Option Discovery in the Absence of Rewards with Manifold Analysis

Amitay Bar, Ronen Talmon and Ron Meir

Viterbi Faculty of Electrical Engineering

Technion - Israel Institute of Technology

Technion
Israel Institute of Technology

THE ANDREW & ERNA VITERBI
FACULTY OF
ELECTRICAL
ENGINEERING

Option Discovery

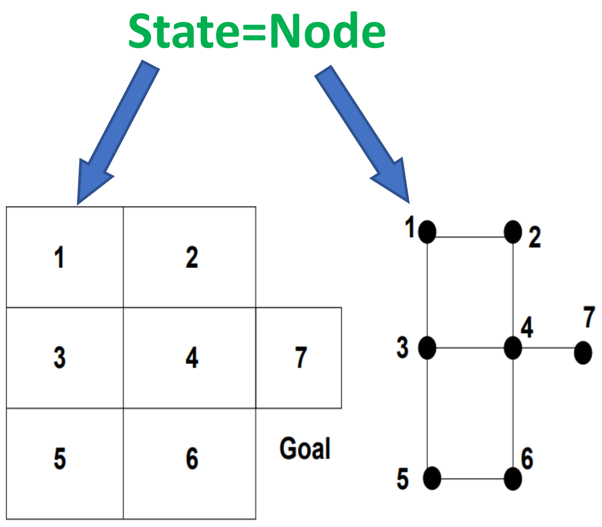
- **We address the problem of option discovery**
- **Options** (a.k.a. skills) are a predefined sequence of primitive actions [Sutton et al. ‘99]
- Options were shown to improve both learning and exploration
- **Setting**
  - Not associated with any specific task
  - Acquired without receiving any reward
  - Important and challenging problem in RL

Contribution

- A new approach to option discovery with theoretical foundation
  - Based on manifold analysis
- The analysis includes novel results in manifold learning
- We propose an algorithm for option discovery
  - Outperforms competing options

Graph Based Approach

- The finite domain is represented by a graph [Mahadevan '07]
    - Nodes - the states ($\mathbb{S}$ is the set of states)
    - Edges - according to the state's connectivity
- The graph is a discrete representation of a manifold



M – Adjacency matrix

$$
\begin{array}{c}
\\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7
\end{array}
\begin{array}{c}
\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
\begin{pmatrix}
0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0
\end{pmatrix}
\end{array}
$$

D – Degree matrix

$$
\begin{array}{c}
\\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7
\end{array}
\begin{array}{c}
\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
\begin{pmatrix}
2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 3 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 4 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}
$$

The Proposed Algorithm

1. Compute the random walk matrix $\boldsymbol{W} = \frac{1}{2}(\boldsymbol{I} - \boldsymbol{M}\boldsymbol{D}^{-1})$
2. Apply EVD to $\boldsymbol{W}$ and obtain its left and right eigenvectors $\{\phi_i\}, \{\tilde{\phi}_i\}$, and its eigenvalues $\{\omega_i\}$
3. Construct $\boxed{f_t: \mathbb{S} \to \mathbb{R}, \quad f_t(s) = \left\| \sum_{i \geq 2} \omega_i^t \phi_i(s) \tilde{\phi}_i \right\|^2}$

