Sparse Multi-Task Reinforcement Learning

Daniele Calandriello, Alessandro Lazaric, Marcello Restelli
SequeL - INRIA Lille, Politecnico di Milano

SequeL – INRIA Lille

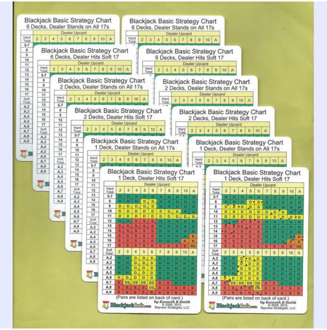
Seminars

Lille, October 2014

Sparse Multi-Task Reinforcement Learning

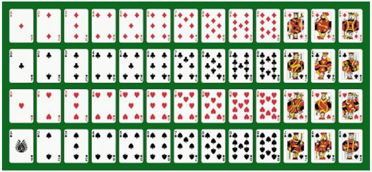
Sparse Multi-Task Reinforcement Learning

Sparse Multi-Task Reinforcement Learning

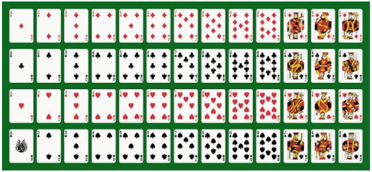


Sparse Multi-Task Reinforcement Learning

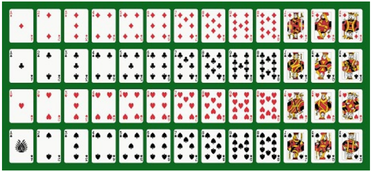
Sparse Multi-Task Reinforcement Learning

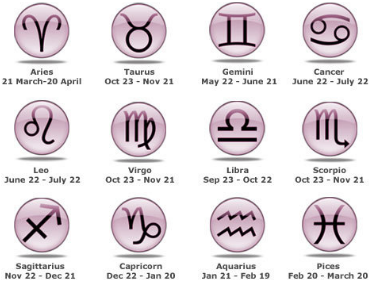


Sparse Multi-Task Reinforcement Learning



Sparse Multi-Task Reinforcement Learning





Talk Overview

- Reinforcement Learning $\rightarrow$ Linear Fitted Q Iteration (LinFQI)
- Sparse Markov Decision Process $\rightarrow$ LASSO FQI
- Multi-Task (Group) Sparsity $\rightarrow$ Group-LASSO FQI
- Learning Sparse Representations $\rightarrow$ Feature Learning FQI
- Experiments

Reinforcement Learning

Markov Decision Process (MDP): $\mathcal{M} = (\mathcal{X}, \mathcal{A}, R, P, \gamma)$

- $\mathcal{X}$ is a bounded closed subset of the Euclidean space
- $\mathcal{A}$ is finite (i.e., $|\mathcal{A}| < \infty$)
- $R : \mathcal{X} \times \mathcal{A} \to [0, 1]$
- $P : \mathcal{X} \times \mathcal{A} \to \mathcal{P}(\mathcal{X})$
- γ: discount factor

Policy: $\pi : \mathcal{X} \to \mathcal{A}$

Reinforcement Learning

Optimal Action-Value Function:

$Q^*(x,a) = \max_\pi \mathbb{E}_\pi[\sum_{i=1}^\infty \gamma^i r_i | r_i \sim R(x_i, \pi(x_i)), x_0 = x, a_0 = a]$

$\pi^*(x) = \arg\max_{a \in \mathcal{A}} Q^*(x,a)$

Reinforcement Learning

Optimal Action-Value Function:

$$Q^*(x,a) = \max_\pi \mathbb{E}_\pi[\sum_{i=1}^\infty \gamma^i r_i | r_i \sim R(x_i, \pi(x_i)), x_0 = x, a_0 = a]$$
$$\pi^*(x) = \arg\max_{a \in \mathcal{A}} Q^*(x,a)$$

Optimal Bellman Operator:

$$\mathcal{T}Q(x,a) = R(x,a) + \gamma \sum_y P(y|x,a) \max_{a'} Q(y,a')$$
$$\mathcal{T}Q^* = Q^*$$

Value Iteration

Exact Value Iteration:

$$
\begin{aligned}
&Q^0\\
&\mathcal{T}Q^0 = Q^1\\
&\mathcal{T}Q^1 = Q^2\\
&\dots\\
&\mathcal{T}Q^K = Q^*
\end{aligned}
$$

Value Iteration

Exact Value Iteration:

$$Q^0$$
$$\mathcal{T}Q^0 = Q^1$$
$$\mathcal{T}Q^1 = Q^2$$
$$\dots$$
$$\mathcal{T}Q^K = Q^*$$

Approximate Value Iteration:

$$\widetilde{Q}^0$$
$$\mathcal{T}\widetilde{Q}^0 \rightsquigarrow \widehat{Q}^1 \rightsquigarrow \widetilde{Q}^1$$
$$\mathcal{T}\widetilde{Q}^1 \rightsquigarrow \widehat{Q}^2 \rightsquigarrow \widetilde{Q}^2$$
$$\dots$$
$$\mathcal{T}\widetilde{Q}^K \rightsquigarrow \widehat{Q}^* \rightsquigarrow \widetilde{Q}^*$$

Fitted Q Iteration

Approximation:

Fitted Q Iteration

Approximation:

Approximate $\mathcal{T}$: use samples

$z^k_{i,a,t} = r^k_{i,a,t} + \gamma \max_{a'} \widetilde{Q}^k_t(y^k_{i,a,t}, a')$

$\mathcal{D}^k_{a,t} = \{(x_{i,t}, a), z^k_{i,a,t}\}_{i=1}^{n_x}$

Fitted Q Iteration

Approximation:

Approximate $\mathcal{T}$: use samples

$$z^k_{i,a,t} = r^k_{i,a,t} + \gamma \max_{a'} \widetilde{Q}^k_t(y^k_{i,a,t}, a')$$

$$\mathcal{D}^k_{a,t} = \{(x_{i,t}, a), z^k_{i,a,t}\}_{i=1}^{n_x}$$

Approximate $\rightsquigarrow$: use regression

Fitted Q Iteration

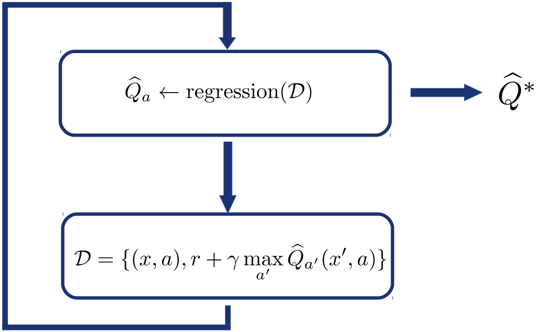
Approximation:

Approximate $\mathcal{T}$: use samples

$z^k_{i,a,t} = r^k_{i,a,t} + \gamma \max_{a'} \widetilde{Q}^k_t(y^k_{i,a,t}, a')$

$\mathcal{D}^k_{a,t} = \{(x_{i,t}, a), z^k_{i,a,t}\}_{i=1}^{n_x}$

Approximate $\rightsquigarrow$: use regression



Linear Fitted Q Iteration

Linear Approximation:

