Combinatorial Bandits Revisited

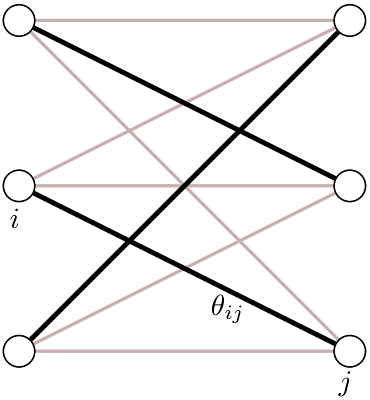
Richard Combes [1], Sadegh Talebi [2], Alexandre Proutière [2] and Marc Lelarge [3]

[1]Centrale-Supélec / L2S, France
[2]KTH, Royal Institute of Technology, Sweden
[3]INRIA, France.

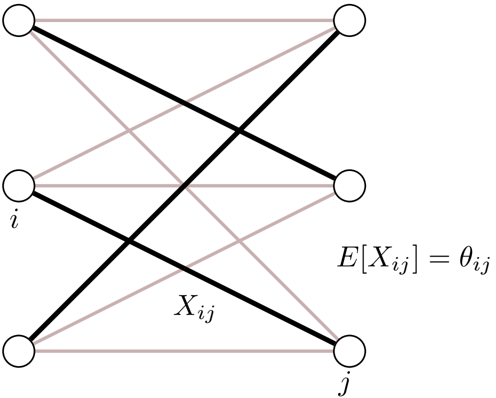
Nov. 2015

Combinatorial optimization

- Decisions: $\mathcal{M} \subset \{0,1\}^d$, weights $\theta \in [0,1]^d$
- Maximize $\mu(M) = M^T\theta = \sum_{i=1}^d M_i\theta_i$ over $\mathcal{M}$



Combinatorial optimization ... in the bandit setting



Bandit vs Semi-bandit

Choose $M(n)$ based on previous samples:

$$X(n) = (0, 1, 1, 0, 1, 0)$$

$$M(n) = (1, 1, 0, 0, 0, 1) \in \mathcal{M}$$

Bandit vs Semi-bandit

Choose $M(n)$ based on previous samples:

$$X(n) = (0, 1, 1, 0, 1, 0)$$

$$M(n) = (1, 1, 0, 0, 0, 1) \in \mathcal{M}$$

Bandit	Semi-Bandit
$M(n)^T X(n) = 2$	$X(n) = (0, 1, \times, \times, \times, 0)$

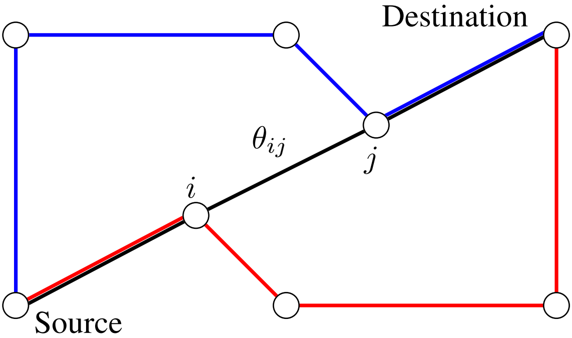
Regret

- Minimize regret:

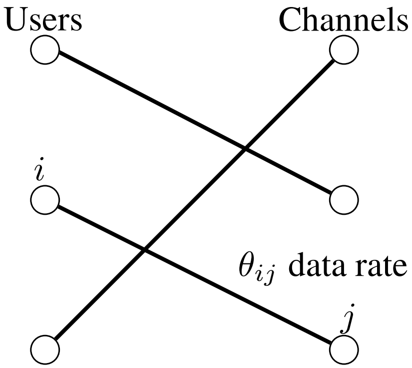
$$R^{\pi}(T) = \underbrace{\max_{M \in \mathcal{M}} \mathbb{E}\left[\sum_{n=1}^{T} M^T X(n)\right]}_{\text{oracle}} - \underbrace{\mathbb{E}\left[\sum_{n=1}^{T} M(n)^T X(n)\right]}_{\text{your algorithm}}.$$

- Stochastic: $X(n)$ i.i.d. , $\mathbb{E}[X(n)] = \theta$, $X_1(n), ..., X_d(n)$ independent.
- Adversarial: $X(n)$ chosen beforehand by an adversary.

