Spectral Thompson Sampling

Tomáš Kocák — *Inria Lille - Nord Europe, France*
Michal Valko — *Inria Lille - Nord Europe, France*
Rémi Munos — *Inria Lille - Nord Europe & Microsoft Research NE*
Shipra Agrawal — *Microsoft Research India*

SequeL – INRIA Lille

AAAI 2014

Quebec, July 2014

Movie recommendation: (in each time step)

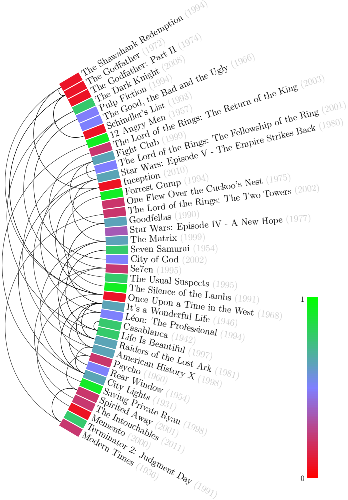
- Recommend movies to a **single user**.
- Good prediction after a few steps ($T \ll N$).

Goal:

- Maximize overall reward (sum of ratings).

Assumptions:

- Unknown reward function $f : V(G) \to \mathbb{R}$.
- Function f is **smooth** on a graph.
- Neighboring movies $\Rightarrow$ similar preferences.
- Similar preferences $\not\Rightarrow$ neighboring movies.



Smooth graph function

- Graph G with vertex set $V(G) = \{1, \ldots, N\}$ and edge set $E(G)$.
- $f_1, \ldots, f_N$: Values of the function on the vertices of the graph.
- $w_{i,j}$: Weight of the edge connecting nodes i and j.
- **Smoothness of the function:**

$$S_G(f) = \frac{1}{2} \sum_{i,j \leq N} w_{i,j}(f_i - f_j)^2$$

  - Smaller value of $S_G(f)$, smoother the function f is.

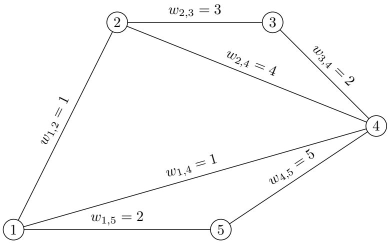
- **Examples:**
  - **Complete graph:** Only constant function has smoothness 0.
  - **Edgeless graph:** Every function has smoothness 0.
  - **Constant function:** Smoothness 0 for every graph.

Graph Laplacian

- $\mathcal{W}$: $N \times N$ matrix of the edge weights $w_{i,j}$.
- $\mathcal{D}$: Diagonal matrix with the entries $d_i = \sum_j w_{i,j}$.
- $\mathcal{L} = \mathcal{D} - \mathcal{W}$: Graph Laplacian.
  - Positive semidefinite matrix.
  - Diagonally dominant matrix.

Example:

$$\mathcal{L} = \begin{pmatrix} 4 & -1 & 0 & -1 & -2 \\ -1 & 8 & -3 & -4 & 0 \\ 0 & -3 & 5 & -2 & 0 \\ -1 & -4 & -2 & 12 & -5 \\ -2 & 0 & 0 & -5 & 7 \end{pmatrix}$$



Smoothness of the function and Laplacian

- $\boldsymbol{f} = (f_1, \ldots, f_N)^\mathsf{T}$: Vector of function values.
- Let $\mathcal{L} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\mathsf{T}$ be the eigendecomposition of the Laplacian.
  - Diagonal matrix $\mathbf{\Lambda}$ whose diagonal entries are eigenvalues of $\mathcal{L}$.
  - Columns of $\mathbf{Q}$ are eigenvectors of $\mathcal{L}$.
  - Columns of $\mathbf{Q}$ form a basis.
- $\boldsymbol{\mu}$: Unique vector such that $\mathbf{Q}\boldsymbol{\mu} = \boldsymbol{f}$ Note: $\mathbf{Q}^\mathsf{T}\boldsymbol{f} = \boldsymbol{\mu}$

$$S_G(f) = \boldsymbol{f}^\mathsf{T}\mathcal{L}\boldsymbol{f} = \boldsymbol{f}^\mathsf{T}\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\mathsf{T}\boldsymbol{f} = \boldsymbol{\mu}^\mathsf{T}\mathbf{\Lambda}\boldsymbol{\mu} = \|\boldsymbol{\mu}\|_{\mathbf{\Lambda}} = \sum_{i=1}^{N} \lambda_i(\mu_i)^2$$