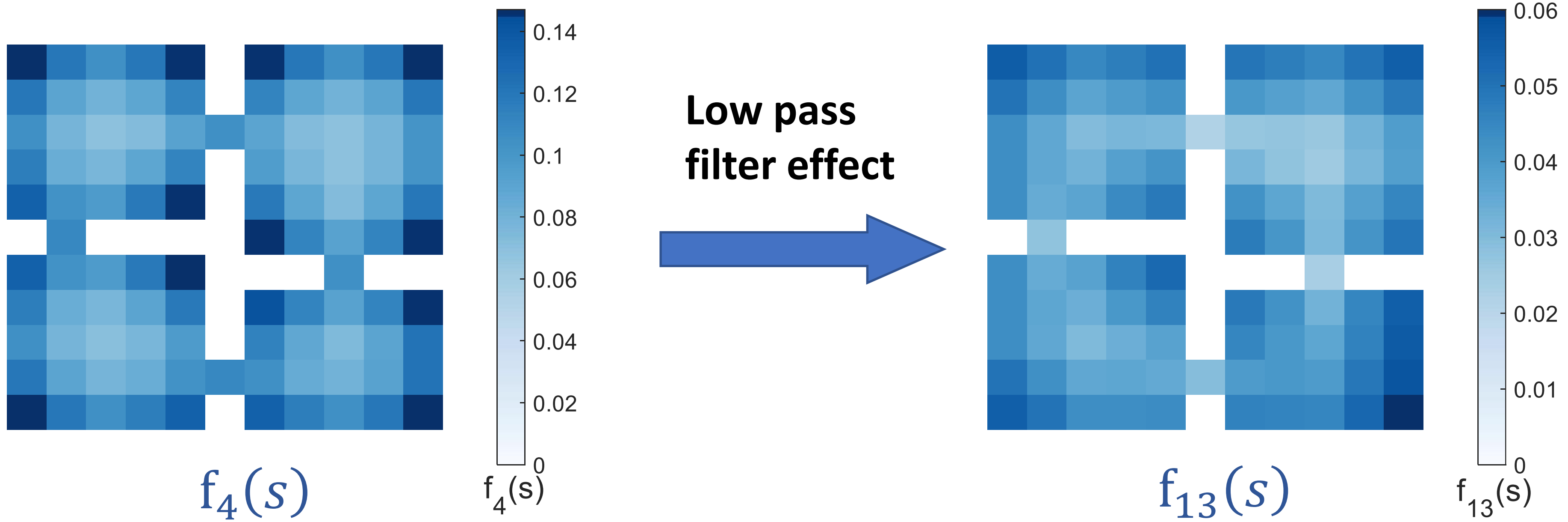
   To be motivated later
4. Find the local maxima of $f_t(s)$, denoted as $\left\{s_o^{(i)}\right\} \subset \mathbb{S}$
5. For each local maximum, $s_o^{(i)}$, build an option leading to it

f_t allows the identification of goal states

Demonstrating the Score Function

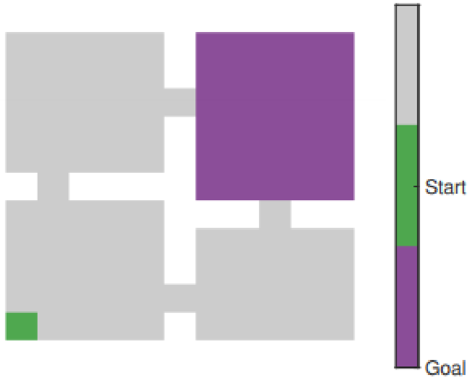
$$f_t(s) = \left\| \sum_{i \geq 2} \omega_i^t \phi_i(s) \tilde{\phi}_i \right\|^2$$

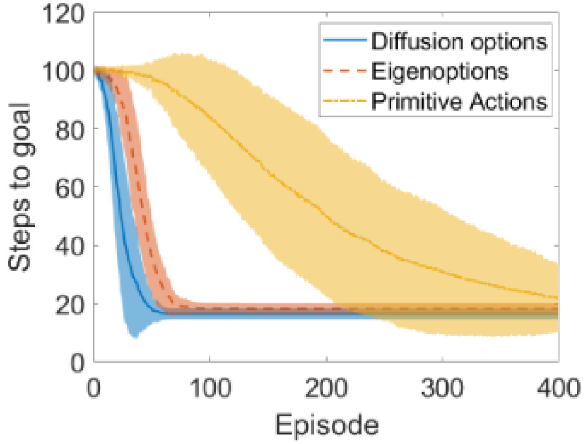
- **4Rooms** [Sutton et al. '99]
- The local maxima of $f_t(s)$ are at states that are "far away" from all other states
  - Corner states and bottleneck states



