{t \leq T} \boldsymbol{\mu}_t^+ \cdot \mathbf{Y}_t \right)} + 12 \sqrt{\alpha n} + 12\alpha n \leq 12 \sqrt{\alpha n \left(\sum_{t \leq T} \boldsymbol{\mu}_t^+ \cdot \mathbf{x}_t \right)} + 12 \sqrt{\alpha n} + 12\alpha n.$$

Proof of Eq. 6.10: Recall that for each round t , the UCB vector $\boldsymbol{\mu}_t^+$ is determined by the random variables $\mathcal{G}_{t-1} = \{\mathbf{Y}_{t'} : \forall t' < t\}$. Further, conditional on a realization of \mathcal{G}_{t-1} , the random variables $\{Y_t(a) : a \in \mathcal{A}\}$ are negatively correlated from the property of RRS. Let $\tilde{\zeta}_t(a) := \mu_t^+(a) Y_t(a)$, $a \in \mathcal{A}$. Note that we have $\mathbb{E}[\tilde{\zeta}_t(a) | \mathcal{G}_{t-1}] = \mu_t^+(a) x_t(a)$. Define

$$\zeta_t(a) := \frac{1 + \mu_t^+(a) Y_t(a) - \mu_t^+(a) x_t(a)}{2}.$$

From Claim 6.5, we have that $\{\zeta_t(a) : a \in \mathcal{A}\}$ conditioned on \mathcal{G}_{t-1} satisfy (2.1). Further, $\mathbb{E}[\zeta_t(a) | \mathcal{G}_{t-1}] = \frac{1}{2}$. Therefore, the family $\{\zeta_t(a) : t \in [T], a \in \mathcal{A}\}$ satisfies the assumptions in Theorem 6.6 and hence satisfies Eq. 6.2 for some absolute constant c . Plugging back the $\tilde{\zeta}_t(a)$'s, we obtain an upper-tail concentration bound:

$$\Pr \left[\frac{1}{nT} \left(\sum_{t=1}^T \sum_{a \in \mathcal{A}} \tilde{\zeta}_t(a) - \mu_t^+(a) x_t(a) \right) \geq \eta \right] \leq c \cdot e^{-2nT\eta^2}.$$

To obtain a corresponding concentration bound for the lower tail, we apply a similar argument to

$$\zeta'_t(a) = \frac{1 + \mu_t^+(a) x_t(a) - \tilde{\zeta}_t(a)}{2}.$$

Once again from Claim 6.5, we have that $\{\zeta'_t(a) : a \in \mathcal{A}\}$ conditioned on \mathcal{G}_{t-1} satisfy (2.1). The family $\{\zeta'_t(a) : t \in [T], a \in \mathcal{A}\}$ satisfies the assumptions in Theorem 6.6 and hence satisfies Eq. 6.2. Plugging back the $\tilde{\zeta}_t(a)$'s, we obtain a lower-tail concentration bound:

$$\Pr \left[\frac{1}{nT} \left(\sum_{t=1}^T \sum_{a \in \mathcal{A}} \mu_t^+(a) x_t(a) - \tilde{\zeta}_t(a) \right) \geq \eta \right] \leq c \cdot e^{-2nT\eta^2}.$$

Combining these two we have,

$$\Pr \left[\frac{1}{nT} \left| \sum_{t=1}^T \sum_{a \in \mathcal{A}} \mu_t^+(a) Y_t(a) - \mu_t^+(a) x_t(a) \right| \geq \eta \right] \leq 2c \cdot e^{-2nT\eta^2}. \quad (6.11)$$

Hence setting $\eta = \sqrt{\frac{\alpha}{nT}}$, we obtain Eq. 6.10 with probability at least $1 - e^{-\Omega(\alpha)}$.

Combining Eq. (6.8), (6.9) and (6.10) Let us denote $H := \sqrt{\sum_{t \in [T]} \mu_t^+ \cdot \mathbf{x}_t}$. Adding the three equations we get

$$\left| \sum_{t \in [T]} r_t - H^2 \right| \leq \sqrt{\alpha} H + \alpha + \sqrt{\alpha n} H + O(\alpha n) + \sqrt{\alpha n T} \quad (6.12)$$

Rearranging and solving for H , we have

$$H \leq \sqrt{\sum_{t \in [T]} r_t} + O(\sqrt{\alpha n}) + (\alpha n T)^{1/4}$$

Plugging this back into Eq. 6.12, we get Lemma 6.9. \square

6.3.2 “Clean event” for resource consumption

We define a similar “clean event” for consumption of each resource j . By a slight abuse of notation, for each round t let $\mathbf{C}_t(j) = (C_t(a, j) : a \in \mathcal{A})$ be the vector of realized consumption of resource j . Let $\chi_t(j)$ denote algorithm’s consumption for resource j at round t (i.e., $\chi_t(j) = \mathbf{C}_t(j) \cdot \mathbf{Y}_t$).