Smoothness and <u>regularization</u>: Small value of

(a) $S_G(f)$ **(b)** $\mathbf{\Lambda}$ norm of $\boldsymbol{\mu}$ **(c)** μ_i for large λ_i

Setting

Problem structure

- Underlying graph structure encoded in the graph laplacian $\mathcal{L}$.
- Eigendecomposition of graph laplacian $\mathcal{L} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\mathsf{T}$ where $\mathbf{Q}$ is the matrix with eigenvectors in columns.
- the i-th row $\mathbf{b}_i$ of the matrix $\mathbf{Q}$ corresponds to the arm i.

Learning setting

- In each time step choose a node $a(t)$.
- Obtain noisy reward $r_t = \mathbf{b}_{a(t)}^\mathsf{T}\boldsymbol{\mu} + \varepsilon_t$.   Note: $\mathbf{b}_{a(t)}^\mathsf{T}\boldsymbol{\mu} = f_{a(t)}$
  - ε_t is R-sub-Gaussian noise. $\forall \xi \in \mathbb{R}, \mathbb{E}[e^{\xi\varepsilon_t}] \leq \exp\left(\xi^2 R^2/2\right)$
- Minimize cumulative regret

$$R_T = T\max_a\left(\mathbf{b}_a^\mathsf{T}\boldsymbol{\mu}\right) - \sum_{t=1}^{T}\mathbf{b}_{a(t)}^\mathsf{T}\boldsymbol{\mu}.$$

Solutions

- **Linear bandit algorithms** (Existing solutions)
  - **LinUCB** (Li et al., 2010)
    - Regret bound $\approx D\sqrt{T \ln T}$
  - **LinearTS** (Agrawal and Goyal, 2013)
    - Regret bound $\approx D\sqrt{T \ln N}$

Note: D is ambient dimension, in our case N, length of $\mathbf{b}_i$.

Solutions

- **Linear bandit algorithms** (Existing solutions)
  - **LinUCB** (Li et al., 2010)
    - Regret bound $\approx D\sqrt{T \ln T}$
  - **LinearTS** (Agrawal and Goyal, 2013)
    - Regret bound $\approx D\sqrt{T \ln N}$

Note: D is ambient dimension, in our case N, length of $\mathbf{b}_i$.

- **Spectral bandit algorithms** (Our solutions)
  - **SpectralUCB** (Valko et al., 2014)
    - Regret bound $\approx d\sqrt{T \ln T}$

Note: d is effective dimension, usually much smaller than D.

Solutions

- **Linear bandit algorithms** (Existing solutions)
  - **LinUCB** (Li et al., 2010)
    - Regret bound $\approx D\sqrt{T \ln T}$
  - **LinearTS** (Agrawal and Goyal, 2013)
    - Regret bound $\approx D\sqrt{T \ln N}$

Note: D is ambient dimension, in our case N, length of $\mathbf{b}_i$.

- **Spectral bandit algorithms** (Our solutions)
  - **SpectralUCB** (Valko et al., 2014)
    - Regret bound $\approx d\sqrt{T \ln T}$
    - Operations per step: $D^2 N$

Note: d is effective dimension, usually much smaller than D.

Solutions

- **Linear bandit algorithms** (Existing solutions)
  - **LinUCB** (Li et al., 2010)
    - Regret bound $\approx D\sqrt{T \ln T}$
  - **LinearTS** (Agrawal and Goyal, 2013)
    - Regret bound $\approx D\sqrt{T \ln N}$

Note: D is ambient dimension, in our case N, length of $\mathbf{b}_i$.