Application 1: Shortest path routing



Application 2: Channel Allocation



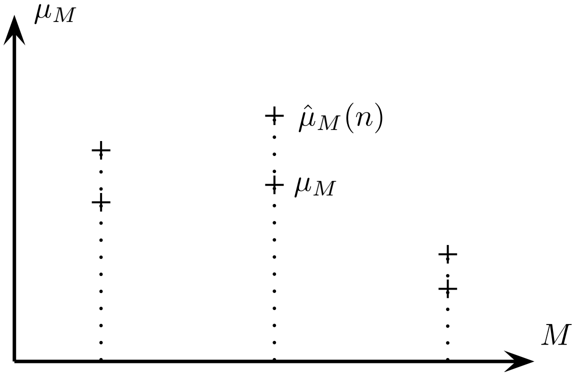
Stochastic case: prior work

How do we quantify problem size ?

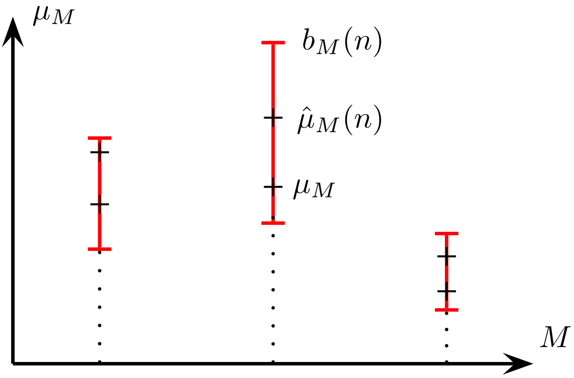
- Dimension $d = \dim \mathcal{M}$,
- Decision size $m = \max_{M \in \mathcal{M}} \sum_{i=1}^{d} M_i$,
- Optimality gap $\Delta = \min_{M \neq M^*} (\mu^* - \mu(M))$
- Regret upper bounds:

LLR (Gai, 2012)	CUCB (Chen, 2013)	CUCB (Kveton, 2014)	ESCB (Present work)
$\mathcal{O}\left(\frac{m^3 d \Delta_{\max}}{\Delta_{\min}^2} \log(T)\right)$	$\mathcal{O}\left(\frac{m^2 d}{\Delta_{\min}} \log(T)\right)$	$\mathcal{O}\left(\frac{md}{\Delta_{\min}} \log(T)\right)$	$\mathcal{O}\left(\frac{\sqrt{m}d}{\Delta_{\min}} \log(T)\right)$

Optimism in the face of uncertainty



Optimism in the face of uncertainty

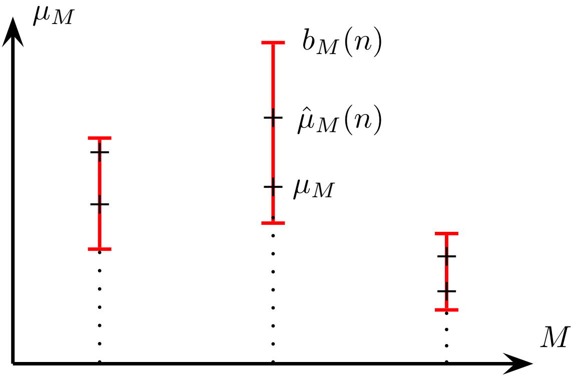


Choose $M(n) \in \arg\max b_M(n)$

Analysis idea:

$$\mathbb{E}[t_M(T)] \leq \underbrace{\sum_{n=1}^{T} \mathbb{P}[b_{M^\star}(n) \leq \mu^*]}_{o(\log(T))} + \underbrace{\sum_{n=1}^{T} \mathbb{P}[M(n) = M, b_M(n) \geq \mu^\star]}_{\text{dominant term}}.$$