- $\phi(\cdot) = [\varphi_1(\cdot), \varphi_2(\cdot), \ldots, \varphi_d(\cdot)]^\mathsf{T}$
- $\varphi_i : \mathcal{X} \to \mathbb{R}$
- $\sup_x ||\phi(x)||_2 \leq L$
- $\mathcal{F} = \{f_w(x, a) = \phi(x)^\mathsf{T} w_a, x \in \mathcal{X}, a \in \mathcal{A}, w_a \in \mathbb{R}^d\}$

Linear Fitted Q Iteration

Linear Approximation:

- $\phi(\cdot) = [\varphi_1(\cdot), \varphi_2(\cdot), \ldots, \varphi_d(\cdot)]^\mathsf{T}$
- $\varphi_i : \mathcal{X} \to \mathbb{R}$
- $\sup_x ||\phi(x)||_2 \leq L$
- $\mathcal{F} = \{f_w(x, a) = \phi(x)^\mathsf{T} w_a, x \in \mathcal{X}, a \in \mathcal{A}, w_a \in \mathbb{R}^d\}$

Matrix notation for samples:

- $\Phi = [\phi(x_1)^\mathsf{T}; \cdots; \phi(x_{n_x})^\mathsf{T}] \in \mathbb{R}^{n_x \times d}$
- $\Phi_a'^k = [\phi(y_{i,a}^k)^\mathsf{T}; \cdots; \phi(y_{n_x,a}^k)^\mathsf{T}] \in \mathbb{R}^{n_x \times d}$
- $R_a^k = [r_{1,a}^k, \ldots, r_{n_x,a}^k] \in \mathbb{R}^{n_x}$
- $Z_a^k = [z_{1,a}^k, \ldots, z_{n_x,a}^k] \in \mathbb{R}^{n_x}$, with $Z_a^k = R_a^k + \gamma \max_{a'}(\Phi_{a'}'^k w_{a'}^{k-1})$

Least Squares Regression

Linear Model Interpretation

$$z_{i,a}^k = \mathcal{T}\widehat{Q}^{k-1}(x_i, a) + \eta_{i,a}^k = \phi(x_i)^\mathsf{T} w_a + \eta_{i,a}^k$$

Least Squares Regression

Linear Model Interpretation

$$z_{i,a}^k = \mathcal{T}\widehat{Q}^{k-1}(x_i, a) + \eta_{i,a}^k = \phi(x_i)^\mathsf{T} w_a + \eta_{i,a}^k$$

Unbiased estimator: squared loss

$$\widehat{w}_a^k = \arg\min_{w \in \mathbb{R}^d} \frac{1}{n_x} \sum_{i=1}^{n_x} \left(\phi(x_i)^\mathsf{T} w - z_{i,a}^k\right)^2 = \frac{1}{n_x} \left\|\Phi w_a^k - Z_a^k\right\|_2^2$$

Least Squares Regression

Linear Model Interpretation

$$z_{i,a}^k = \mathcal{T}\widehat{Q}^{k-1}(x_i, a) + \eta_{i,a}^k = \phi(x_i)^\mathsf{T} w_a + \eta_{i,a}^k$$

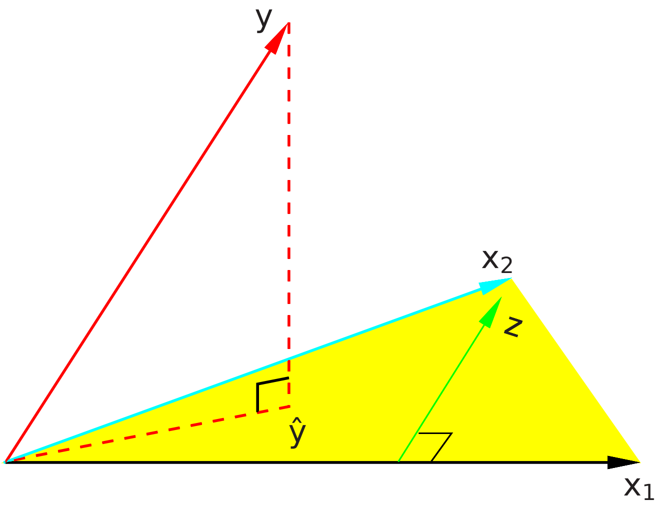
Unbiased estimator: squared loss

$$\widehat{w}_a^k = \arg\min_{w \in \mathbb{R}^d} \frac{1}{n_x} \sum_{i=1}^{n_x} \left(\phi(x_i)^\mathsf{T} w - z_{i,a}^k\right)^2 = \frac{1}{n_x} \left\|\Phi w_a^k - Z_a^k\right\|_2^2$$

Asymptotically: $\left\|w_a^k - \widehat{w}_a^k\right\|_2$ is small

Ordinary Least Square (OLS):

$$\widehat{w}_a^k = (\Phi^\mathsf{T}\Phi)^{-1}\Phi^\mathsf{T} Z_a^k$$



Talk Overview

- ~~Reinforcement Learning $\rightarrow$ Linear Fitted Q Iteration (LinFQI)~~
- Sparse Markov Decision Process $\rightarrow$ LASSO FQI
- Multi-Task (Group) Sparsity $\rightarrow$ Group-LASSO FQI
- Learning Sparse Representations $\rightarrow$ Feature Learning FQI
- Experiments

High Dimensional Assumption

? **Problem:** the regression problem must approximate $\mathcal{T}Q$ well

High Dimensional Assumption

? **Problem:** the regression problem must approximate $\mathcal{T}Q$ well

! **Solution:** use large number of features, rich feature space captures everything

High Dimensional Assumption

? **Problem:** the regression problem must approximate $\mathcal{T}Q$ well

! **Solution:** use large number of features, rich feature space captures everything

Assumption

The space $\mathcal{F}$ is such that for any function $f_w \in \mathcal{F}$, the image of the Bellman operator $\mathcal{T}$ is always in $\mathcal{F}$, i.e., $\mathcal{T} f_w \in \mathcal{F}$.

High Dimensional Assumption

? **Problem:** the regression problem must approximate $\mathcal{T}Q$ well

! **Solution:** use large number of features, rich feature space captures everything

Assumption

The space $\mathcal{F}$ is such that for any function $f_w \in \mathcal{F}$, the image of the Bellman operator $\mathcal{T}$ is always in $\mathcal{F}$, i.e., $\mathcal{T} f_w \in \mathcal{F}$.

? **Problem:** when $d > n_x$, the OLS projection $(\Phi^\mathsf{T}\Phi)^{-1}$ is not defined

High Dimensional Assumption

? **Problem:** the regression problem must approximate $\mathcal{T}Q$ well

! **Solution:** use large number of features, rich feature space captures everything

Assumption

The space $\mathcal{F}$ is such that for any function $f_w \in \mathcal{F}$, the image of the Bellman operator $\mathcal{T}$ is always in $\mathcal{F}$, i.e., $\mathcal{T} f_w \in \mathcal{F}$.

? **Problem:** when $d > n_x$, the OLS projection $(\Phi^\mathsf{T}\Phi)^{-1}$ is not defined

! **Solution:** Get more samples

High Dimensional Assumption

? **Problem:** the regression problem must approximate $\mathcal{T}Q$ well

! **Solution:** use large number of features, rich feature space captures everything

Assumption

The space $\mathcal{F}$ is such that for any function $f_w \in \mathcal{F}$, the image of the Bellman operator $\mathcal{T}$ is always in $\mathcal{F}$, i.e., $\mathcal{T} f_w \in \mathcal{F}$.