- **Spectral bandit algorithms** (Our solutions)
  - **SpectralUCB** (Valko et al., 2014)
    - Regret bound $\approx d\sqrt{T \ln T}$
    - Operations per step: $D^2 N$
  - **SpectralTS** – New! –
    - Regret bound $\approx d\sqrt{T \ln N}$
    - Operations per step: $D^2 + DN$

Note: d is effective dimension, usually much smaller than D.

Effective dimension

- **Effective dimension:** Largest d such that

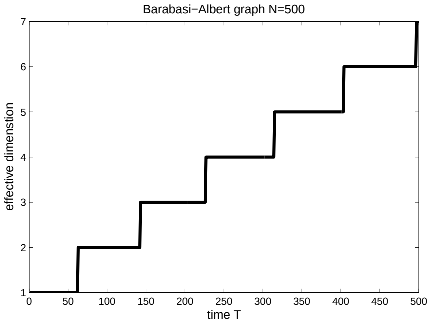
$$(d-1)\lambda_d \leq \frac{T}{\log(1+T/\lambda)}.$$

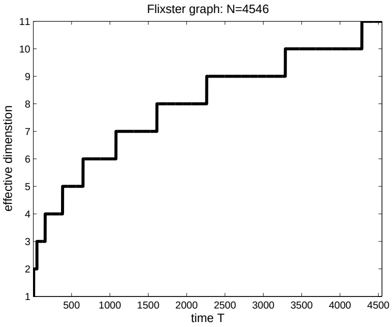
- Function of time horizon and graph properties
- λ_i : i-th smallest eigenvalue of $\mathbf{\Lambda}$.
- λ: Regularization parameter of the algorithm.

Properties:

- d is small when the coefficients λ_i grow rapidly above time.
- d is related to the number of "non-negligible" dimensions.
- Usually d is much smaller than D in real world graphs.
- Can be computed beforehand.

Effective dimension vs. Ambient dimension





$$d \ll D$$

Note: In our setting $T < N = D$.

SpectralTS algorithm

Input:
N: number of arms, T: number of pulls
$\{\mathbf{\Lambda}_{\mathcal{L}}, \mathbf{Q}\}$: spectral basis of graph Laplacian $\mathcal{L}$
λ, δ: regularization and confidence parameters
R, C: upper bounds on noise and $\|\boldsymbol{\mu}\|_{\mathbf{\Lambda}}$
Initialization:
$v = R\sqrt{6d\log((\lambda + T)/\delta\lambda)} + C$
$\hat{\boldsymbol{\mu}} = 0_N$, $\boldsymbol{f} = 0_N$, $\mathbf{B} = \mathbf{\Lambda}_{\mathcal{L}} + \lambda \mathbf{I}_N$
Run:
for $t = 1$ **to** T **do**
Sample $\tilde{\boldsymbol{\mu}} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}, v^2\mathbf{B}^{-1})$
$a(t) \leftarrow \arg\max_a \mathbf{b}_a^{\mathsf{T}}\tilde{\boldsymbol{\mu}}$
Observe a noisy reward $r(t) = \mathbf{b}_{a(t)}^{\mathsf{T}}\boldsymbol{\mu} + \varepsilon_t$
$\boldsymbol{f} \leftarrow \boldsymbol{f} + \mathbf{b}_{a(t)} r(t)$
Update $\mathbf{B} \leftarrow \mathbf{B} + \mathbf{b}_{a(t)}\mathbf{b}_{a(t)}^{\mathsf{T}}$
Update $\hat{\boldsymbol{\mu}} \leftarrow \mathbf{B}^{-1}\boldsymbol{f}$
end for