Lemma 6.10. *Consider SemiBwK without stopping. Then with probability at least $1 - nT e^{-\Omega(\alpha)}$:*

$$\forall j \in [d] \quad \left| \sum_{t \in [T]} \chi_t(j) - \sum_{t \in [T]} \mathbf{C}_t^-(j) \cdot \mathbf{x}_t \right| \leq \sqrt{\alpha n B_\epsilon} + \alpha n + \sqrt{\alpha n T}.$$

Proof. The proof is similar to Lemma 6.9. We will split the proof into following three equations. Fix an arbitrary resource $j \in [d]$. With probability at least $1 - nT e^{-\Omega(\alpha)}$ the following holds:

$$\left| \sum_{t \leq T} \chi_t(j) - \sum_{t \leq T} \mathbf{C}(j) \cdot \mathbf{Y}_t \right| \leq 3nT \text{Rad} \left(\frac{1}{nT} \sum_{t \leq T} \mathbf{C}(j) \cdot \mathbf{Y}_t, nT \right). \quad (6.13)$$

$$\left| \sum_{t \leq T} \mathbf{C}(j) \cdot \mathbf{Y}_t - \mathbf{C}_t^-(j) \cdot \mathbf{Y}_t \right| \leq 12 \sqrt{\alpha n \left(\sum_{t \leq T} \mathbf{C}(j) \cdot \mathbf{Y}_t \right)} + 12 \sqrt{\alpha n} + 12 \alpha n. \quad (6.14)$$

$$\left| \sum_{t \leq T} \mathbf{C}_t^-(j) \cdot \mathbf{Y}_t - \mathbf{C}_t^-(j) \cdot \mathbf{x}_t \right| \leq \sqrt{\alpha n T}. \quad (6.15)$$

Using the parts 6.13, 6.14 and 6.15 we can find an upper bound on $\sqrt{\sum_{t \leq T} \mathbf{C}_t(j) \cdot \mathbf{Y}_t}$. Hence, combining Lemmas 6.13, 6.14 and 6.15 with this bound and taking an Union Bound over all the resources, we get Lemma 6.10.

Proof of Eq. 6.13. We have that $\{C_t(a, j) : a \in \mathcal{A}\}$ is a set of independent random variables over a probability space C_Ω . Note that, $\mathbb{E}_{C_\Omega} C_t(a, j) Y_t(a) = C(a, j) Y_t(a)$. Hence, we can invoke Theorem 6.8 on *independent random variables* to get with probability $1 - nT e^{-\Omega(\alpha)}$

$$|\sum_{t \leq T} \chi_t(j) - \sum_{t \leq T} \mathbf{C}(j) \cdot \mathbf{Y}_t| \leq 3nT \text{Rad} \left(\frac{1}{nT} \sum_{t \leq T} \mathbf{C}(j) \cdot \mathbf{Y}_t, nT \right).$$

Proof of Eq. 6.14. This is very similar to proof of 6.9 and we will skip the repetitive parts. Hence, we have with probability $1 - nT e^{-\Omega(\alpha)}$

$$\begin{aligned} |\sum_{t \leq T} \mathbf{C}(j) \cdot \mathbf{Y}_t - \mathbf{C}_t^-(j) \cdot \mathbf{Y}_t| &\leq 12\sqrt{\alpha n(\mathbf{C}(j) \cdot (N_T + \mathbf{1}))} + 12\alpha n \\ &\leq 12\sqrt{\alpha n \left(\sum_{t \leq T} \mathbf{C}(j) \cdot \mathbf{Y}_t \right)} + 12\sqrt{\alpha n} + 12\alpha n. \end{aligned}$$

Proof of Eq. 6.15. Recall that for each round t and each resource j , the LCB vector $\mathbf{C}_t^-(j)$ is determined by the random variables $\mathcal{G}_{t-1} = \{\mathbf{Y}_{t'} : \forall t' < t\}$. Similar to the proof of Eq. 6.10, random variables $\{Y_t(a) : a \in \mathcal{A}\}$ obtained from the RRS are negatively correlated given \mathcal{G}_{t-1} . As before define $\tilde{\zeta}_t(a) = C_t^-(a) Y_t(a)$, $a \in \mathcal{A}$. We have that $\mathbb{E}[\zeta_t(a) | \mathcal{G}_{t-1}] = C_t^-(a) x_t(a)$.