? **Problem:** when $d > n_x$, the OLS projection $(\Phi^{\mathsf{T}}\Phi)^{-1}$ is not defined

! ~~**Solution:** Get more samples~~

High Dimensional Assumption

? **Problem:** the regression problem must approximate $\mathcal{T}Q$ well

! **Solution:** use large number of features, rich feature space captures everything

Assumption

The space $\mathcal{F}$ is such that for any function $f_w \in \mathcal{F}$, the image of the Bellman operator $\mathcal{T}$ is always in $\mathcal{F}$, i.e., $\mathcal{T} f_w \in \mathcal{F}$.

? **Problem:** when $d > n_x$, the OLS projection $(\Phi^\mathsf{T}\Phi)^{-1}$ is not defined

! ~~**Solution:** Get more samples~~

! **Solution:** Use less features

High Dimensional Assumption

? **Problem:** the regression problem must approximate $\mathcal{T}Q$ well

! **Solution:** use large number of features, rich feature space captures everything

Assumption

The space $\mathcal{F}$ is such that for any function $f_w \in \mathcal{F}$, the image of the Bellman operator $\mathcal{T}$ is always in $\mathcal{F}$, i.e., $\mathcal{T} f_w \in \mathcal{F}$.

? **Problem:** when $d > n_x$, the OLS projection $(\Phi^\mathsf{T}\Phi)^{-1}$ is not defined

! ~~**Solution:** Get more samples~~

! ~~**Solution:** Use less features~~

High Dimensional Assumption

? **Problem:** the regression problem must approximate $\mathcal{T}Q$ well

! **Solution:** use large number of features, rich feature space captures everything

Assumption

The space $\mathcal{F}$ is such that for any function $f_w \in \mathcal{F}$, the image of the Bellman operator $\mathcal{T}$ is always in $\mathcal{F}$, i.e., $\mathcal{T}f_w \in \mathcal{F}$.

? **Problem:** when $d > n_x$, the OLS projection $(\Phi^\mathsf{T}\Phi)^{-1}$ is not defined

! ~~**Solution:** Get more samples~~

! ~~**Solution:** Use less features~~

! **Solution:** Let the regression select useful features

Regularization and Sparsity

Sparse representation:

$$|J(w)| = \|w\|_0 = s << d$$

Regularization and Sparsity

Sparse representation:

$$|J(w)| = \|w\|_0 = s << d$$

$$\overbrace{\underbrace{1, 0.64, 0.49, 0.36, 0.25}_{s} \; 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \ldots}^{d} \qquad s << d$$

Regularization and Sparsity

Sparse representation:

$$|J(w)| = \|w\|_0 = s << d$$

$$\overbrace{\underbrace{\boxed{1, 0.64, 0.49, 0.36, 0.25}}_{s}\boxed{0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \ldots}}^{d} \qquad s << d$$

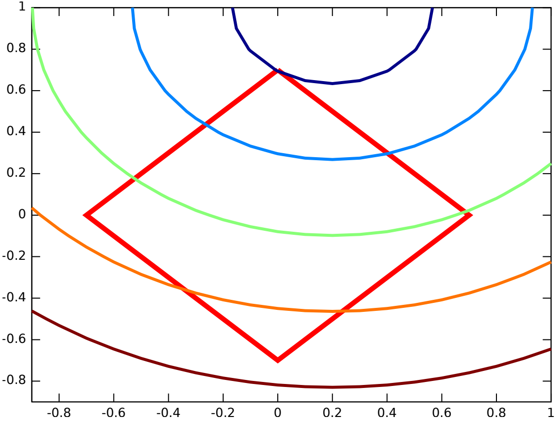
$$\min_{w \in \mathbb{R}^{d_x}} \frac{1}{n_x} \sum_{i=1}^{n_x} \left(\phi(x_i)^\mathsf{T} w - z_{i,a}^k \right)^2 \text{ s.t. } ||w||_0 < \lambda$$

$$\Updownarrow$$

$$\min_{w \in \mathbb{R}^{d_x}} \frac{1}{n_x} \sum_{i=1}^{n_x} \left(\phi(x_i)^\mathsf{T} w - z_{i,a}^k \right)^2 + \lambda' ||w||_0$$

ℓ_1 regularization and LASSO

$$\widehat{w}_a^k = \arg\min_{w \in \mathbb{R}^{d_x}} \frac{1}{n_x} \sum_{i=1}^{n_x} \left(\phi(x_i)^\mathsf{T} w - z_{i,a}^k \right)^2 + \lambda ||w||_1.$$



Sparse Value Iteration

Exact Value Iteration:

Q^0: w^0, $|J(w^0)| = s^0$

Sparse Value Iteration

Exact Value Iteration:

Q^0: w^0, $|J(w^0)| = s^0$

$\mathcal{T}Q^0 = Q^1$: w^1, $|J(w^1)| = s^1$, depends on s^0

Sparse Value Iteration

Exact Value Iteration:

Q^0: w^0, $|J(w^0)| = s^0$

$\mathcal{T}Q^0 = Q^1$: w^1, $|J(w^1)| = s^1$, depends on s^0

$\mathcal{T}Q^1 = Q^2$: w^2, $|J(w^2)| = s^2$, depends on s^1

Sparse Value Iteration

Exact Value Iteration:

Q^0: w^0, $|J(w^0)| = s^0$

$\mathcal{T}Q^0 = Q^1$: w^1, $|J(w^1)| = s^1$, depends on s^0

$\mathcal{T}Q^1 = Q^2$: w^2, $|J(w^2)| = s^2$, depends on s^1

$\ldots$

Sparse Value Iteration

Exact Value Iteration:

Q^0: w^0, $|J(w^0)| = s^0$

$\mathcal{T}Q^0 = Q^1$: w^1, $|J(w^1)| = s^1$, depends on s^0

$\mathcal{T}Q^1 = Q^2$: w^2, $|J(w^2)| = s^2$, depends on s^1

$\dots$

$\mathcal{T}Q^K = Q^*$: w^*, $|J(w^*)| = s^*$, depends on s^K

Sparse MDPs

Assumption (Sparse MDPs)

Given the sets of states $\mathcal{S} = \{x_i\}_{i=1}^{n_x}$ *used in* FQI*, there exists a set* J *(the set of useful features) for MDP* $\mathcal{M}$*, with* $|J| = s \ll d$*, such that for any* $i \notin J$*, any* $j \in [d_x]$ *and any policy* π

$$\sum_{x \in \mathcal{S}} \varphi_i(x) \int_{x' \in \mathcal{X}} P(\mathrm{d}x'|x, \pi(x))\varphi_j(x')\mathrm{d}x = 0, \tag{1}$$

and there exists a function $f_{w^R} = R$ *such that* $J(w^R) \subseteq J$.

Sparse MDPs

Lemma

Under High Dimensional Assumption and Sparse MDPs Assumption, the application of the Bellman operator $\mathcal{T}$ to any function $f_w \in \mathcal{F}$, produces a function $f_{w'} = \mathcal{T} f_w \in \mathcal{F}$ such that $J(w') \subseteq J$.