SpectralTS algorithm

Input:
N: number of arms, T: number of pulls
$\{\mathbf{\Lambda}_{\mathcal{L}}, \mathbf{Q}\}$: spectral basis of graph Laplacian $\mathcal{L}$
λ, δ: regularization and confidence parameters
R, C: upper bounds on noise and $\|\boldsymbol{\mu}\|_{\mathbf{\Lambda}}$
Initialization:
$v = R\sqrt{6d\log((\lambda + T)/\delta\lambda)} + C$
$\hat{\boldsymbol{\mu}} = 0_N$, $\boldsymbol{f} = 0_N$, $\mathbf{B} = \mathbf{\Lambda}_{\mathcal{L}} + \lambda\mathbf{I}_N$
Run:
for $t = 1$ **to** T **do**
Sample $\tilde{\boldsymbol{\mu}} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}, v^2\mathbf{B}^{-1})$
$a(t) \leftarrow \arg\max_a \mathbf{b}_a^{\mathsf{T}}\tilde{\boldsymbol{\mu}}$
Observe a noisy reward $r(t) = \mathbf{b}_{a(t)}^{\mathsf{T}}\boldsymbol{\mu} + \varepsilon_t$
$\boldsymbol{f} \leftarrow \boldsymbol{f} + \mathbf{b}_{a(t)}r(t)$
Update $\mathbf{B} \leftarrow \mathbf{B} + \mathbf{b}_{a(t)}\mathbf{b}_{a(t)}^{\mathsf{T}}$
Update $\hat{\boldsymbol{\mu}} \leftarrow \mathbf{B}^{-1}\boldsymbol{f}$
end for

SpectralTS algorithm

Input:
- N: number of arms, T: number of pulls
- $\{\mathbf{\Lambda}_{\mathcal{L}}, \mathbf{Q}\}$: spectral basis of graph Laplacian $\mathcal{L}$
- λ, δ: regularization and confidence parameters
- R, C: upper bounds on noise and $\|\boldsymbol{\mu}\|_{\mathbf{\Lambda}}$

Initialization:
- $v = R\sqrt{6d\log((\lambda + T)/\delta\lambda)} + C$
- $\hat{\boldsymbol{\mu}} = 0_N$, $\boldsymbol{f} = 0_N$, $\mathbf{B} = \mathbf{\Lambda}_{\mathcal{L}} + \lambda \mathbf{I}_N$

Run:

for $t = 1$ **to** T **do**
- Sample $\tilde{\boldsymbol{\mu}} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}, v^2\mathbf{B}^{-1})$
- $a(t) \leftarrow \arg\max_a \mathbf{b}_a^{\mathsf{T}}\tilde{\boldsymbol{\mu}}$
- Observe a noisy reward $r(t) = \mathbf{b}_{a(t)}^{\mathsf{T}}\boldsymbol{\mu} + \varepsilon_t$
- $\boldsymbol{f} \leftarrow \boldsymbol{f} + \mathbf{b}_{a(t)}r(t)$
- Update $\mathbf{B} \leftarrow \mathbf{B} + \mathbf{b}_{a(t)}\mathbf{b}_{a(t)}^{\mathsf{T}}$
- Update $\hat{\boldsymbol{\mu}} \leftarrow \mathbf{B}^{-1}\boldsymbol{f}$

end for

SpectralTS algorithm

Input:
N: number of arms, T: number of pulls
$\{\boldsymbol{\Lambda}_{\mathcal{L}}, \mathbf{Q}\}$: spectral basis of graph Laplacian $\mathcal{L}$
λ, δ: regularization and confidence parameters
R, C: upper bounds on noise and $\|\boldsymbol{\mu}\|_{\boldsymbol{\Lambda}}$
Initialization:
$v = R\sqrt{6d\log((\lambda + T)/\delta\lambda)} + C$
$\hat{\boldsymbol{\mu}} = 0_N$, $\boldsymbol{f} = 0_N$, $\mathbf{B} = \boldsymbol{\Lambda}_{\mathcal{L}} + \lambda\mathbf{I}_N$
Run:
for $t = 1$ **to** T **do**
Sample $\tilde{\boldsymbol{\mu}} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}, v^2\mathbf{B}^{-1})$
$a(t) \leftarrow \arg\max_a \mathbf{b}_a^{\mathsf{T}}\tilde{\boldsymbol{\mu}}$
Observe a noisy reward $r(t) = \mathbf{b}_{a(t)}^{\mathsf{T}}\boldsymbol{\mu} + \varepsilon_t$
$\boldsymbol{f} \leftarrow \boldsymbol{f} + \mathbf{b}_{a(t)} r(t)$
Update $\mathbf{B} \leftarrow \mathbf{B} + \mathbf{b}_{a(t)}\mathbf{b}_{a(t)}^{\mathsf{T}}$
Update $\hat{\boldsymbol{\mu}} \leftarrow \mathbf{B}^{-1}\boldsymbol{f}$
end for