Optimism in the face of uncertainty



Choose $M(n) \in \arg\max b_M(n)$

Analysis idea:

$$\mathbb{E}[t_M(T)] \leq \underbrace{\sum_{n=1}^{T} \mathbb{P}[b_{M^\star}(n) \leq \mu^*]}_{o(\log(T))} + \underbrace{\sum_{n=1}^{T} \mathbb{P}[M(n) = M, b_M(n) \geq \mu^\star]}_{\text{dominant term}} .$$

Index construction

- Empirical mean $\hat{\theta}_i(n)$, number of observations: $t_i(n)$.
- Naive approach: consider each component of θ separately
- Hoeffding's inequality, w.p. $1 - O(T^{-1})$:

$$\left|\hat{\theta}_i(n) - \theta_i\right| \leq \sqrt{\frac{2\log(T)}{t_i(n)}}$$

- Index:

$$b_M(n) = \hat{\mu}_M(n) + \sum_{i=1}^{d} M_i \sqrt{\frac{2\log(T)}{t_i(n)}}.$$

- Too conservative: we ignored independence !
- Intuition: $\mathbf{var}(\hat{\mu}_M(n)) = O(\sum_{i=1}^{d} \frac{M_i}{t_i(n)})$.

Index construction

- Empirical mean $\hat{\theta}_i(n)$, number of observations: $t_i(n)$.
- Naive approach: consider each component of θ separately
- Hoeffding's inequality, w.p. $1 - O(T^{-1})$:

$$\left|\hat{\theta}_i(n) - \theta_i\right| \leq \sqrt{\frac{2\log(T)}{t_i(n)}}$$

- Index:

$$b_M(n) = \hat{\mu}_M(n) + \sum_{i=1}^{d} M_i \sqrt{\frac{2\log(T)}{t_i(n)}}.$$

- Too conservative: we ignored independence !
- Intuition: **var**$(\hat{\mu}_M(n)) = O(\sum_{i=1}^{d} \frac{M_i}{t_i(n)})$.

Index construction

- Empirical mean $\hat{\theta}_i(n)$, number of observations: $t_i(n)$.
- Naive approach: consider each component of θ separately
- Hoeffding's inequality, w.p. $1 - O(T^{-1})$:

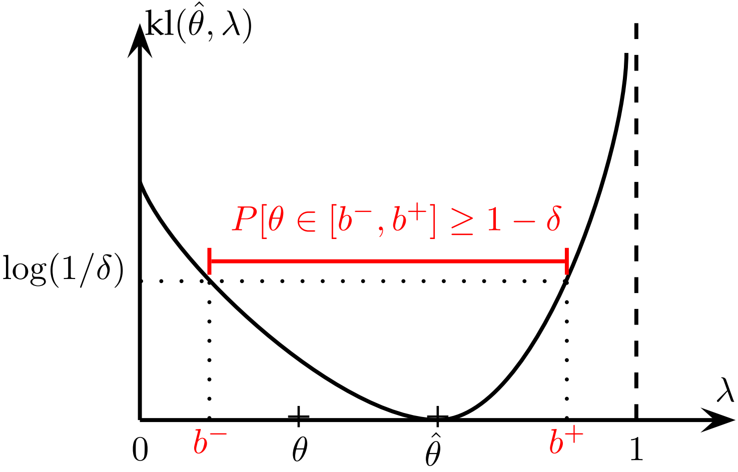
$$\left|\hat{\theta}_i(n) - \theta_i\right| \leq \sqrt{\frac{2\log(T)}{t_i(n)}}$$

- Index:

$$b_M(n) = \hat{\mu}_M(n) + \sum_{i=1}^{d} M_i \sqrt{\frac{2\log(T)}{t_i(n)}}.$$

- Too conservative: <u>we ignored independence</u> !
- Intuition: $\textbf{var}(\hat{\mu}_M(n)) = O(\sum_{i=1}^{d} \frac{M_i}{t_i(n)})$.