LASSO-FQI, Theoretical Guarantees

Theorem (LASSO-FQI)

If LASSO-FQI is run for K iterations with a regularizer

$$\lambda = \delta Q_{\max}\sqrt{\frac{\log d}{n}},$$

for any numerical constant $\delta > 8$, then with probability at least $(1 - 2d^{1-\delta/8})^K$, the performance loss is bounded as

$$\left\| Q^* - Q^{\pi^K} \right\|_{2,\mu}^2 \leq \mathcal{O}\left(\frac{1}{(1-\gamma)^4} \left[\frac{Q_{\max}^2 L^2}{\kappa_{\min}^4(s)} \frac{s \log d}{n} + \gamma^K Q_{\max}^2 \right] \right)$$

Talk Overview

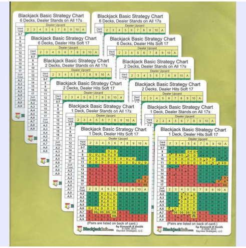
- ~~Reinforcement Learning $\rightarrow$ Linear Fitted Q Iteration (LinFQI)~~
- ~~Sparse Markov Decision Process $\rightarrow$ LASSO FQI~~
- Multi-Task (Group) Sparsity $\rightarrow$ Group-LASSO FQI
- Learning Sparse Representations $\rightarrow$ Feature Learning FQI
- Experiments

Multi-Task

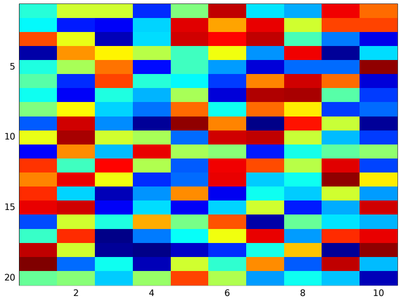
Multi-Task MDP $\mathcal{M}_t = (\mathcal{X}, \mathcal{A}, P_t, R_t, \gamma_t)$, $t \in [T] = \{1, \dots, T\}$

Performance measure:

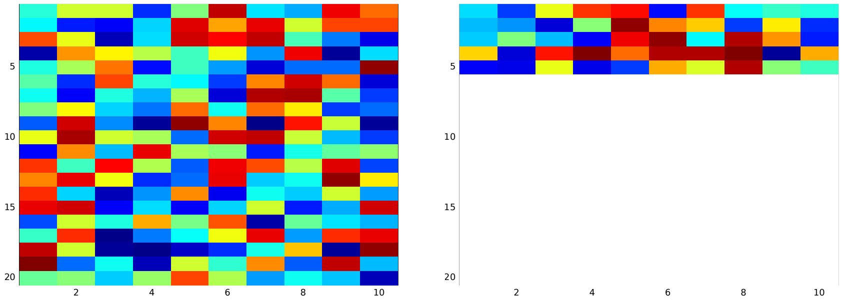
$$\frac{1}{T} \sum_{t=1}^{T} \left\| Q_t^* - Q_t^{\pi_t^K} \right\|_{2,\mu}^2$$



Group Sparsity and Group Lasso

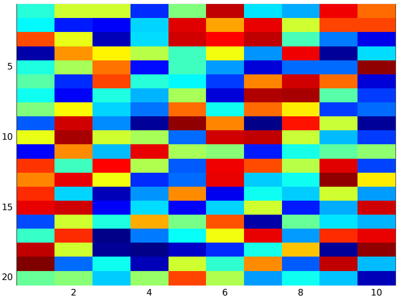


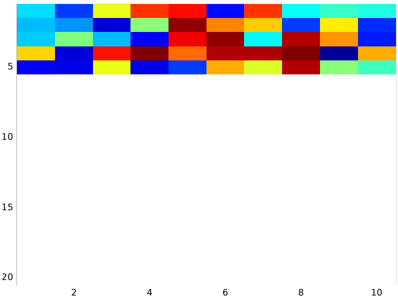
Group Sparsity and Group Lasso



Group Sparsity and Group Lasso

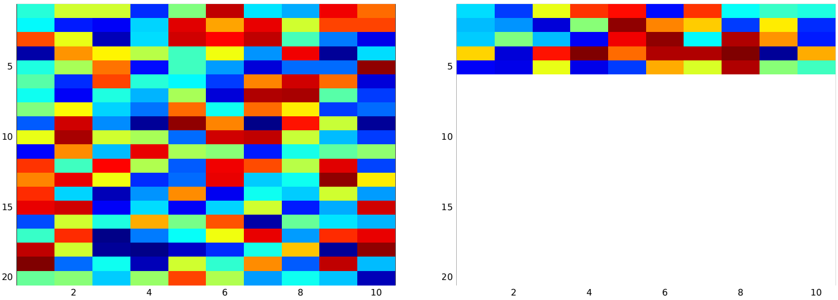
$$\ell_{2,1}\text{-norm } \|W\|_{2,1} = \sum_{i=1}^{d} \|[W]^i\|_2$$





Group Sparsity and Group Lasso

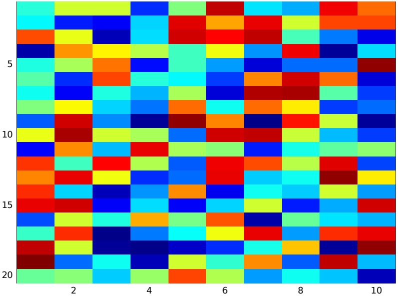
$$\ell_{2,1}\text{-norm } \|W\|_{2,1} = \sum_{i=1}^{d} \|[W]^i\|_2$$

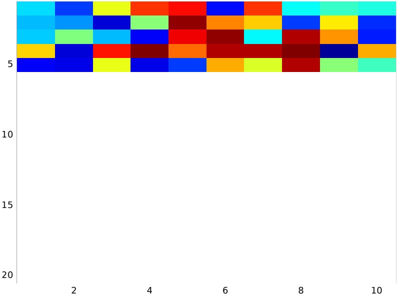


$$\text{high} \leftarrow \|W\|_{2,1} \rightarrow \text{low}$$

Group Sparsity and Group Lasso

$$\ell_{2,1}\text{-norm } \|W\|_{2,1} = \sum_{i=1}^{d} \|[W]^i\|_2$$





$$\text{high} \leftarrow \|W\|_{2,1} \rightarrow \text{low}$$

$$\widehat{W}_a^k = \arg\min_{W_a} \sum_{t=1}^{T} \left\| Z_{a,t}^k - \Phi_t w_{a,t} \right\|_2^2 + \lambda \|W_a\|_{2,1} .$$

Group Sparsity

Assumption

We assume that the joint useful features across all the tasks are such that $|J| = \tilde{s} \ll d$.