By Claim 6.5, random variables

$$\zeta_t(a) = \frac{1 + \tilde{\zeta}_t(a) - C_t^-(a) x_t(a)}{2}, \quad a \in \mathcal{A}$$

satisfy (2.1), given \mathcal{G}_{t-1} . We conclude that family $\{\zeta_t(a) : t \in [T], a \in \mathcal{A}\}$ satisfies the assumptions in Theorem 6.6, and therefore satisfies Eq. 6.2 for some absolute constant c . Therefore, we obtain an upper-tail concentration bound for $\tilde{\zeta}_t(a)$'s:

$$\Pr \left[\frac{1}{nT} \left(\sum_{t=1}^T \sum_{a \in \mathcal{A}} \tilde{\zeta}_t(a) - C_t^-(a) x_t(a) \right) \geq \eta \right] \leq c \cdot e^{-2nT\eta^2}.$$

To obtain a corresponding concentration bound for the lower tail, we apply a similar argument to

$$\zeta'_t(a) = \frac{1 + C_t^-(a) x_t(a) - \tilde{\zeta}_t(a)}{2}.$$

Once again, invoking Claim 6.5 we have that $\{\zeta'_t(a) : a \in \mathcal{A}\}$ conditioned on \mathcal{G}_{t-1} satisfy (2.1). Thus, family $\{\zeta_t(a) : t \in [T], a \in \mathcal{A}\}$ satisfies the assumptions in Theorem 6.6, and therefore satisfies Eq. 6.2. We obtain:

$$\Pr \left[\frac{1}{nT} \left(\sum_{t=1}^T \sum_{a \in \mathcal{A}} C_t^-(a) x_t(a) - \tilde{\zeta}_t(a) \right) \geq \eta \right] \leq c \cdot e^{-2nT\eta^2}.$$

Combing the two tails we have,

$$\Pr \left[\frac{1}{nT} \left| \sum_{t=1}^T \sum_{a \in \mathcal{A}} C_t^-(a) Y_t(a) - C_t^-(a) x_t(a) \right| \geq \eta \right] \leq 2c \cdot e^{-2nT\eta^2}. \quad (6.16)$$

Once again, setting $\eta = \sqrt{\frac{\alpha}{nT}}$, we obtain Eq. 6.15 with probability at least $1 - e^{-\Omega(\alpha)}$.

Proof of Lemma 6.10. Denote $G = \sqrt{\sum_{t \leq T} \mathbf{C}(j) \cdot \mathbf{Y}_t}$. From Equation 6.13, 6.14 and 6.15, we have that $G^2 - 2\Omega(\sqrt{\alpha n})G \leq \sum_{t \leq T} \mathbf{C}_t^-(j) \cdot \mathbf{x}_t + O(\alpha n) + \sqrt{\alpha n T}$. Note that $\sum_{t \leq T} \mathbf{C}_t^-(j) \cdot \mathbf{x}_t \leq B_\epsilon$. Hence, $G^2 - 2\Omega(\sqrt{\alpha n})G \leq B_\epsilon + O(\alpha n) + \sqrt{\alpha n T}$. Hence, re-arranging this gives us $G \leq \sqrt{B_\epsilon} + O(\sqrt{\alpha n}) + (\alpha n T)^{1/4}$. Plugging this back in Equations 6.13, 6.14 and 6.15, we get Lemma 6.10. \square

6.4 Putting it all together

Similar to Agrawal and Devanur [2014b], we will handle the hard constraint on budget, by choosing an appropriate value of ϵ . We then combine the above Lemma on "rewards" clean event to compare the reward of the algorithm with that of the optimal value of LP to obtain the regret bound in Theorem 6.1. Additionally, we use the Lemma on "consumption" clean event to argue that the algorithm doesn't exhaust the resource budget before round T . Formally, consider the following.

