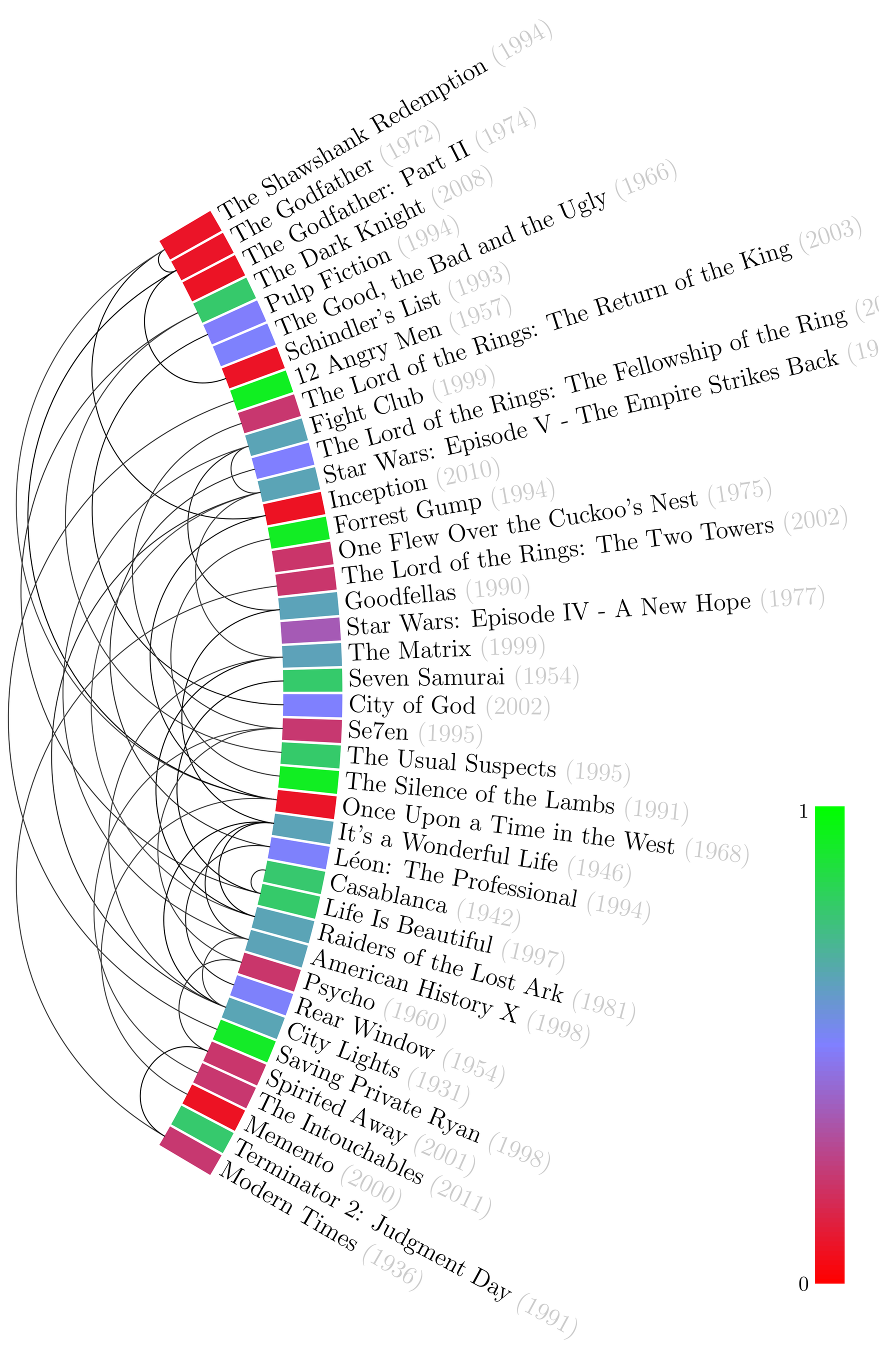
Spectral Bandits for Smooth Graph Functions