GL-FQI, Theoretical Guarantees

Theorem (GL–FQI)

If GL–FQI is run jointly on all T tasks for K iterations for any numerical constant $\delta > 0$, then with probability at least $(1 - \log(d)^{-\delta})^K$, the performance loss is bounded as

$$\frac{1}{T}\sum_{t=1}^{T}\left\|Q_t^* - Q_t^{\pi_t^K}\right\|_{2,\mu}^2$$

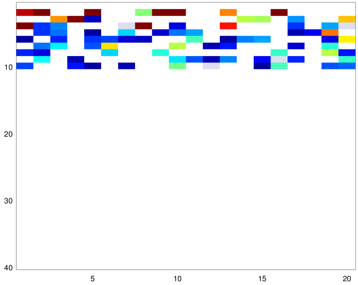
$$\leq \mathcal{O}\left(\frac{1}{(1-\gamma)^4}\left[\frac{L^2 Q_{\max}^2}{\kappa^4(2\tilde{s})}\frac{\tilde{s}}{n}\left(1 + \frac{(\log d)^{3/2+\delta}}{\sqrt{T}}\right) + \gamma^K Q_{\max}^2\right]\right).$$

Multi-Task vs Single Task: pros and cons

$$\text{GL–FQI}: \; \widetilde{\mathcal{O}}\Big(\frac{\tilde{s}}{n}\big(1+\frac{\log d}{\sqrt{T}}\big)\Big), \quad \text{LASSO-FQI}: \; \widetilde{\mathcal{O}}\Big(\frac{\bar{s}\log d}{n}\Big), \bar{s} = \frac{\sum_t s_t}{T}$$

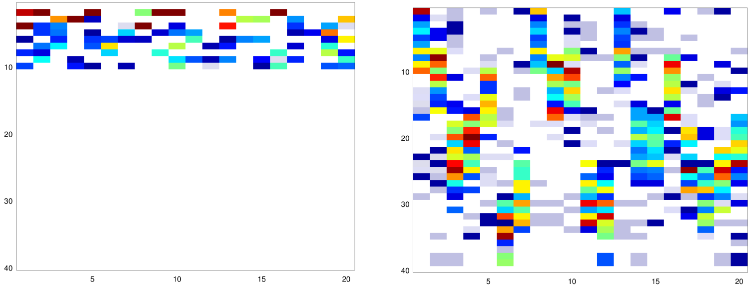
Multi-Task vs Single Task: pros and cons

$$\text{GL–FQI}: \ \widetilde{\mathcal{O}}\Big(\frac{\tilde{s}}{n}\big(1+\frac{\log d}{\sqrt{T}}\big)\Big), \quad \text{LASSO-FQI}: \ \widetilde{\mathcal{O}}\Big(\frac{\bar{s}\log d}{n}\Big), \bar{s} = \frac{\sum_t s_t}{T}$$



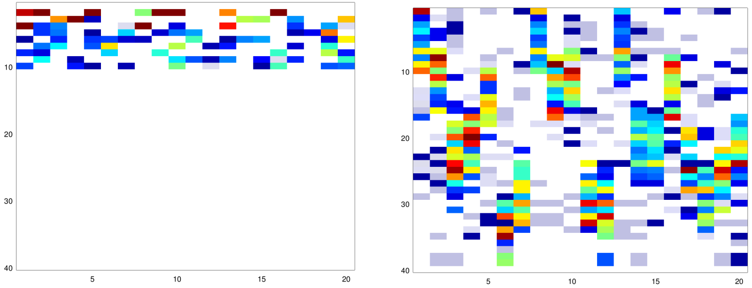
Multi-Task vs Single Task: pros and cons

$$\text{GL–FQI}: \; \widetilde{\mathcal{O}}\Big(\frac{\tilde{s}}{n}\big(1 + \frac{\log d}{\sqrt{T}}\big)\Big), \quad \text{LASSO-FQI}: \; \widetilde{\mathcal{O}}\Big(\frac{\bar{s}\log d}{n}\Big), \bar{s} = \frac{\sum_t s_t}{T}$$



Multi-Task vs Single Task: pros and cons

$$\text{GL–FQI}: \ \widetilde{\mathcal{O}}\Big(\frac{\tilde{s}}{n}\big(1 + \frac{\log d}{\sqrt{T}}\big)\Big), \quad \text{LASSO-FQI}: \ \widetilde{\mathcal{O}}\Big(\frac{\bar{s}\log d}{n}\Big), \bar{s} = \frac{\sum_t s_t}{T}$$



same $\bar{s}$, different $\tilde{s}$

Talk Overview

- ~~Reinforcement Learning $\rightarrow$ Linear Fitted Q Iteration (LinFQI)~~
- ~~Sparse Markov Decision Process $\rightarrow$ LASSO FQI~~
- ~~Multi-Task (Group) Sparsity $\rightarrow$ Group-LASSO FQI~~
- Learning Sparse Representations $\rightarrow$ Feature Learning FQI
- Experiments

Sparsity and Representation

Properties and change of feature representation:

Sparsity and Representation

Properties and change of feature representation:

- $\surd$ Bounded

Sparsity and Representation

Properties and change of feature representation:

- $\surd$ Bounded
- $\surd$ Smooth

Sparsity and Representation

Properties and change of feature representation:

- √ Bounded
- √ Smooth
- √ Measurable

Sparsity and Representation

Properties and change of feature representation:

- √ Bounded
- √ Smooth
- √ Measurable
- X Sparse

Sparsity and Representation

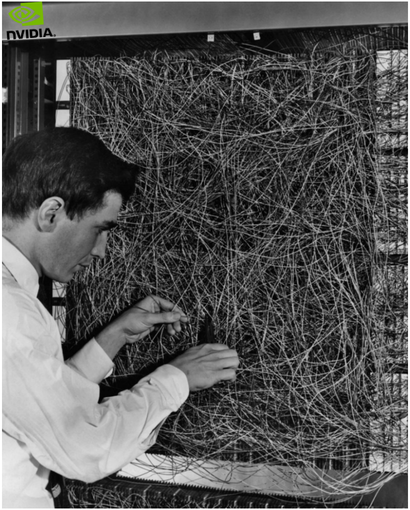
Properties and change of feature representation:

- $\surd$ Bounded
- $\surd$ Smooth
- $\surd$ Measurable
- **X** Sparse

Idea: learn a series of transformations to recover sparsity.

Deep Sparse Fitted Value Iteration

Deep Learning is recurrent in Machine Learning.



Multi-Task Feature Learning

Deep Learning: extremely powerful

Multi-Task Feature Learning

Deep Learning: extremely powerful
So powerful one layer is enough

Multi-Task Feature Learning

Deep Learning: extremely powerful
So powerful one layer is enough
even a linear layer

Multi-Task Feature Learning

Deep Learning: extremely powerful
So powerful one layer is enough
even a linear layer
Single Level Deep Linear Feature Learning: can a linear transformation do something useful?

$$(\widehat{U}_a^k, \widehat{A}_a^k) = \arg \min_{U_a \in \boldsymbol{O}^d} \min_{A_a \in \mathbb{R}^{d \times T}} \sum_{t=1}^{T} ||Z_{a,t}^k - \Phi_t U_a [A_a]_t||^2 + \lambda \|A\|_{2,1} .$$

Multi-Task Feature Learning: an interpretation

Proposition

Given $A, W \in \mathbb{R}^{d \times T}$, $U \in \boldsymbol{O}^d$, the following equality holds

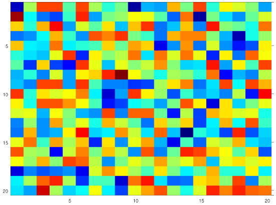
$$\min_{A,U} \sum_{t=1}^{T} ||Z^k_{a,t} - \Phi_t U_a [A_a]_t||^2 + \lambda \|A\|_{2,1}$$

$$= \min_{W} \sum_{t=1}^{T} ||Z^k_{a,t} - \Phi_t [W_a]_t||^2 + \lambda \|W\|_1.$$

The relationship between the optimal solutions is $W^* = UA^*$.