Recall that from Lemma 6.2, we have $\text{OPT}_{\text{ALG}} \geq \frac{1}{T}(1 - \epsilon) \text{OPT}$. Let us define the performance of the algorithm as $\text{ALG} = \sum_{t \leq T} r_t$. From Lemma 6.9, that with probability at least $1 - ndT e^{-\Omega(\alpha)}$

$$\begin{aligned} \text{ALG} &\geq (1 - \epsilon) \text{OPT} - O(\sqrt{\alpha n \text{ALG}}) - O(\alpha n) - \sqrt{\alpha n T} \\ &\geq (1 - \epsilon) \text{OPT} - O(\sqrt{\alpha n \text{OPT}}) - O(\alpha n) - \sqrt{\alpha n T} \quad (\text{since } \text{ALG} \leq \text{OPT}). \end{aligned}$$

Choosing $\epsilon = \sqrt{\frac{\alpha n}{B}} + \frac{\alpha n}{B} + \frac{\sqrt{\alpha n T}}{B}$ and using the assumption that $B > 3(\alpha n + \sqrt{\alpha n T})$, we derive Eq. 6.1. For any given δ , we set $\alpha = \Omega(\log(\frac{ndT}{\delta}))$ to obtain a success probability of at least $1 - \delta$.

Now we will argue that the algorithm does not exhaust the resource budget before round T with probability at least $1 - ndT e^{-\Omega(\alpha)}$. Note that for every resource $j \in [d]$,

$$\sum_{t \leq T} C_t^-(j) \cdot x_t \leq (1 - \epsilon)B.$$

Hence, combining this with Lemma 6.10, we have $\sum_{t \leq T} C_t(j) Y_t \leq (1 - \epsilon)B + \epsilon B \leq B$.

7 Details for Numerical Simulations

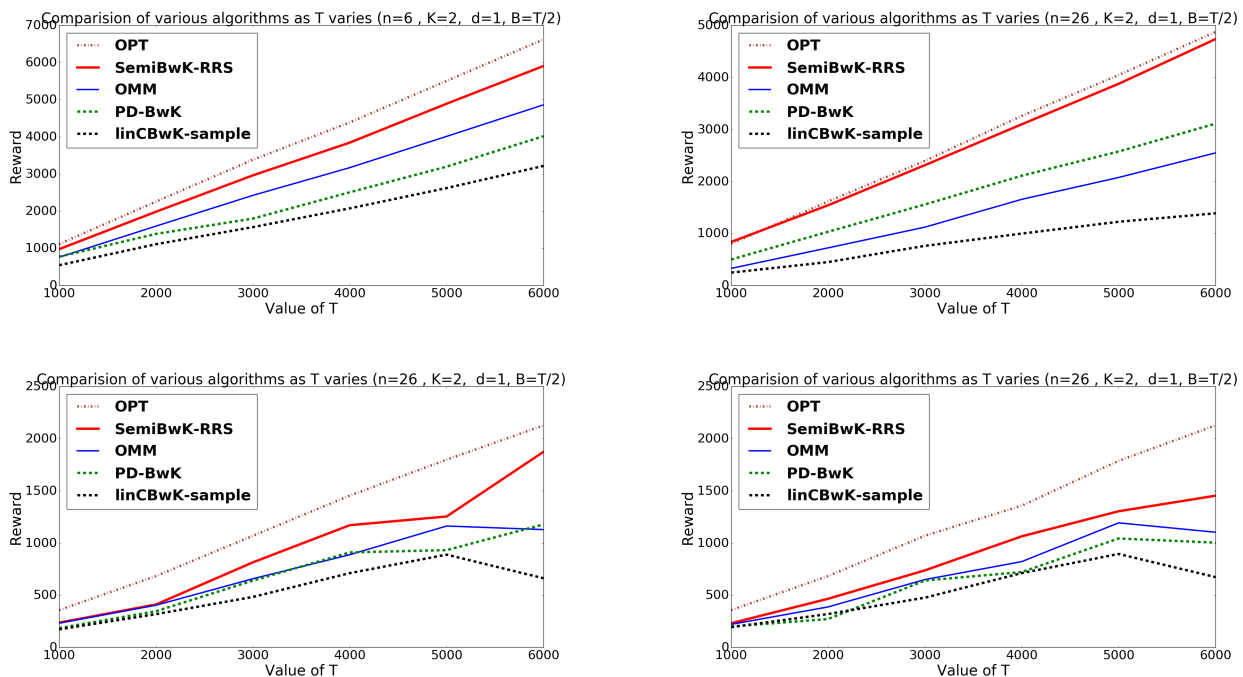


Figure 3: Experimental Results for Uniform matroid (left plots) and Partition matroid (right plots) on independent (upper) and correlated (lower) instances for $n = 26$. Here OPT denotes LP_{OPT} .