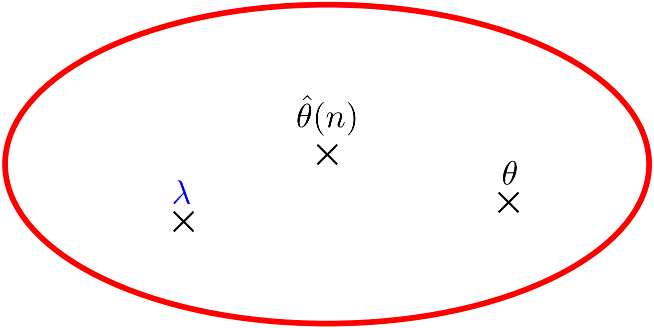
Confidence intervals: one dimension



Confidence intervals: multiple dimensions

Idea: concentration of empirical KL-divergence.

$$P[M^T\theta \leq \max_{\lambda \in B} M^T\lambda] \geq 1 - \delta$$



$$B = \{\lambda : \sum_i t_i(n) kl(\hat{\theta}_i, \lambda_i) \leq \log(1/\delta)\}$$

Proposed algorithm: ESCB

$$b_M(n) = \max_{\lambda \in [0,1]^d} \sum_{i=1}^{d} M_i \lambda_i$$

$$\text{s.t.} \sum_{i=1}^{d} M_i t_i(n) kl(\hat{\theta}_i(n), \lambda_i) \leq \log(T).$$

- b_M computed by a line search (KKT conditions)
- UCB-like index:

$$c_M(n) = \hat{\mu}_M(n) + \sqrt{\sum_{i=1}^{d} M_i \frac{\log(T)}{2t_i(n)}}.$$

Proposed algorithm: ESCB

$$b_M(n) = \max_{\lambda \in [0,1]^d} \sum_{i=1}^{d} M_i \lambda_i$$

$$\text{s.t.} \sum_{i=1}^{d} M_i t_i(n) kl(\hat{\theta}_i(n), \lambda_i) \leq \log(T).$$

- b_M computed by a line search (KKT conditions)
- UCB-like index:

$$c_M(n) = \hat{\mu}_M(n) + \sqrt{\sum_{i=1}^{d} M_i \frac{\log(T)}{2t_i(n)}}.$$

Regret analysis

Theorem

The regret under ESCB satisfies $R(T) = \mathcal{O}\left(\frac{d\sqrt{m}\log(T)}{\Delta}\right)$

- Idea: $c_M(n) \geq b_M(n) \geq M^T\theta$ with high probability
- Crucial concentration inequality:

Lemma (Combes et al., COLT 2014)

$$\mathbb{P}\left[\max_{n\leq T}\sum_{i=1}^{d} M_i t_i(n) kl(\hat{\theta}_i(n), \theta_i) \geq \delta\right] \leq C_m(\log(T)\delta)^m e^{-\delta}$$

- Proof: multi-dimensional peeling.

Regret Lower Bound

How far are we from the optimal algorithm ?

- Uniformly good algorithm: $R^\pi(T) = O(\log(T))$ for all θ.

Theorem

For any uniformly good algorithm, $\liminf_{T\to\infty} \frac{R^\pi(T)}{\log(T)} \geq c(\theta)$*, with* $c(\theta)$ *solution to:*

$$\inf_{c\in(\mathbb{R}^+)^{|\mathcal{M}|}} \sum_{M\in\mathcal{M}} c_M(\mu^* - \mu(M))$$

$$s.t. \sum_{i=1}^{d} kl(\theta_i, \lambda_i) \sum_{M\in\mathcal{M}} c_M M_i \geq 1 \;, \; \forall \lambda \in B(\theta).$$

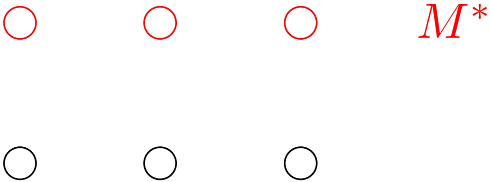
- Proof idea: Graves and Lai's LP.
- Decision M must be sampled at least $c_M \log(T)$ times.