SpectralTS algorithm

Input:
N: number of arms, T: number of pulls
$\{\mathbf{\Lambda}_{\mathcal{L}}, \mathbf{Q}\}$: spectral basis of graph Laplacian $\mathcal{L}$
λ, δ: regularization and confidence parameters
R, C: upper bounds on noise and $\|\boldsymbol{\mu}\|_{\mathbf{\Lambda}}$
Initialization:
$v = R\sqrt{6d\log((\lambda + T)/\delta\lambda)} + C$
$\hat{\boldsymbol{\mu}} = 0_N$, $\boldsymbol{f} = 0_N$, $\mathbf{B} = \mathbf{\Lambda}_{\mathcal{L}} + \lambda \mathbf{I}_N$
Run:
for $t = 1$ **to** T **do**
Sample $\tilde{\boldsymbol{\mu}} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}, v^2\mathbf{B}^{-1})$
$a(t) \leftarrow \arg\max_a \mathbf{b}_a^{\mathsf{T}} \tilde{\boldsymbol{\mu}}$
Observe a noisy reward $r(t) = \mathbf{b}_{a(t)}^{\mathsf{T}} \boldsymbol{\mu} + \varepsilon_t$
$\boldsymbol{f} \leftarrow \boldsymbol{f} + \mathbf{b}_{a(t)} r(t)$
Update $\mathbf{B} \leftarrow \mathbf{B} + \mathbf{b}_{a(t)} \mathbf{b}_{a(t)}^{\mathsf{T}}$
Update $\hat{\boldsymbol{\mu}} \leftarrow \mathbf{B}^{-1} \boldsymbol{f}$
end for

SpectralTS regret bound

- d: Effective dimension.
- λ: Minimal eigenvalue of $\mathbf{\Lambda} = \mathbf{\Lambda}_{\mathcal{L}} + \lambda\mathbf{I}$.
- C: Smoothness upper bound, $\|\boldsymbol{\mu}\|_{\mathbf{\Lambda}} \leq C$.
- $\mathbf{b}_i^\mathsf{T}\boldsymbol{\mu} \in [-1, 1]$ for all i.

The **cumulative regret** R_T of **SpectralTS** is with probability $1 - \delta$ bounded as

$$\mathcal{R}_T \leq \frac{11g}{p}\sqrt{\frac{4+4\lambda}{\lambda} dT \log\frac{\lambda+T}{\lambda}} + \frac{1}{T} + \frac{g}{p}\left(\frac{11}{\sqrt{\lambda}} + 2\right)\sqrt{2T\log\frac{2}{\delta}},$$

where $p = 1/(4e\sqrt{\pi})$ and

$$g = \sqrt{4\log TN}\left(R\sqrt{6d\log\left(\frac{\lambda+T}{\delta\lambda}\right)} + C\right) + R\sqrt{2d\log\left(\frac{(\lambda+T)T^2}{\delta\lambda}\right)} + C.$$

$$R_T \approx d\sqrt{T\log N}$$

SpectralTS analysis sketch

Divide arms into two groups

- $\Delta_i = \mathbf{b}_*^\mathsf{T}\boldsymbol{\mu} - \mathbf{b}_i^\mathsf{T}\boldsymbol{\mu} \leq g\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}$ arm i is **unsaturated**
- $\Delta_i = \mathbf{b}_*^\mathsf{T}\boldsymbol{\mu} - \mathbf{b}_i^\mathsf{T}\boldsymbol{\mu} > g\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}$ arm i is **saturated**