Tomas.Kocak@inria.fr, Michal.Valko@inria.fr, Remi.Munos@inria.fr, and Branislav.Kveton@technicolor.com

Motivation - Movie recommendation

- **Goal:** Movie recommendation based on similarities
- **Challenges:** Good prediction after just a few steps ($T \ll N$)
- **Prior knowledge:** The preferences of movies are smooth over a given weighted similarity graph



- Colors represent *single*-user preferences.
- Connected (similar) movies have similar user ratings

Smooth graph functions

- **Graph function:** mapping from set of the graph vertices $V(G)$ into real numbers
- **Smoothness of a graph function** $S_G(f)$:
  - eigendecomposition of graph Laplacian: $\mathcal{L} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\intercal$

$$S_G(f) = \frac{1}{2}\sum_{u,v\in V(G)} w_{u,v}(f(u)-f(v))^2 = \mathbf{f}^\intercal\mathcal{L}\mathbf{f}$$

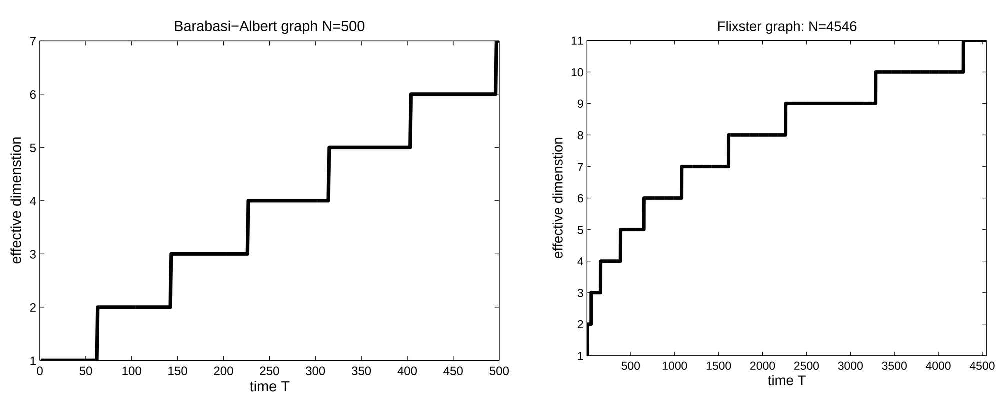
$$= \mathbf{f}^\intercal\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\intercal\mathbf{f} = \boldsymbol{\alpha}^{*\intercal}\mathbf{\Lambda}\boldsymbol{\alpha}^* = \|\boldsymbol{\alpha}^*\|_{\mathbf{\Lambda}} = \sum_{i=1}^N \lambda_i\alpha_i^2$$

- **Observation:** $S_G(\mathbf{q}_i) = \lambda_i$
- **Smoothness and regularization:** Small value of **(a)** $S_G(f)$ **(b)** $\mathbf{\Lambda}$ norm of $\boldsymbol{\alpha}^*$ **(c)** α_i for large λ_i

Effective Dimension

Definition 1. *Let the* ***effective dimension*** d *be the largest* d *such that*