$f_4(s)$ $f_{13}(s)$

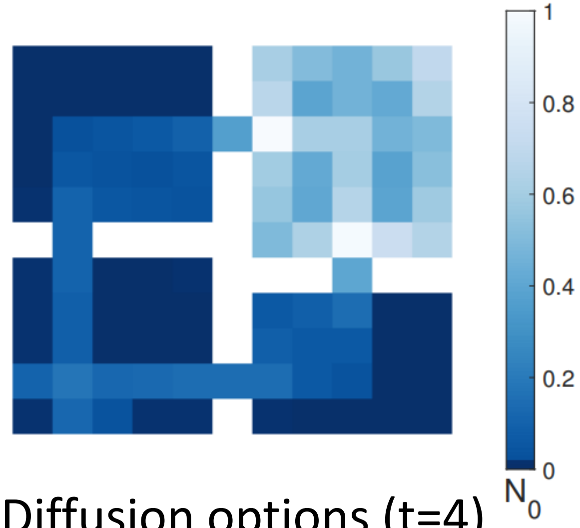
Experimental Results - Learning



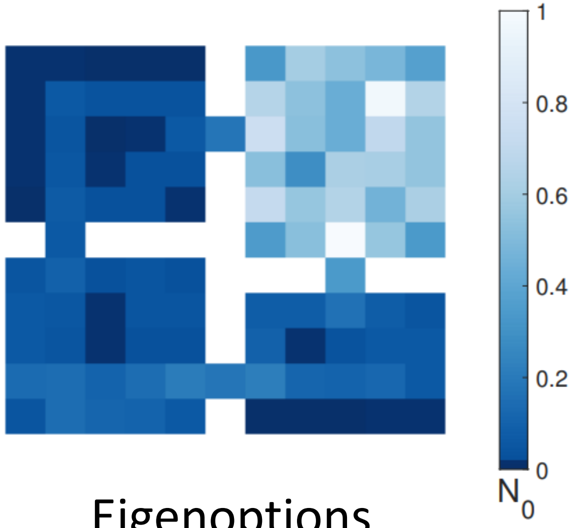


- Q learning
  [Watkins and Dayan, ‘92]
- Eigenoptions
  [Machado et al. ‘17]

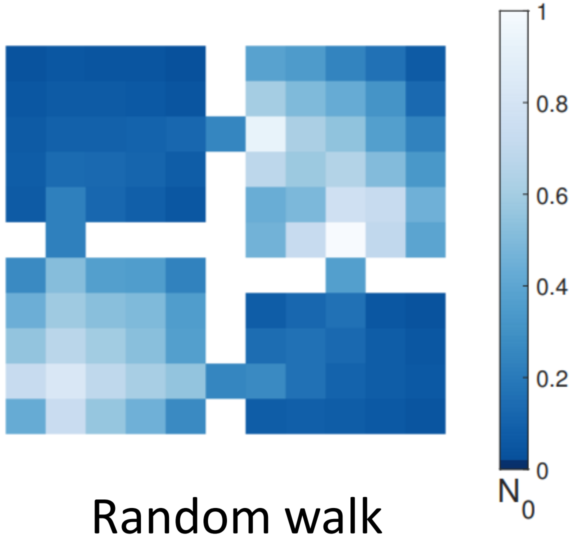
Normalized visitation during learning



Diffusion options (t=4)



Eigenoptions

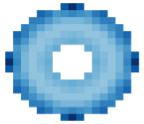


Random walk

* Further results in paper

Experimental Results - Exploration

- Exploration
  - Median number of steps between every two states [Machado et al. ‘17]

Domain (#states)	t	#options	Diffusion Options	Eigenoptions	Cover Options	Random Walk
 Ring (192)	4	32	**217**	301	361	565
	13	28	**219**	279	363	565
 Maze (148)	4	19	**282**	446	525	1280
	13	14	**249**	641	498	1280
 4Rooms (104)	4	20	**147**	160	179	487
	13	15	**140**	162	175	487

[Machado et al. ’17] [Jinnai et al. ‘19]

Theoretical Analysis

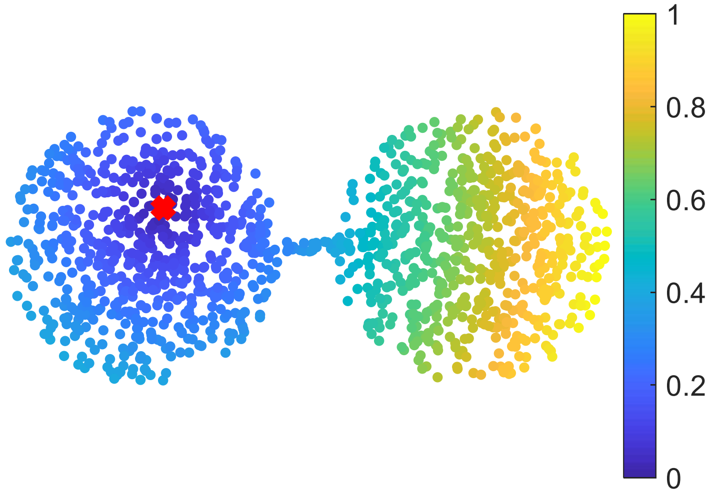
- We use **manifold learning** results and concepts
  - **Diffusion distance** [Coifman and Laffon ‘06]
  - New concept – considering the entire spectrum [Cheng and Mishne ‘18]
- **Comparison to existing work -** eigenoptions [Machado et al. ‘17] and cover options [Jinnai et al. ‘19]
  - Use only the principal components instead of all/many
  - Consider only one eigenvector at a time, instead of incorporating them together