SpectralTS analysis sketch

Divide arms into two groups

- $\Delta_i = \mathbf{b}_*^\mathsf{T}\boldsymbol{\mu} - \mathbf{b}_i^\mathsf{T}\boldsymbol{\mu} \leq g\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}$ arm i is **unsaturated**
- $\Delta_i = \mathbf{b}_*^\mathsf{T}\boldsymbol{\mu} - \mathbf{b}_i^\mathsf{T}\boldsymbol{\mu} > g\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}$ arm i is **saturated**

Saturated arm

- Small standard deviation $\rightarrow$ accurate regret estimate.
- **High regret** on playing the arm $\rightarrow$ **Low probability** of picking

SpectralTS analysis sketch

Divide arms into two groups

- $\Delta_i = \mathbf{b}_*^\mathsf{T}\boldsymbol{\mu} - \mathbf{b}_i^\mathsf{T}\boldsymbol{\mu} \leq g\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}$ arm i is **unsaturated**
- $\Delta_i = \mathbf{b}_*^\mathsf{T}\boldsymbol{\mu} - \mathbf{b}_i^\mathsf{T}\boldsymbol{\mu} > g\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}$ arm i is **saturated**

Saturated arm

- Small standard deviation $\rightarrow$ accurate regret estimate.
- **High regret** on playing the arm $\rightarrow$ **Low probability** of picking

Unsaturated arm

- **Low regret** bounded by a factor of standard deviation
- **High probability** of picking

SpectralTS analysis sketch

- Confidence ellipsoid for estimate $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ (with probability $1 - \delta/T^2$)
  - Using analysis of OFUL algorithm (Abbasi-Yadkori et al., 2011)

$$|\mathbf{b}_i^\mathsf{T}\hat{\boldsymbol{\mu}} - \mathbf{b}_i^\mathsf{T}\boldsymbol{\mu}| \leq \left(R\sqrt{2\log\left(\frac{|\mathbf{B}_T|^{1/2}T^2}{|\boldsymbol{\Lambda}|^{1/2}\delta}\right)} + C \right) \|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}$$

SpectralTS analysis sketch

- Confidence ellipsoid for estimate $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ (with probability $1 - \delta/T^2$)
  - Using analysis of OFUL algorithm (Abbasi-Yadkori et al., 2011)

$$|\mathbf{b}_i^\mathsf{T}\hat{\boldsymbol{\mu}} - \mathbf{b}_i^\mathsf{T}\boldsymbol{\mu}| \leq \left(R\sqrt{2\log\left(\frac{|\mathbf{B}_T|^{1/2}T^2}{|\boldsymbol{\Lambda}|^{1/2}\delta}\right)} + C \right) \|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}$$

- Our key result coming from spectral properties of $\mathbf{B}_t$.

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

SpectralTS analysis sketch

- Confidence ellipsoid for estimate $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ (with probability $1 - \delta/T^2$)
  - Using analysis of OFUL algorithm (Abbasi-Yadkori et al., 2011)

$$|\mathbf{b}_i^\mathsf{T}\hat{\boldsymbol{\mu}} - \mathbf{b}_i^\mathsf{T}\boldsymbol{\mu}| \leq \left(R\sqrt{2\,d\log\left(\frac{(\lambda + T)T^2}{\delta\lambda}\right)} + C\right)\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}} = \ell\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}$$

- Our key result coming from spectral properties of $\mathbf{B}_t$.

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

SpectralTS analysis sketch

- Confidence ellipsoid for estimate $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ (with probability $1 - \delta/T^2$)
  - Using analysis of OFUL algorithm (Abbasi-Yadkori et al., 2011)

$$|\mathbf{b}_i^\mathsf{T}\hat{\boldsymbol{\mu}} - \mathbf{b}_i^\mathsf{T}\boldsymbol{\mu}| \leq \left(R\sqrt{2\,d\log\left(\frac{(\lambda+T)T^2}{\delta\lambda}\right)} + C\right)\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}} = \ell\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}$$

- Our key result coming from spectral properties of $\mathbf{B}_t$.

