Supplementary Materials

The following content was not necessarily subject to peer review.

C Auxiliary Results

Lemma 1. *Let χ_m^2 be a chi-squared variable with degree m and $t > 0$, then $\mathbb{P}(\chi_m^2 \geq t) \leq 3^{m/2}\exp(-t/3)$.*

Proof. The moment generating function of χ_m^2 is given by $\mathbb{E}[\exp(\zeta\chi_m^2)] = (1-2\zeta)^{-m/2}$ for $\zeta < 1/2$. Through Markov Inequality we have

$$\mathbb{P}(\chi_m^2 \geq t) \leq \mathbb{E}[\exp(\zeta\chi_m^2)]e^{-\zeta t} = (1-2\zeta)^{-m/2}e^{-\zeta t} \tag{41}$$

The proof is completed by picking $\zeta = 1/3$. □

D Proof of Proposition 1

Define $n_k = n_{K-1}$ for $k > K-1$. Imagine that in each episode, the algorithm pulls all arms still remaining in $\mathcal{X}$ (including virtual arms and known arms). Then the total number of arm pulls is $\sum_{k=1}^{K+L} n_k$.

If an "arm" i is the k-th rejected arm, then it is pulled exactly n_k times. If it is not rejected, then it is pulled n_{K-1} times. Hence to obtain the actual total number of arm pulls, we only need to subtract those n_k's corresponding to virtual arms and known arms from the summation. We conclude that the total number of arm pulls is at most

$$\max_{\substack{\mathcal{J}\subset[K+L]\\|\mathcal{J}|=K-K_0+L}}\left(\sum_{k=1}^{K+L} n_k - \sum_{k\in\mathcal{J}} n_k\right) = \sum_{k=K-K_0+L+1}^{K+L} n_k \tag{42}$$

$$\leq \sum_{k=K-K_0+L+1}^{K+L}\left(1 + \frac{1}{\Psi(K_0,L)}\frac{N-K_0}{\max(2,K+1-k)}\right) \tag{43}$$

$$= K_0 + (N-K_0)\cdot\frac{1}{\Psi(K_0,L)}\sum_{j=L}^{K_0+L-1}\frac{1}{\max(2,K_0-j)} = N \tag{44}$$

E Proof of Proposition 2

The "if" part is clear by the definition of $\mathcal{I}^*$. We establish the "only if" part as follows.

Consider a basis $\mathcal{J} \neq \mathcal{I}^*$. Suppose that $\Delta_{\mathcal{J}}^2 = 0$. Then, there exists a sequence of reward-cost vectors $(\tilde{r}^{(n)}, \tilde{c}^{(n)})_{n=1}^\infty$ such that both of the following hold: (i) $(\tilde{r}^{(n)}, \tilde{c}^{(n)}) \to (r, c)$; (ii) $(\tilde{\mathbf{A}}_{\mathcal{I}^*}^{(n)})^{-1}\mathbf{b} \geq 0, (\tilde{\mathbf{A}}_{\mathcal{J}}^{(n)})^{-1}\mathbf{b} \geq 0,\ (\tilde{\mu}_{\mathcal{J}}^{(n)})^T(\tilde{\mathbf{A}}_{\mathcal{J}}^{(n)})^{-1}\mathbf{b} \geq (\tilde{\mu}_{\mathcal{I}^*}^{(n)})^T(\tilde{\mathbf{A}}_{\mathcal{I}^*}^{(n)})^{-1}\mathbf{b}$.

Set $\mathbf{x}^{(n)} \in \mathbb{R}^{K+L}$ to be the basic feasible solution of $(\tilde{\mathbf{A}}^{(n)}, \mathbf{b})$ corresponding to basis $\mathcal{J}$, i.e. $\mathbf{x}_{\mathcal{J}}^{(n)} = (\tilde{\mathbf{A}}_{\mathcal{J}}^{(n)})^{-1}\mathbf{b}$ and $\mathbf{x}_i^{(n)} = 0$ for $i \notin \mathcal{J}$. Note that $(\mathbf{x}^{(n)})_{n=1}^\infty$ is a uniformly bounded sequence of finite dimensional vectors. By taking subsequences, without lost of generality, assume that $\mathbf{x}^{(n)} \to \mathbf{x}^{(\infty)}$. We have

$$\mathbf{x}^{(\infty)} \geq 0, \quad \mathbf{A}\mathbf{x}^{(\infty)} = \lim_{n\to\infty}(\tilde{\mathbf{A}}_{\mathcal{J}}^{(n)})\mathbf{x}_{\mathcal{J}}^{(n)} = \mathbf{b}, \tag{45}$$

meaning that $\mathbf{x}^{(\infty)}$ is a feasible solution of (SFLP). By taking the limit of the last inequality in (ii) we have

$$\mu^T\mathbf{x}^{(\infty)} = \lim_{n\to\infty}(\tilde{\mu}_{\mathcal{J}}^{(n)})^T(\tilde{\mathbf{A}}_{\mathcal{J}}^{(n)})^{-1}\mathbf{b} \geq \limsup_{n\to\infty}(\tilde{\mu}_{\mathcal{I}^*}^{(n)})^T(\tilde{\mathbf{A}}_{\mathcal{I}^*}^{(n)})^{-1}\mathbf{b}. \tag{46}$$