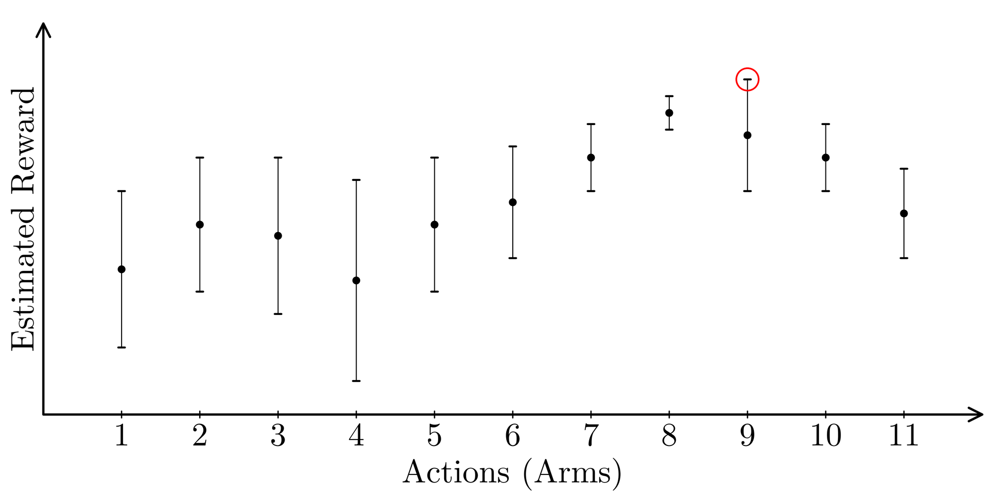
$$(d-1)\lambda_d \le \frac{T}{\log(1+T/\lambda)}.$$



- d is small when the coefficients λ_i grow rapidly above time.
- d is related to the number of "non-negligible" dimensions

Upper Confidence Bound algorithms

- Pick an arm with the highest upper confidence bound.
- Update estimates and confidence intervals.



Setting

- **Task:** Each time t, pick an action (node) to get a reward.
- **Reward:** $\mathbf{x}_i^\intercal\boldsymbol{\alpha}^* + \varepsilon_t$ (with unknown parameter $\boldsymbol{\alpha}^*$)
  - $\mathbf{x}_i$ is the i-th row of $\mathbf{Q}$
  - reward is a combination of smooth eigenvectors
- **Goal:** Minimize the cumulative regret w.r.t. the best node

$$R_T = T\max_v f_{\boldsymbol{\alpha}^*}(v) - \sum_{t=1}^T f_{\boldsymbol{\alpha}^*}(\pi(t))$$

Algorithm 1 - SpectralUCB

Input:
N : the number of nodes, T : the number of pulls
$\{\mathbf{\Lambda}_\mathcal{L}, \mathbf{Q}\}$ spectral basis of $\mathcal{L}$
λ, δ : regularization and confidence parameters
R, C : upper bounds on the noise and $\|\boldsymbol{\alpha}^*\|_{\mathbf{\Lambda}}$
Run:
$\mathbf{\Lambda} \leftarrow \mathbf{\Lambda}_\mathcal{L} + \lambda\mathbf{I}$
$d \leftarrow \max\{d : (d-1)\lambda_d \le T/\log(1+T/\lambda)\}$
for $t = 1$ **to** T **do**
Update the basis coefficients $\hat{\boldsymbol{\alpha}}$:
$\mathbf{X}_t \leftarrow [\mathbf{x}_1, \dots, \mathbf{x}_{t-1}]^\intercal$
$r \leftarrow [r_1, \dots, r_{t-1}]^\intercal$
$\mathbf{V}_t \leftarrow \mathbf{X}_t\mathbf{X}_t^\intercal + \mathbf{\Lambda}$
$\hat{\boldsymbol{\alpha}}_t \leftarrow \mathbf{V}_t^{-1}\mathbf{X}_t^\intercal r$
$c_t \leftarrow 2R\sqrt{d\log(1+t/\lambda) + 2\log(1/\delta)} + C$
Choose the node v_t ($\mathbf{x}_{v_t}$-th row of $\mathbf{Q}$):
$v_t \leftarrow \arg\max_v \left(f_{\hat{\boldsymbol{\alpha}}}(v) + c_t\|\mathbf{x}_v\|_{\mathbf{V}_t^{-1}}\right)$
Observe the reward r_t
end for