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

- Concentration of sample $\tilde{\boldsymbol{\mu}}$ around mean $\hat{\boldsymbol{\mu}}$ (with probability $1 - 1/T^2$)
  - Using concentration inequality for Gaussian random variable.

$$|\mathbf{b}_i^\mathsf{T}\tilde{\boldsymbol{\mu}} - \mathbf{b}_i^\mathsf{T}\hat{\boldsymbol{\mu}}| \leq \left(R\sqrt{6d\log\left(\frac{\lambda+T}{\delta\lambda}\right)} + C\right)\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}\sqrt{4\log(TN)} = v\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}\sqrt{4\log(TN)}$$

SpectralTS analysis sketch

Define $\text{regret}'(t) = \text{regret}(t) \cdot \mathbb{1}\{|\mathbf{b}_i^\mathsf{T}\hat{\boldsymbol{\mu}}(t) - \mathbf{b}_i^\mathsf{T}\boldsymbol{\mu}| \le \ell\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}\}$

$$\text{regret}'(t) \le \frac{11g}{p}\|\mathbf{b}_{a(t)}\|_{\mathbf{B}_t^{-1}} + \frac{1}{T^2}$$

SpectralTS analysis sketch

Define $\text{regret}'(t) = \text{regret}(t) \cdot \mathbb{1}\{|\mathbf{b}_i^\mathsf{T} \hat{\boldsymbol{\mu}}(t) - \mathbf{b}_i^\mathsf{T} \boldsymbol{\mu}| \leq \ell \|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}\}$

$$\text{regret}'(t) \leq \frac{11g}{p} \|\mathbf{b}_{a(t)}\|_{\mathbf{B}_t^{-1}} + \frac{1}{T^2}$$

Super-martingale (i.e. $\mathbb{E}[Y_t - Y_{t-1}|\mathcal{F}_{t-1}] \leq 0$)

$$X_t = \text{regret}'(t) - \frac{11g}{p} \|\mathbf{b}_{a(t)}\|_{\mathbf{B}_t^{-1}} - \frac{1}{T^2}$$

$$Y_t = \sum_{w=1}^{t} X_w.$$

$(Y_t;\ t = 0, \ldots, T)$ is a **super-martingale** process w.r.t. history $\mathcal{F}_t$.

SpectralTS analysis sketch

Define $\text{regret}'(t) = \text{regret}(t) \cdot \mathbb{1}\{|\mathbf{b}_i^\mathsf{T}\hat{\boldsymbol{\mu}}(t) - \mathbf{b}_i^\mathsf{T}\boldsymbol{\mu}| \leq \ell\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}\}$

$$\text{regret}'(t) \leq \frac{11g}{p}\|\mathbf{b}_{a(t)}\|_{\mathbf{B}_t^{-1}} + \frac{1}{T^2}$$

Super-martingale (i.e. $\mathbb{E}[Y_t - Y_{t-1}|\mathcal{F}_{t-1}] \leq 0$)

$$X_t = \text{regret}'(t) - \frac{11g}{p}\|\mathbf{b}_{a(t)}\|_{\mathbf{B}_t^{-1}} - \frac{1}{T^2}$$

$$Y_t = \sum_{w=1}^{t} X_w.$$

$(Y_t;\ t = 0, \ldots, T)$ is a **super-martingale** process w.r.t. history $\mathcal{F}_t$.

Azuma-Hoeffding inequality for super-martingale, w. p. $1 - \delta/2$:

$$\sum_{t=1}^{T}\text{regret}'(t) \leq \frac{11g}{p}\sum_{t=1}^{T}\|\mathbf{b}_{a(t)}\|_{\mathbf{B}_t^{-1}} + \frac{1}{T} + \frac{g}{p}\left(\frac{11}{\sqrt{\lambda}} + 2\right)\sqrt{2T\ln\frac{2}{\delta}}$$