Under Assumption 1, $\mathbf{A}_{\mathcal{I}^*}$ is invertible and hence the mapping $\tilde{c} \to \tilde{\mathbf{A}}_{\mathcal{I}^*}^{-1}$ is continuous at $\tilde{\mathbf{c}} = \mathbf{c}$. Therefore, through (i) we conclude that $\limsup_{n\to\infty}(\tilde{\mu}_{\mathcal{I}^*}^{(n)})^T(\tilde{\mathbf{A}}_{\mathcal{I}^*}^{(n)})^{-1}\mathbf{b} = \mu_{\mathcal{I}^*}^T \mathbf{A}_{\mathcal{I}^*}^{-1}\mathbf{b}$, i.e. the optimal value of (SFLP). Therefore, (46) means that $\mathbf{x}^{(\infty)}$ is also an optimal solution of (SFLP), which contradicts with the uniqueness assumption. (Let $i \in \mathcal{I}^*\backslash\mathcal{J}$, we have $\mathbf{x}_i^{(\infty)} = 0 \neq \mathbf{x}_i^*$ and hence $\mathbf{x}^{(\infty)} \neq \mathbf{x}^*$.)

F Proof of Θ being dense in $\mathbb{R}^{(L+1)\times K}$

If $\theta' \in \mathbb{R}^{(L+1)\times K}\backslash\Theta$ (i.e. θ' is a feasible instance that violates Assumption 1), then it is necessary that one of the following statements is true: (i) $\det(\mathbf{A}'_{\mathcal{I}}) = 0$ for some basis $\mathcal{I}$; or (ii) $(\mu'_{\mathcal{I}})^T(\mathbf{A}'_{\mathcal{I}})^{-1}\mathbf{b} - (\mu'_{\mathcal{J}})^T(\mathbf{A}'_{\mathcal{J}})^{-1}\mathbf{b} = 0$ for some bases $\mathcal{I} \neq \mathcal{J}$; or (iii) certain coordinate of $(\mathbf{A}'_{\mathcal{I}})^{-1}\mathbf{b}$ is zero for some basis $\mathcal{I}$. In either case, we have $h(\theta') = 0$ for some non-zero polynomial function $h : \mathbb{R}^{(L+1)\times K} \mapsto \mathbb{R}$. Since the set of zeroes of any non-zero polynomial function cannot contain any open ball, we conclude that $\mathbb{R}^{(L+1)\times K}\backslash\Theta$ does not contain any open set, i.e. Θ is dense in $\mathbb{R}^{(L+1)\times K}$.

G Comparing `SFSR` and `SFSR-L` Algorithms

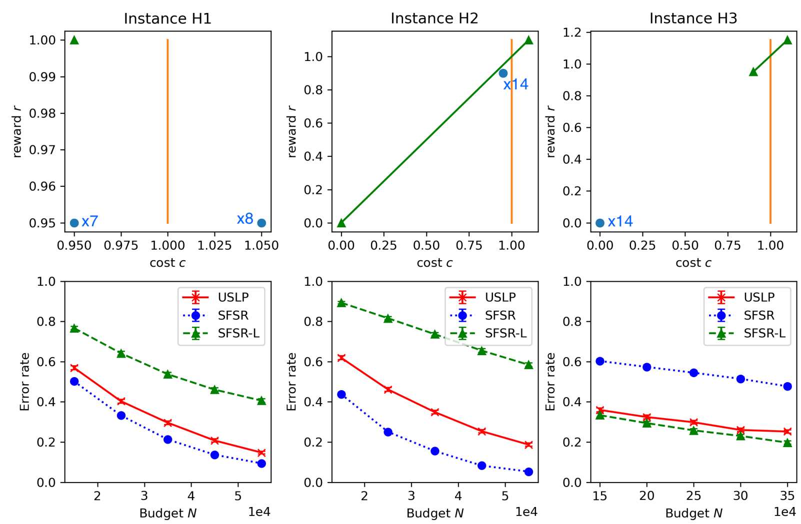

Figure 3: Top: Three hard 16-arm instances. Arms in the optimal support are labeled with green triangles. Bottom: Empirical results for three algorithms under varying budgets. 95% confidence intervals are indicated and tight.