Towards an explicit lower bound

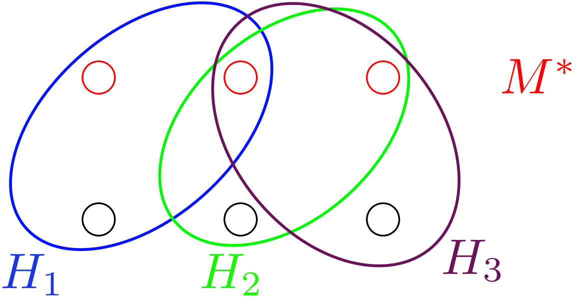
How does $c(\theta)$ scale with d,m and Δ ?

- The LP is not explicit, we must work harder ...
- Idea: consider a covering of the suboptimal items

Towards an explicit lower bound



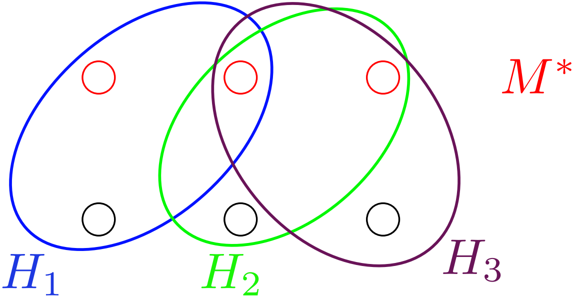
Towards an explicit lower bound



$$\mathcal{H} = \{H_1, H_2, H_3\}$$

- Proposition: $c(\theta) = \Omega(|\mathcal{H}|/\Delta)$
- For most problems $|\mathcal{H}| = \Omega(d - m)$.

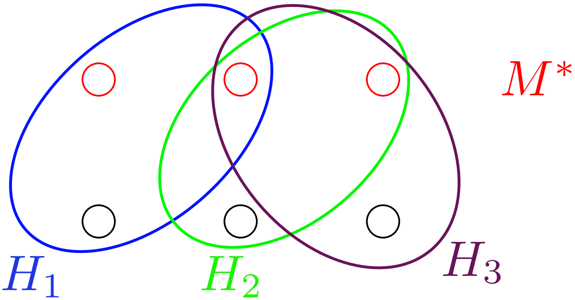
Towards an explicit lower bound



$$\mathcal{H} = \{H_1, H_2, H_3\}$$

- Proposition: $c(\theta) = \Omega(|\mathcal{H}|/\Delta)$
- For most problems $|\mathcal{H}| = \Omega(d - m)$.

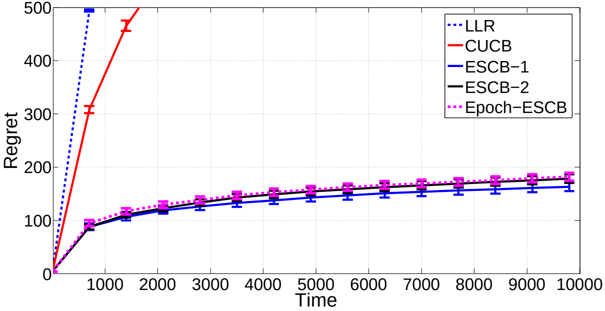
Towards an explicit lower bound

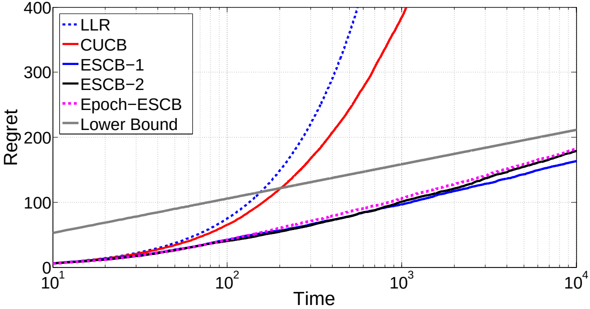


$$\mathcal{H} = \{H_1, H_2, H_3\}$$

- Proposition: $c(\theta) = \Omega(|\mathcal{H}|/\Delta)$
- For most problems $|\mathcal{H}| = \Omega(d - m)$.

Some numerical experiments, matchings





Conclusion

- Efficient algorithms for online combinatorial optimization
- Much more in the paper: adversarial + bandit feedback
- `http://arxiv.org/abs/1502.03475`
- A (personal) conjecture: ESCB is asymptotically optimal

Thank you for your attention !