SpectralTS analysis sketch

Define $\text{regret}'(t) = \text{regret}(t) \cdot \mathbb{1}\{|\mathbf{b}_i^\top \hat{\boldsymbol{\mu}}(t) - \mathbf{b}_i^\top \boldsymbol{\mu}| \leq \ell \|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}\}$

$$\text{regret}'(t) \leq \frac{11g}{p}\|\mathbf{b}_{a(t)}\|_{\mathbf{B}_t^{-1}} + \frac{1}{T^2}$$

Super-martingale (i.e. $\mathbb{E}[Y_t - Y_{t-1}|\mathcal{F}_{t-1}] \leq 0$)

$$X_t = \text{regret}'(t) - \frac{11g}{p}\|\mathbf{b}_{a(t)}\|_{\mathbf{B}_t^{-1}} - \frac{1}{T^2}$$

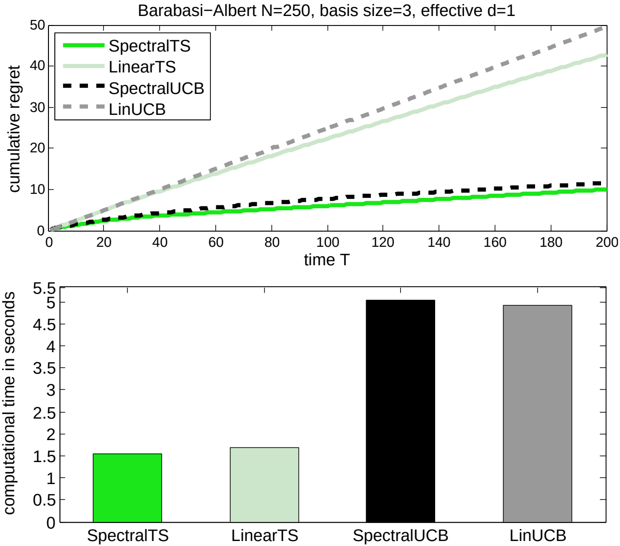
$$Y_t = \sum_{w=1}^{t} X_w.$$

$(Y_t;\ t = 0, \ldots, T)$ is a **super-martingale** process w.r.t. history $\mathcal{F}_t$.

Azuma-Hoeffding inequality for super-martingale, w. p. $1 - \delta/2$:

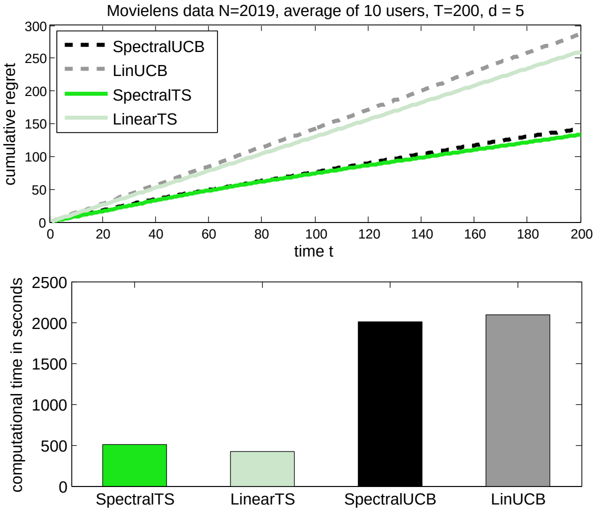
$$\sum_{t=1}^{T} \text{regret}'(t) \leq \frac{11g}{p} \sum_{t=1}^{T} \|\mathbf{b}_{a(t)}\|_{\mathbf{B}_t^{-1}} + \frac{1}{T} + \frac{g}{p}\left(\frac{11}{\sqrt{\lambda}} + 2\right)\sqrt{2T \ln \frac{2}{\delta}}$$

Synthetic experiment



Real world experiment

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



Conclusion

- **New algorithm** for spectral bandit setting.
- **SpectralTS**
  - **Regret bound** $\approx d\sqrt{T \log N}$
    - Bound scales with **effective dimension** $d \ll D$.
    - Comparable to SpectralUCB
  - **Computational complexity** $\approx D^2 + DN$
    - Better than SpectralUCB $\approx D^2N$

Thank you!

Tomáš Kocák

tomas.kocak@inria.fr

sequel.lille.inria.fr