While both flavors of `SFSR` can achieve good performance on an average instance, in Figure 3, we see that in certain carefully constructed hard instances ($H1$-$H3$) while one of the two flavors, `SFSR` or `SFSR-L` performs well, the other does not (the guarantee in any case is probabilistic). In fact, the $H3$ instance in Figure 3 shows that it is hard enough that `SFSR-L` struggles to perform much better than the `USLP` algorithm. Below, we provide a detailed explanation on why `SFSR-L` fails on instance $H2$.

Consider a CBMAI instance with K arms and one type of cost. The mean reward and cost are shown as in the left of Figure 4: Arm 1 has low reward and low cost, arm 2 has high reward and near feasible cost, and arm 3 to K all have the same mean reward and cost: The cost is feasible but close to cost bound $\bar{c}$, and the reward is chosen such that the best mixed arm is formed by a mixture of arm 1 and 2. The (negative) Lagrangian reward $f_a^{\mathrm{L}}(\hat{\mathbf{r}}, \hat{\mathbf{c}})$ of an arm $a \in [K]$ can be visualized in Figure 4 as the vertical distance between arm a to the "frontier" (i.e. the extended line formed by cost-reward vectors of two arms in the empirical optimal support).

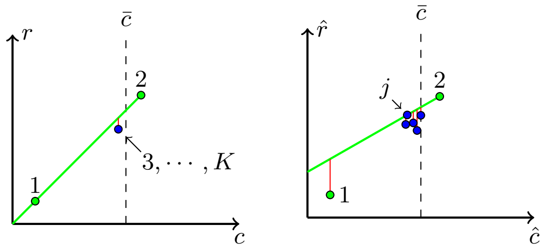

Figure 4: Illustration of `SFSR-L` in a 1-constraint instance. Left: True mean reward and cost. Right: Empirical means after the first episode. Despite that the empirical means do not deviate from the the true mean by too much, arm 1 (a member of the optimal support) ends up having the lowest empirical Lagrangian reward, and is eliminated as a result.

Now, consider the end of episode 1 of `SFSR-L`, and the empirical means of rewards and costs are shown as on the right of Figure 4. Now, arm 2 and some arm $3 \leq j \leq K$ forms the empirical frontier. The empirical dual optimal solution (symbolized by the slope of the frontier) is very different from the true dual optimizer. More importantly, the empirical Lagrangian reward of arm 1 is now the lowest among all arms, and arm 1 is rejected by the `SFSR-L` algorithm in round 1 as a result. Note the identification error happens despite the fact that the arm 1 did not underperform (i.e. $\hat{r}_1 < r_1, \hat{c}_1 > c_1$) its mean.

While in elimination style algorithms there's always the possibility of erroneously rejecting optimal arms, we note that the type of event as shown on the right of Figure 4 is not unlikely: We only require *one of* the $K - 2$ arms to slightly outperform its true mean for arm 1 to be eliminated. In comparison, in unconstrained BAI problems, for the optimal arm (arm 1) to be rejected in episode 1 in an elimination-style algorithm (Audibert & Bubeck, 2010; Karnin et al., 2013), it requires *all of* the other arms (including the worst arm) to empirically outperform arm 1.

H Additional Empirical Results

In addition to the experiments in Section 6, we also applied the three algorithms (`SFSR`, `SFSR-L`, and `USLP`) to 6 instances with $L = 2$ constraints. We set $K = K_0 = 24, \sigma_r = 1, \sigma_c = 0.5$ and $\bar{c}_1 = \bar{c}_2 = 1.0$. In all of the three instances, we set the costs of the 24 arms to be the 24 combinations of $c_{1,i} \in \{0.4, 0.6, 0.8, 1.0, 1.2, 1.4\}$ and $c_{2,i} \in \{0.7, 0.9, 1.1, 1.3\}$. Then, to define the rewards for each instance, we first pick a noise vector $(W_i)_{i=1}^{24}$ (which we will describe later) independently for each instance. In instance $D1$, we set $r_i = 1.0 - W_i$. In instance $D2$, we set $r_i = c_{1,i} - W_i$. In instance $D3$, we set $r_i = c_{1,i} + c_{2,i} - W_i$. We run the randomizations for a few times until the optimal support of each instance Dj has exactly j arms. Finally, we increment the reward of each arm in the optimal support by 0.02 to ensure that the optimal support is unique and the instance is not overly difficult for any CBMAI algorithm.

We consider two ways of choosing the random vector $(W_i)_{i=1}^{24}$: (i) a random permutation of $\{0.0, 0.02, \cdots, 0.46\}$. (ii) i.i.d. uniform random choices from $\{0.0, 0.02, \cdots, 0.28\}$. For the former choice, we will refer to the instance as DjP. For the latter, we will use DjI. The specific instances we used are reported in Table 1.