Algorithm 2 - SpectralEliminator

Input:
N : the number of nodes, T : the number of pulls
$\{\mathbf{\Lambda}_\mathcal{L}, \mathbf{Q}\}$ spectral basis of $\mathcal{L}$
λ : regularization parameter
$\beta, \{t_j\}_j^J$ parameters of the elimination and phases
$A_1 \leftarrow \{\mathbf{x}_1, \dots, \mathbf{x}_K\}$.
for $j = 1$ **to** J **do**
$\mathbf{V}_{t_j} \leftarrow \gamma\mathbf{\Lambda}_\mathcal{L} + \lambda\mathbf{I}$
for $t = t_j$ **to** $\min(t_{j+1}-1, T)$ **do**
Play $\mathbf{x}_t \in A_j$ with the largest width to observe r_t:
$\mathbf{x}_t \leftarrow \arg\max_{\mathbf{x}\in A_j}\|\mathbf{x}\|_{\mathbf{V}_t^{-1}}$
$\mathbf{V}_{t+1} \leftarrow \mathbf{V}_t + \mathbf{x}_t\mathbf{x}_t^\intercal$
end for
Eliminate the arms that are not promising:
$\hat{\boldsymbol{\alpha}}_t \leftarrow \mathbf{V}_t^{-1}[\mathbf{x}_{t_j}, \dots, \mathbf{x}_t][r_{t_j}, \dots, r_t]^\intercal$
$p \leftarrow \max_{\mathbf{x}\in A_j}\left[\langle\hat{\boldsymbol{\alpha}}_t, \mathbf{x}\rangle - \|\mathbf{x}\|_{\mathbf{V}_t^{-1}}\beta\right]$
$A_{j+1} \leftarrow \left\{\mathbf{x}\in A_j, \langle\hat{\boldsymbol{\alpha}}_t, \mathbf{x}\rangle + \|\mathbf{x}\|_{\mathbf{V}_t^{-1}}\beta \ge p\right\}$
end for

Main results

SpectralUCB regret bound

Theorem 1. *Let be the minimum eigenvalue of* $\mathbf{\Lambda}$. *If* $\|\boldsymbol{\alpha}^*\|_{\mathbf{\Lambda}} \le C$ *and for all* $\mathbf{x}_v$, $\langle\mathbf{x}_v, \boldsymbol{\alpha}^*\rangle \in [-1,1]$, *then the cumulative regret of SpectralUCB is with probability at least* $1-\delta$ *bounded as*

$$R_T \le \left[8R\sqrt{d\log(1+T/\lambda) + 2\log(1/\delta)} + 4C + 4\right] \times \sqrt{dT\log(1+T/\lambda)} \approx d\sqrt{T}$$

Setting $\mathbf{\Lambda} = \mathbf{I}$ we recover LinUCB. Since $\log(|\mathbf{V}_T|/|\mathbf{\Lambda}|)$ can be upperbounded by $D\log T$ [1], we obtain $\tilde{\mathcal{O}}(D\sqrt{T})$ for LinUCB.

SpectralEliminator regret bound

Theorem 2. *Choose the phases starts as* $t_j = 2^{j-1}$. *Assume all rewards are in* $[0,1]$ *and* $\|\boldsymbol{\alpha}^*\|_{\mathbf{\Lambda}} \le C$. *For any* $\delta > 0$, *with probability at least* $1-\delta$, *the cumulative regret of SpectralEliminator algorithm run with parameter* $\beta = 2R\sqrt{14\log(2K\log_2 T/\delta)} + C$ *is bounded as:*

$$R_T \le \frac{4}{\log 2}\left(2R\sqrt{14\log\frac{2K\log_2 T}{\delta}} + C\right)\sqrt{dT\log\left(1+\frac{T}{\lambda}\right)} \approx \sqrt{dT}$$

If $\mathbf{\Lambda} = \mathbf{I}$, we get a competitor to SupLinRel [2], with $\tilde{\mathcal{O}}(\sqrt{DT})$ regret, with significantly more elegant algorithm and analysis.

Linear vs. Spectral bandits

Linear	*Spectral*
LinUCB $D\sqrt{T}\ln T$	**SpectralUCB** $d\sqrt{T}\ln T$
SupLinRel $\sqrt{DT}\ln T$	***SpectralEliminator*** $\sqrt{dT}\ln T$

Analyses sketch

- Derivation of the confidence ellipsoid for estimate $\hat{\boldsymbol{\alpha}}$.