Details of heuristic implementation of linCBwK. We now briefly describe the heuristic we use to simulate the linCBwK algorithm. Note that even though the per-time-step running time of linCBwK is reasonable, it takes a significant time when we want to perform simulations for many time-steps. The time-consuming step in the linCBwK algorithm is the solution to

a convex program for computing the optimistic estimates (namely $\tilde{\mu}_t$ and \tilde{W}_t). Hence, the heuristic gives a faster way to obtain this estimate. We sample multiple times from a multi-variate Gaussian with mean $\hat{\mu}$ and covariance M_t (to obtain estimate $\tilde{\mu}_t$) and with mean \hat{w}_{tj} and covariance M_t (to obtain estimate \tilde{w}_{tj} for each resource j). We use these samples to compute the objective to choose the action at time-step t . For each sample, we compute the best action based on the objective in `linCBwK`. We finally choose the action that occurs majority number of times in these samples. The number of samples we choose is set to 30.

Language Details of algorithms. All algorithms except `linCBwK` were implemented in Python. The `linCBwK` algorithm was implemented in MATLAB. This difference is crucial when we compare running times since language construct can speed-up or slow down algorithms in practice. However, it is known that¹⁰ for matrix operations commonly encountered in engineering and statistics, MATLAB implementations runs several orders of magnitude faster than the corresponding python implementation. Since `linCBwK` is the slowest of the four algorithms, our comparison of running times across languages is justified.

Further results. We now show additional plots omitted in the main section in Figures 3 and 4. In particular we show the variation of rewards of various algorithms when $n = 6$ and $n = 26$ on problem instances to be defined. As before, our algorithm has the best performance across the algorithms in all settings.