For each combination of instance-algorithm-budget, we run the simulation for 5000 times independently and obtain the error rate as the proportion of times the algorithm output the wrong support. The results are provided in Figure 5. Each figure takes about 2 hours on an Apple M1 MacBook Air.

We can see that the `SFSR-L` algorithm on these two constraints instances either has the same performance as the `SFSR` algorithm or does a bit better in terms of having a lower error rate.

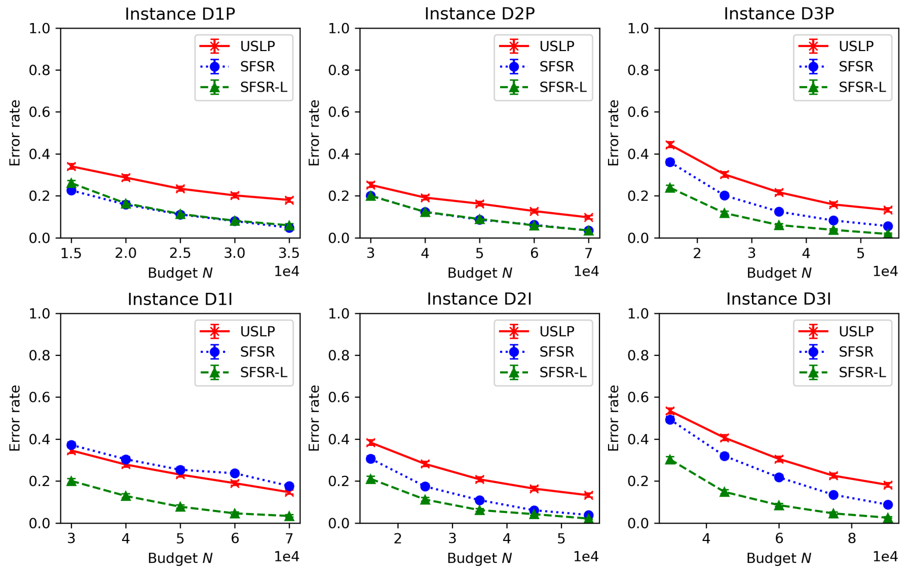

Figure 5: Simulation results for 6 instances with $L = 2$ under varying budget. 95% confidence intervals are indicated and tight.

Table 1: Description of the mean rewards and costs of instances. The rewards of arms in the optimal support is shown in bold. Top Row (from left to right): $D1P, D2P, D3P$. Bottom Row (from left to right): $D1I, D2I, D3I$.

$c_1 \backslash c_2$	0.7	0.9	1.1	1.3
0.4	0.88	0.80	0.82	0.66
0.6	0.72	**1.02**	0.70	0.54
0.8	0.92	0.74	0.94	0.84
1.0	0.76	0.60	0.56	0.86
1.2	0.98	0.64	0.68	0.78
1.4	0.62	0.96	0.90	0.58

$c_1 \backslash c_2$	0.7	0.9	1.1	1.3
0.4	0.08	0.28	-0.02	0.22
0.6	0.20	0.46	0.54	0.40
0.8	0.42	0.52	**0.80**	0.34
1.0	0.92	0.78	0.96	0.70
1.2	0.94	0.76	1.10	0.86
1.4	**1.42**	1.16	1.04	1.24

$c_1 \backslash c_2$	0.7	0.9	1.1	1.3
0.4	1.04	1.22	1.28	1.26
0.6	0.98	1.22	1.60	1.54
0.8	1.26	1.40	**1.88**	1.84
1.0	**1.72**	1.76	1.64	1.92
1.2	1.78	1.70	1.96	2.08
1.4	1.94	**2.30**	2.32	2.50

$c_1 \backslash c_2$	0.7	0.9	1.1	1.3
0.4	0.84	**1.02**	0.74	0.76
0.6	0.84	0.88	0.96	0.90
0.8	0.90	0.98	0.92	0.80
1.0	0.98	0.98	0.90	0.74
1.2	0.94	0.90	0.88	0.72
1.4	0.78	0.82	0.88	0.84

$c_1 \backslash c_2$	0.7	0.9	1.1	1.3
0.4	**0.40**	0.18	0.28	0.14
0.6	0.44	0.54	0.40	0.32
0.8	0.56	0.52	0.68	0.64
1.0	0.84	0.80	0.82	0.74
1.2	0.94	1.18	1.02	1.12
1.4	**1.42**	1.16	1.24	1.24

$c_1 \backslash c_2$	0.7	0.9	1.1	1.3
0.4	0.92	1.12	1.32	1.42
0.6	1.14	1.42	1.62	1.68
0.8	1.30	**1.70**	1.68	**2.06**
1.0	1.46	1.82	2.02	2.04
1.2	1.84	2.02	2.12	2.24
1.4	2.06	**2.28**	2.32	2.58