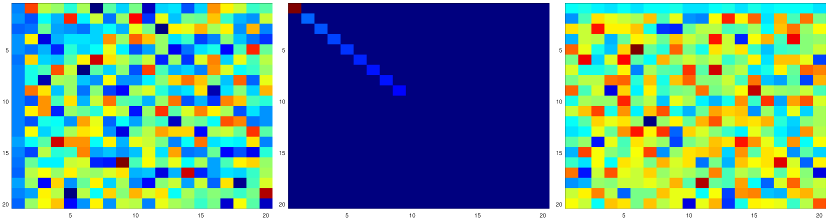
Low Rank and Trace Norm

$$\|W\|_1 = \sum \sigma(W)$$



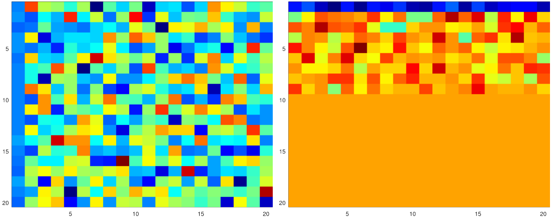
Low Rank and Trace Norm

$$\|W\|_1 = \sum \sigma(W)$$



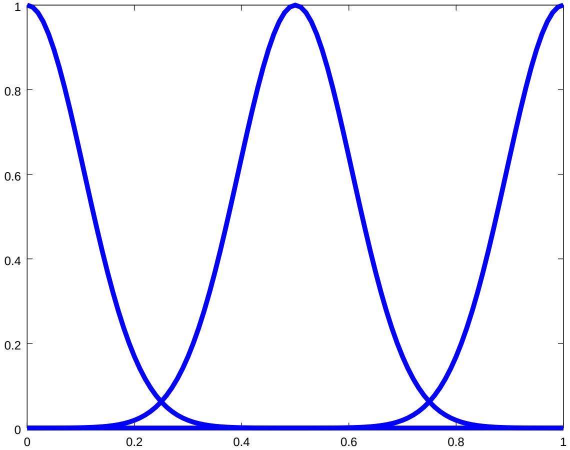
Low Rank and Trace Norm

$$\|W\|_1 = \sum \sigma(W)$$



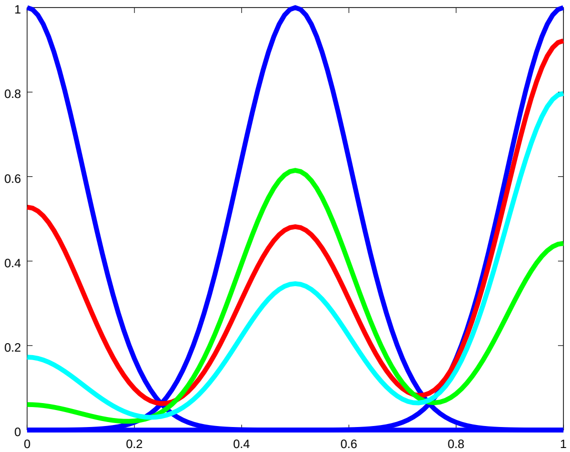
Low Rank and Task correlation

Dictionary of base tasks



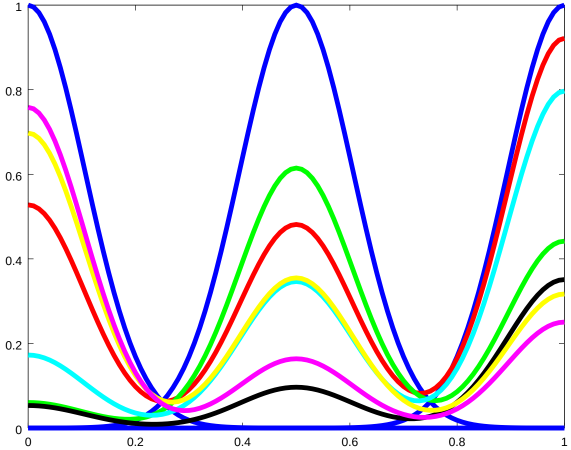
Low Rank and Task correlation

Dictionary of base tasks



Low Rank and Task correlation

Dictionary of base tasks



Low Rank Assumption

Assumption

There exists an orthogonal matrix $U \in \boldsymbol{O}^d$ *such that the weight matrix* A^* *obtained as a transformation of* W^* *(i.e.,* $A^* = U^{-1}W^*$*) is jointly sparse, i.e., has a set of "useful" features* $J(A^*) = \cup_{t=1}^{T} J([A^*]_t)$ *with* $|J(A^*)| = s^* \ll d$*.*

FL-FQI, Theoretical Guarantee

Theorem (FL–FQI)

*Let $T > \mathcal{O}(\log n)$. If FL–FQI (Algorithm **??** with Eq. **??**) is run jointly on all T tasks for K iterations with a regularizer*

$$\lambda \geq 2LQ_{\max}\sqrt{\frac{d+T}{n}},$$

then there exist constants c_1 and c_2 such that with probability at least $(1 - c_1 \exp\{-c_2(d+T)\})^K$, the performance loss is bounded as

$$\frac{1}{T}\sum_{t=1}^{T}\left\|Q_t^* - Q_t^{\pi_t^K}\right\|_{2,\rho}^2 \leq \mathcal{O}\left(\frac{1}{(1-\gamma)^4}\left[\frac{Q_{\max}^2 L^4}{\kappa^2}\frac{s^*}{n}\left(1+\frac{d}{T}\right)+\gamma^K Q_{\max}^2\right]\right).$$

Different sparsities, a comparison

$$\text{LASSO-}\mathrm{F}Q\mathrm{I}\text{:}\ \ \tilde{O}\left(\frac{\bar{s}\log(d)}{n}\right);$$

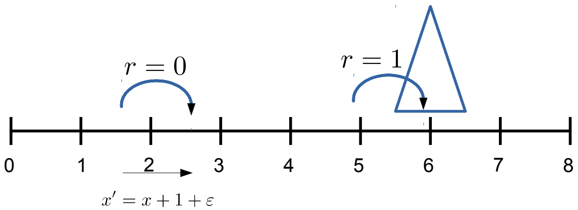
$$\mathrm{GL–F}Q\mathrm{I}\text{:}\ \ \tilde{O}\left(\frac{\tilde{s}}{n}\left(1+\frac{\log(d)}{\sqrt{T}}\right)\right);$$

$$\text{FL-}\mathrm{F}Q\mathrm{I}\text{:}\ \ \tilde{O}\left(\frac{s^*}{n}\left(1+\frac{d}{T}\right)\right),$$

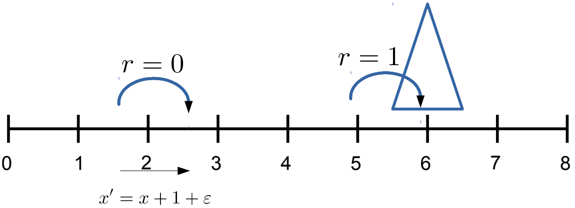
Talk Overview

- ~~Reinforcement Learning $\rightarrow$ Linear Fitted Q Iteration (LinFQI)~~
- ~~Sparse Markov Decision Process $\rightarrow$ LASSO FQI~~
- ~~Multi-Task (Group) Sparsity $\rightarrow$ Group-LASSO FQI~~
- ~~Learning Sparse Representations $\rightarrow$ Feature Learning FQI~~
- Experiments