Diffusion Distance

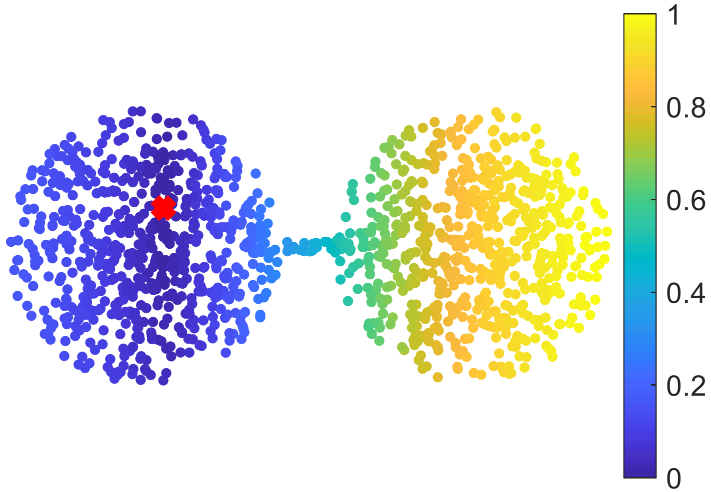
$$W = \frac{1}{2}(\boldsymbol{I} - \boldsymbol{M}\boldsymbol{D}^{-1})$$

- Consider $\boldsymbol{W}^t = \begin{array}{c} \begin{array}{ccc} & \color{red}{\boldsymbol{w}_l^t} & \end{array} \\ \begin{pmatrix} \cdots & \boxed{} & \cdots & \end{pmatrix} \end{array}$

$$D_t(s, s') = \|\boldsymbol{w}_s^t - \boldsymbol{w}_{s'}^t\|$$



Euclidean distance



Diffusion distance

Properties of the Score Function

Proposition 1

The function $f_t: \mathbb{S} \to \mathbb{R}$ can be expressed as

$$\boxed{f_t(s) = \langle D_t^2(s, s') \rangle_{s' \in \mathbb{S}} + const}$$

- $\langle D_t^2(s, s') \rangle_{s' \in \mathbb{S}}$ is the average diffusion distance between state s and all other states

*See ICML paper for the proof

Properties of the Score Function

Proposition 1

The function $f_t : \mathbb{S} \to \mathbb{R}$ can be expressed as

$$\boxed{f_t(s) = \langle D_t^2(s, s') \rangle_{s' \in \mathbb{S}} + const}$$

- Option discovery: $\max f_t(s) = \max \langle D_t^2(s, s') \rangle_{s' \in \mathbb{S}}$

Exploration benefits

- Agent visits different regions
- Avoiding the dithering effect of random walk

*See ICML paper for the proof

Properties of the Score Function

Proposition 2

Relates $f_t(s)$ to $\boldsymbol{\pi}_0$, the stationary distribution of the graph

$$\boxed{f_t(s) = \left\| \boldsymbol{p}_t^{(s)} - \boldsymbol{\pi}_0 \right\|^2}$$

$$\boxed{f_t(s) \leq \omega_2^{2t} \left(\frac{1}{\boldsymbol{\pi}_0(s)} - 1 \right)}$$

- PageRank algorithm [Page et al. ’99, Kleinberg ‘99]

Exploration benefits

- Diffusion options lead to states for which $\boldsymbol{\pi}_0(s)$ is small
- Rarely visited by an uninformed random walk

*See ICML paper for the proof

Extensions and Scaling Up

- Extending diffusion options to stochastic domains
  - Stochastic domains → can lead to asymmetric matrices
  - We use polar decomposition on the graph Laplacian [Mhaskar '18]
- Scaling up to large scale domains/function approximation case
  - [Wu et al. '19], [Jinnai et al. '20]
- See ICML paper for further discussion and results

Summary

- We introduced **theoretically motivated** options
- Analysis based on concepts from **manifold learning**
- Diffusion options **encourage exploration**
  - Lead to distant states in term of **diffusion distance**
  - Compensate for low **stationary distribution** values
- Empirically demonstrated **improved performance**
  - Both **learning** and **exploration**

Thank you

“Option Discovery in the Absence of Rewards with Manifold Analysis”,
A. Bar, R. Talmon and R. Meir, ICML 2020