By self-normalized bound of [1]: w. p. $1-\delta$:

$$|x^\intercal(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)| \le \|x\|_{\mathbf{V}_t^{-1}}\left(R\sqrt{2\log\left(\frac{|\mathbf{V}_t|^{1/2}}{\delta|\mathbf{\Lambda}|^{1/2}}\right)} + C\right)$$

- Our key result coming from spectral properties of $\mathbf{V}_t$:

$$\log\frac{|\mathbf{V}_t|}{|\mathbf{\Lambda}|} \le 2d\log\left(1+\frac{T}{\lambda}\right)$$

SpectralUCB

- Regret in one time step: $\mathbf{x}_*^\intercal\boldsymbol{\alpha}^* - \mathbf{x}_{\pi(t)}^\intercal\boldsymbol{\alpha}^* \le 2c_t\|\mathbf{x}_{\pi(t)}\|_{\mathbf{V}_t^{-1}}$

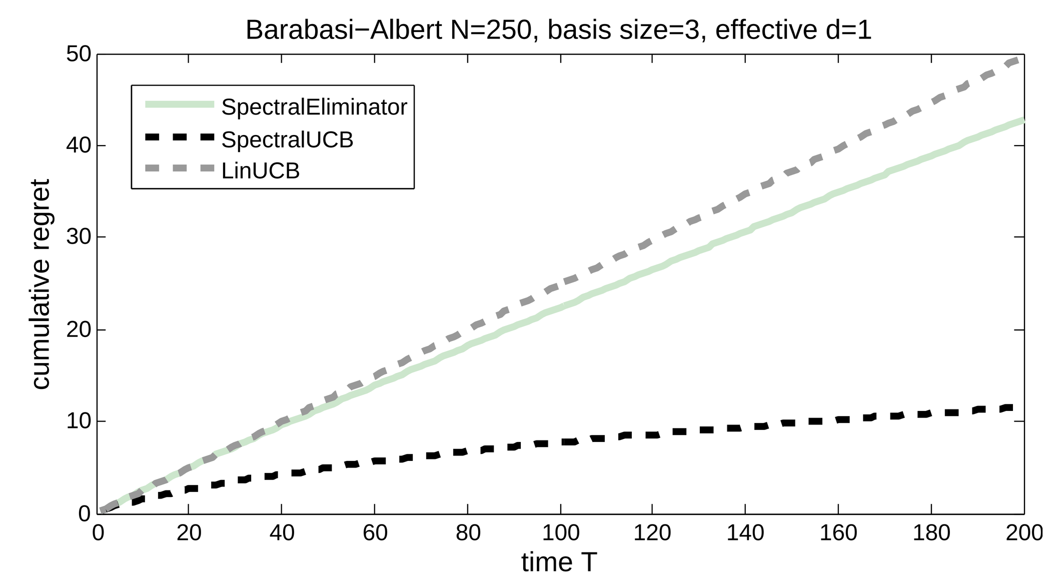
SpectralEliminator

- Divide time into sets ($t_1 = 1 \le t_2 \le \dots$) to introduce independence for Azuma-Hoeffding inequality and observe $R_T \le \sum_{j=0}^J(t_{j+1}-t_j)\left[\langle\mathbf{x}^*-\mathbf{x}_t, \hat{\boldsymbol{\alpha}}_j\rangle + (\|\mathbf{x}^*\|_{\mathbf{V}_j^{-1}} + \|\mathbf{x}_t\|_{\mathbf{V}_j^{-1}})\beta\right]$
- Bound $\langle\mathbf{x}^*-\mathbf{x}_t, \hat{\boldsymbol{\alpha}}_j\rangle$ for each phase
- No bad arms: $\langle\mathbf{x}^*-\mathbf{x}_t, \hat{\boldsymbol{\alpha}}_j\rangle \le (\|\mathbf{x}^*\|_{\mathbf{V}_j^{-1}} + \|\mathbf{x}_t\|_{\mathbf{V}_j^{-1}})\beta$
- By algorithm: $\|\mathbf{x}\|^2_{\mathbf{V}_j^{-1}} \le \frac{1}{t_j - t_{j-1}}\sum_{s=t_{j-1}+1}^{t_j}\|\mathbf{x}_s\|^2_{\mathbf{V}_{s-1}^{-1}}$
- $\sum_{s=t_{j-1}+1}^{t_j}\min\left(1, \|\mathbf{x}_s\|^2_{\mathbf{V}_{s-1}^{-1}}\right) \le \log\frac{|\mathbf{V}_j|}{|\mathbf{\Lambda}|}$