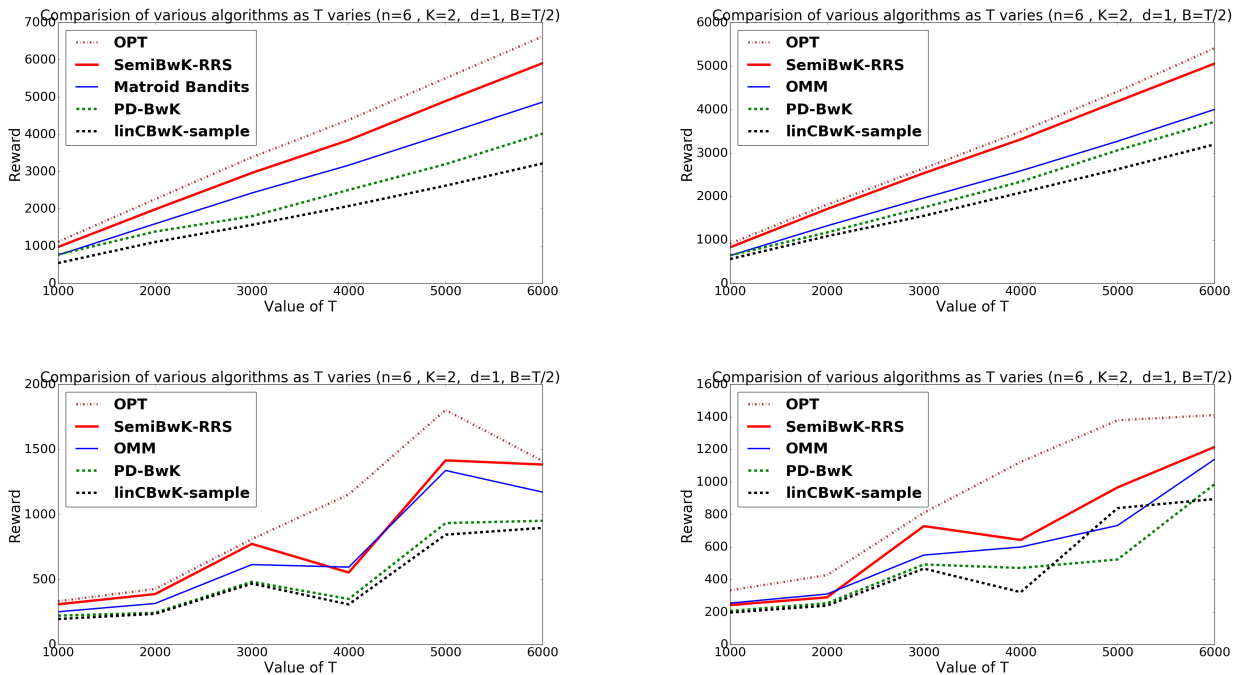


Figure 4: Experimental Results for Uniform matroid (left plots) and partition matroid (right plots) on independent (upper) and correlated (lower) instances for $n = 6$.

The first family was inspired by the dynamic assortment application. As in dynamic assortment, we have n products, and for each product i there is an atom i and a resource i . The (fixed) price for each product is generated as an independent sample from $U_{[0,1]}$, a uniform distribution on $[0, 1]$. At each round, we sample the buyers's valuation from $U_{[0,1]}$, independently for each product. If the valuation for a given product is greater than the price, one item of this product is sold (and then the reward for atom i is the price, and consumption of resource i is 1). Else, we do something different from dynamic assortment: we set reward for atom i and consumption for resource i to be the buyer's valuation.

The second family was inspired by the dynamic pricing application with two products. We have $n/2$ allowed prices, uniformly spaced in the $[0, 1]$ interval. Recall that atoms correspond to price-product pairs, for the total of n atoms. In each round t , the valuation $v_{t,i}$ for each product i is chosen independently from a normal distribution $\mathcal{N}(v_i^0, 1)$ truncated on $[0, 1]$. The mean valuation v_i^0 is drawn (once for all rounds) from $U_{[0,1]}$. If the valuation for a given product i is greater

¹⁰<https://www.mathworks.com/products/matlab/matlab-vs-python.html>

than the offered price p , one item of this product is sold (and then reward for the corresponding atom (p, i) is the price, and consumption of product i is 1). If there is no sale for this product, we do something different from dynamic pricing. For each atom (p, i) , if $p < v_{t,i}$ then the reward for atom (p, i) is drawn independently from $U_{[0,1]}$ and resource consumption is 1; else, reward is 0 and consumption is .3. While dynamic assortment is modeled with a uniform matroid, and dynamic pricing is modeled with a partition matroid, we tried both matroids on each family.

A Proof of Theorem in Preliminaries

Theorem 2.1 follows easily from Theorem 3.3 in Impagliazzo and Kabanets [2010].

Theorem (Theorem 2.1). *Let $\mathcal{X} = (X_1, X_2, \dots, X_m)$ denote a collection of random variables which take values in $[0, 1]$, and let $X := \frac{1}{m} \sum_{i=1}^m X_i$ be their average. Suppose \mathcal{X} satisfies (2.3), i.e., $\mathbb{E}[\prod_{i \in S} X_i] \leq (\frac{1}{2})^{|S|}$ for every $S \subseteq [m]$. Then for some absolute constant c ,*

$$\Pr[X \geq \frac{1}{2} + \eta] \leq c \cdot e^{-2m\eta^2} \quad (\forall \eta > 0). \quad (\text{A.1})$$

Proof. Fix $\eta > 0$. From Theorem 3.3 in Impagliazzo and Kabanets [2010], we have that

$$\Pr[X \geq \frac{1}{2} + \eta] \leq c \cdot e^{-m \mathcal{D}_{\text{KL}}(1/2 + \eta \parallel 1/2)},$$

where $\mathcal{D}_{\text{KL}}(\cdot \parallel \cdot)$ denotes KL-divergence, so that

$$\mathcal{D}_{\text{KL}}(\frac{1}{2} + \eta \parallel \frac{1}{2}) = (\frac{1}{2} + \eta) \log(1 + 2\eta) + (\frac{1}{2} - \eta) \log(1 - 2\eta). \quad (\text{A.2})$$

From Pinsker's inequality we have, $\mathcal{D}_{\text{KL}}(1/2 + \eta \parallel 1/2) \geq 2\eta^2$, which implies (A.1). □

B Matroid constraints

To make this paper more self-contained, we provide more background on matroid constraints and special cases thereof.