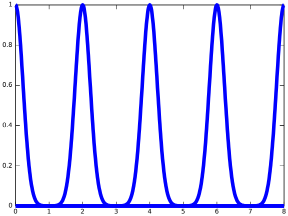
Experiments: Chain Walk



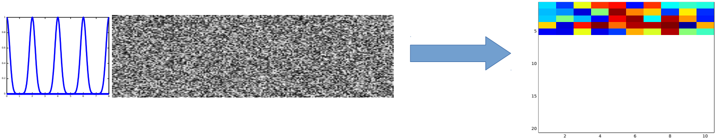
Experiments: Chain Walk

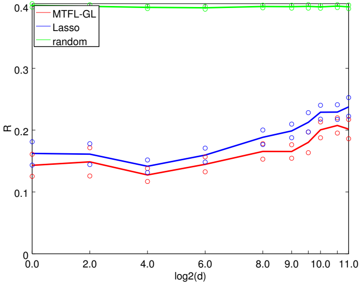


$s = 17$
$x[1, 17] = \mathrm{RBF}$
$x[18, d] = \mathcal{U}(-0.25, 0.25)$
$\log(d) = \{2, 4, 6, 8, 10, 11\}$

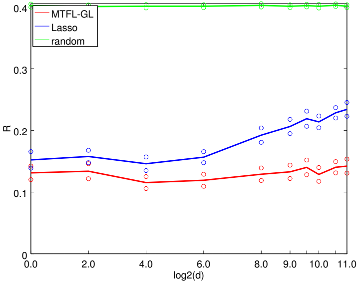


Experiments: Chain Walk



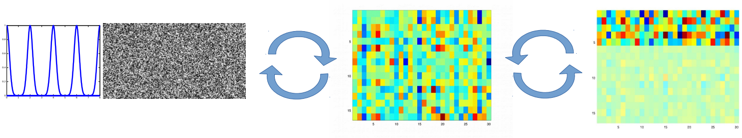


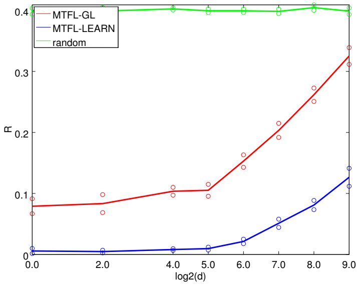
T = 3



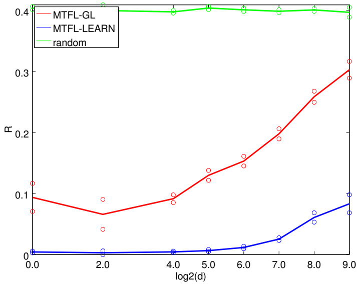
T = 10

Experiments: Chain Walk





T = 100



T = 200

Experiments: BlackJack

Rules:

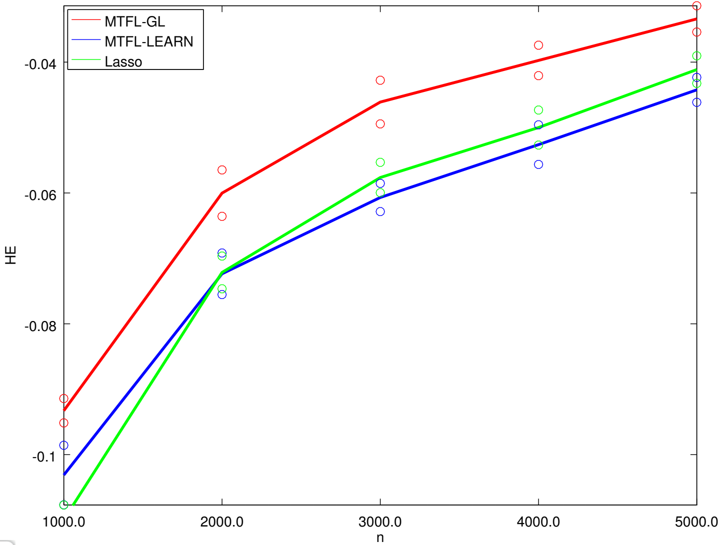
- ♣ Aces $= \{1, 11\}$, $2, \ldots, 9 = [2, \ldots, 9]$, 10,J,Q,K $= 10$.
- ♢ Player can ask another card "HIT" or "STAY"
- ♡ If the player goes over 21, he loses, end of game
- ♠ Dealer has to "HIT" until a threshold, then "STAY"
- ♣ If the dealer goes over 21, player wins.
- ♢ If the player has a strictly higher score than the dealer, player wins

How to make multiple tasks:

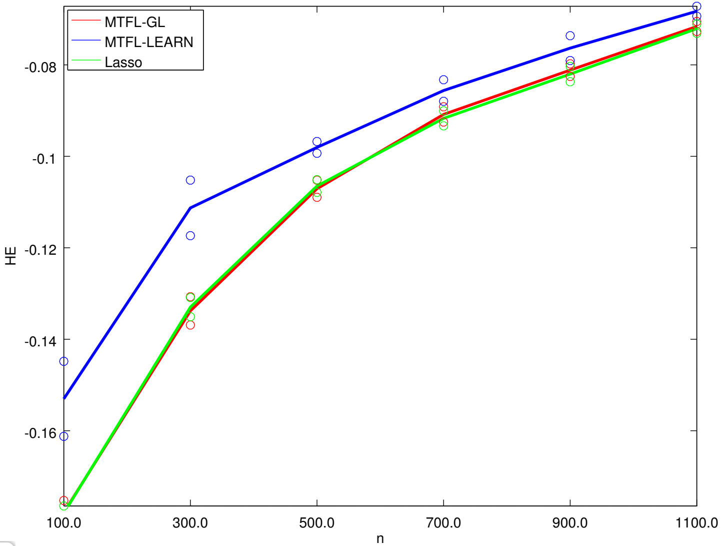
- ♡ Dealer threshold $\{15, 16, 17, 18\}$
- ♠ Number of decks $\{2, 4, 6, 8\}$
- ♣ If the dealer "HIT" when has a soft ace (A=11)

Two variants: Player can "DOUBLE" his bet after seeing the first two card, he receives a card and "STAY"

Experiments: BlackJack



Experiments: BlackJack



Conclusion

Multi-Task and Sparsity formalized for MDPs

Theoretical guarantees for new algorithms

Experimental validation shows assumptions are reasonable

Future Works

Weird regression techniques: have we tried everything?

- More sparse: Sparse Group LASSO, Graph LASSO
- More Learning: Inter-Task Feature Learning

Still, we need gurantees on the reconstruction.

Negative transfer: can we find a method that improves every single task?

Thanks for the attention

Questions?

Restricted Eigenvalues Assumption

Assumption (Restricted Eigenvalues)

For any $s \in [d]$, there exists $\kappa(s) \in \mathbb{R}^+$ such that:

$$\min \left\{ \frac{\|\Phi\Delta\|_2}{\sqrt{n}\,\|\Delta_J\|_2} : |J| \leq s, \Delta \in \mathbb{R}^d \backslash \{\mathbf{0}\}, \|\Delta_{J^c}\|_1 \leq 3\,\|\Delta_J\|_1 \right\} \geq \kappa(s), \tag{2}$$

where n is the number of samples, and J^c denotes the complement of the set of indices J.