Experiments

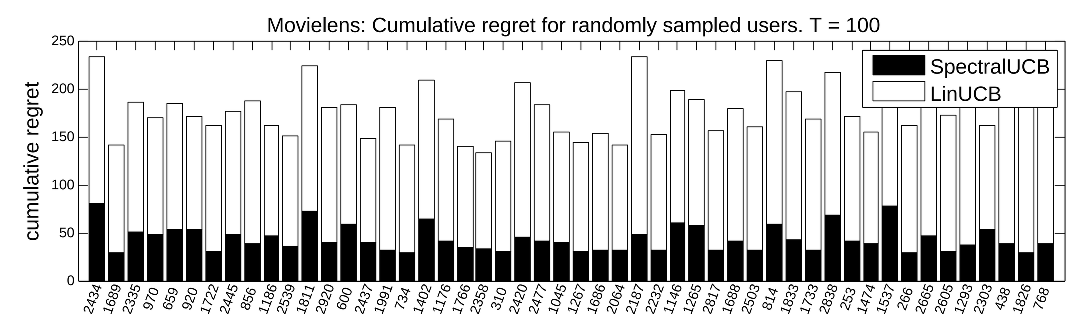
Synthetic Experiment

Barabási-Albert (BA) model with the degree parameter 3.



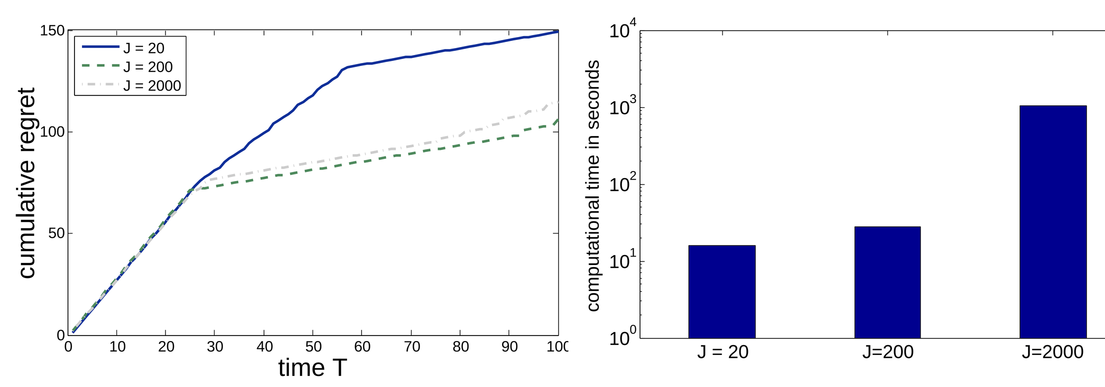
Movie Experiment

MovieLens dataset of 6k users who rated one million movies.



Improving the running time: reduced eigenbasis

- **Reduced basis:** We only need first few eigenvectors.
- **Getting** J **eigenvectors:** $\mathcal{O}(Jm\log m)$ time for m edges
- Computationally less expensive, comparable performance.



References

[1] Yasin Abbasi-Yadkori, David Pal, and Csaba Szepesvari. Improved Algorithms for Linear Stochastic Bandits. In *Neural Information Processing Systems*. 2011.

[2] Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3:397–422, March 2002.

Acknowledgments

intel
CompLACS