Recall that in `SemiBwK`, we have a finite ground set whose elements are called ‘‘atoms’’, and a family \mathcal{F} of ‘‘feasible subsets’’ of the ground set which are the actions. To be consistent with the literature on matroids, the ground set will be denoted E . Family \mathcal{F} of subsets of E is called a matroid if it satisfies the following properties:

- **Empty set:** The empty set \emptyset is present in \mathcal{F}
- **Hereditary property:** For two subsets $X, Y \subseteq E$ such that $X \subseteq Y$, we have that

$$Y \in \mathcal{F} \implies X \in \mathcal{F}$$

- **Exchange property:** For $X, Y \in \mathcal{F}$ and $|X| > |Y|$, we have that

$$\exists e \in X \setminus Y : Y \cup \{e\} \in \mathcal{F}$$

Matroids are *linearizable*, i.e., the convex hull of \mathcal{F} forms a polytope in \mathbb{R}^E . (Here subsets of \mathcal{F} are interpreted as binary vectors in \mathbb{R}^E .) In other words, there exists a set of linear constraints whose set of feasible *integral* solutions is \mathcal{F} . In fact, the convex hull of \mathcal{F} , a.k.a. the *matroid polytope*, can be represented via the following linear system:

$$\begin{aligned} x(S) &\leq \text{rank}(S) && \forall S \subseteq E \\ x_e &\in [0, 1]^E && \forall e \in E. \end{aligned} \quad (\text{LP-Matroid})$$

Here $x(S) := \sum_{e \in S} x_e$, and $\text{rank}(S) = \max\{|Y| : Y \subseteq S, Y \in \mathcal{F}\}$ is the ‘‘rank function’’ for \mathcal{F} .

\mathcal{F} is indeed the set of all feasible integral solutions of the above system. This is a standard fact in combinatorial optimization, e.g., see Theorem 40.2 and its corollaries in Schrijver [2002].

We will now describe some well-studied special cases of matroids. That they indeed are special cases of matroids is well-known, we will not present the corresponding proofs here.

In all LPs presented below, we have variables x_e for each atom $e \in E$, and we use shorthand $x(S) := \sum_{e \in S} x_e$ for $S \subset E$.

Cardinality constraints. Cardinality constraint is defined as follows: a subset S of atoms belongs to \mathcal{F} if and only if $|S| \leq K$ for some fixed K . This is perhaps the simplest constraint that our results are applicable to. In the context of SemiBwK, each action selects at most K atoms.

The corresponding induced polytope is as follows:

$$\begin{aligned} x(E) &\leq K \\ x_e &\in [0, 1] \quad \forall e \in E. \end{aligned} \tag{LP-Cardinality}$$

Partition matroid constraints. A generalization of cardinality constraints, called partition matroid constraints, is defined as follows. Suppose we have a collection B_1, \dots, B_k of disjoint subsets of E , and numbers d_1, \dots, d_k . A subset S of atoms belongs to \mathcal{F} if and only if $|S \cap B_i| \leq d_i$ for every i . Partition matroid constraints appear in several applications of SemiBwK such as dynamic pricing, adjusting repeated auctions, and repeated bidding. In these applications, each action selects one price/bid for each offered product. Also, partition matroid constraints can model clusters of mutually exclusive products in dynamic assortment application.

The induced polytope is as follows:

$$\begin{aligned} x(B_i) &\leq d_i \quad \forall i \in [k] \\ x_e &\in [0, 1] \quad \forall e \in E. \end{aligned} \tag{LP-PartitionMatroid}$$

Spanning tree constraints. Spanning tree constraints describe spanning trees in a given undirected graph $G = (V, E)$, where the atoms correspond to edges in the graph. A spanning tree in G is a subset $E' \subset E$ of edges such that (V, E') is a tree. Action set \mathcal{F} consists of all spanning trees of G .

The induced polytope is as follows:

$$\begin{aligned} x(E_S) &\leq |S| - 1 \quad \forall S \subseteq V \\ x(E_V) &= |V| - 1 \\ x_e &\in [0, 1] \quad \forall e \in E. \end{aligned} \tag{LP-SpanningTree}$$

Here, E_S denotes the edge set in subgraph induced by node set $S \subset V$.