Multi-Task Restricted Eigenvalues

Assumption (Multi-Task Restricted Eigenvalues)

For any $s \in [d]$, there exists $\kappa(s) \in \mathbb{R}^+$ such that:

$$\min\left\{\frac{\|\Phi\,\mathrm{Vec}(\Delta)\|_2}{\sqrt{n}\,\|\mathrm{Vec}(\Delta_J)\|_2} : |J| \leq s, \Delta \in \mathbb{R}^{d\times T}\backslash\{\mathbf{0}\}, \|\Delta_{J^c}\|_{2,1} \leq 3\,\|\Delta_J\|_{2,1}\right\} \geq \kappa(s), \tag{3}$$

where n is the number of samples, J^c denotes the complement of the set of indices J, and Φ indicates the block diagonal matrix composed by the union of the T sample matrices Φ_t.

Restricted Strong Convexity

Assumption (Restricted Strong Convexity)

Under Assumption 4, let $W^* = UDV^\mathsf{T}$ *be a singular value decomposition of the optimal matrix* W^* *of rank* r*, and* U^r, V^r *the submatrices associated with the top* r *singular values. Define* $\mathcal{B} = \{\Delta \in \mathbb{R}^{d\times T} : \mathrm{Row}(\Delta) \perp U^r \text{ and } \mathrm{Col}(\Delta) \perp V^r\}$*, and the projection operator onto this set* $\Pi_\mathcal{B}$*. There exists a positive constant* κ *such that*

$$\min\left\{\frac{\|\Phi\,\mathrm{Vec}(\Delta)\|_2^2}{2nT\|\,\mathrm{Vec}(\Delta)\|_2^2} : \Delta \in \mathbb{R}^{d\times T}, \|\Pi_\mathcal{B}(\Delta)\|_1 \leq 3\|\Delta - \Pi_\mathcal{B}(\Delta)\|_1\right\} \geq \kappa \tag{4}$$

Regression with unbiased samples

$$z_{i,a}^k = r_{i,a}^k + \gamma \max_{a'} \widehat{Q}^{k-1}(y_{i,a}^k, a') = r_{i,a}^k + \gamma \max_{a'} \phi(y_{i,a}^k)^\mathsf{T} w_{a',t}^{k-1} \qquad (5)$$

$$z_{i,a}^k = \mathcal{T}\widehat{Q}^{k-1}(x_i, a) + \eta_{i,a}^k, \qquad (6)$$

$\eta_{i,a}^k$ is random due to the choice of samples, but is 0 mean and bounded by $[-Q_{\max}, Q_{\max}]$ (truncation).
Resampling $y_{i,a}^k, r_{i,a}^k$ keeps the iterations i.i.d.

$$W_a^K \in \mathbb{R}^{d \times T},$$

$$\Phi = [\phi(x_1)^\mathsf{T}; \cdots ; \phi(x_{n_x})^\mathsf{T}] \in \mathbb{R}^{n_x \times d},$$

$$\Phi_a'^k = [\phi(y_{i,a}^k)^\mathsf{T}; \cdots ; \phi(y_{n_x,a}^k)^\mathsf{T}] \in \mathbb{R}^{n_x \times d},$$

$$R_a^k = [r_{1,a}^k, \ldots, r_{n_x,a}^k] \in \mathbb{R}^{n_x},$$

and the vector $Z_a^k = [z_{1,a}^k, \ldots, z_{n_x,a}^k] \in \mathbb{R}^{n_x}$ obtained as

$$Z_a^k = R_a^k + \gamma \max_{a'}(\Phi_{a'}'^k w_{a'}^{k-1}).$$

LASSO-FQI, Under the hood

$$|Q_t^* - Q_t^{\pi_t^K}| \leq \frac{2\gamma(1-\gamma^{K+1})}{(1-\gamma)^2}\left[\sum_{k=0}^{K-1} \alpha_k A_{tk}|\varepsilon_t^k| + \alpha_K A_{tK}|Q_t^* - Q_t^0|\right],$$

$$\begin{aligned}|\varepsilon_t^k(y,b)| &= |f_{w_t^k}(y,b) - f_{\widehat{w}_t^k}(y,b)| = |\phi(y)^\mathsf{T} w_{b,t}^k - \phi(y)^\mathsf{T}\widehat{w}_{b,t}^k| \\ &\leq ||\phi(y)||_2 ||w_{b,t}^k - \widehat{w}_{b,t}^k||_2 \leq L||w_{b,t}^k - \widehat{w}_{b,t}^k||_2,\end{aligned}$$

$$\left\|w_{a,t}^k - \widehat{w}_{a,t}^k\right\|_2^2 \leq \frac{256\delta^2 Q_{\max}^2}{\kappa^4(s_t^k)}\frac{s_t^k \log d}{n}. \quad (7)$$

GL-FQI, Under the hood

Proposition ([?])

For any action $a \in \mathcal{A}$ and any iteration $k < K$, let W_a^k be sparse such that $|J(W_a^k)| \leq \tilde{s}^k$ and satisfy Assumption 6 with $\kappa_t^k = \kappa(2s_t^k)$. Then if Eq. **??** *is run with a regularizer*

$$\lambda = \frac{LQ_{\max}}{\sqrt{nT}} \left(1 + \frac{(\log d)^{\frac{3}{2}+\delta}}{\sqrt{T}}\right)^{\frac{1}{2}},$$

for any numerical constant $\delta > 0$, then with probability at least $1 - \log(d)^{-\delta}$, the function $f_{\widehat{w}_{a,t}^k}$ computed in Eq. **??** *has an error bounded as*

$$\frac{1}{T}\sum_{t=1}^{T} \left\| [W_a^k]_t - [\widehat{W}_a^k]_t \right\|_2^2 = \frac{1}{T}\left\| W^k - \widehat{W}^k \right\|_2^2 \leq \frac{160 L^2 Q_{\max}^2}{\kappa_{Td}^4(2\tilde{s})} \frac{\tilde{s}}{n} \left(1 + \frac{(\log d)^{3/2+\delta}}{\sqrt{T}} \right. \tag{8}$$

Previous approaches

Lasso-TD (Single-Task)

$$\|V^* - V^{\pi_K}\|_\rho \leq \frac{2}{(1-\gamma)^2}$$

$$\left[\gamma^{K/2}\|V^* - V_0\|_\infty + C\left[c_1\lambda \max_{k=1,\dots,K}\|T_k^* Q_0\|_{\mathcal{K}}^2 + \frac{c_2 V_{max}^4}{n\lambda^{d/l}} + \frac{c_3\log(1/\delta)}{nV_{max}^4}\right]^2\right].$$

Kernel-FQI (Non-Parametric, Non-sparse)

$$\|V^\pi - \Phi\widehat{w}\|_n \leq$$

$$\frac{1}{1-\gamma}\inf_u\left[\|V^\pi - \Phi u\|_n + \frac{12\gamma V_{max} L\sqrt{s}}{\psi}\left(\sqrt{\frac{2\log(2d/\delta)}{n}} + \frac{1}{2n}\right)\right]$$