5.22 57 19 7.75 1.58 126
2. How to choose the ~y
V A Markov reward process over state space
OLS-LSTD are made.

Our goal is to be able to estimate how “good”
any 
V
s
sured in terms of the expected accumulated reward that we
velop according to
x
p
i
γ
is a discount factor, and which obeys the linear
fixed point equation (2) weighted by the stationary

equences: the first
k
is associated
with the state. Now there are two questions:

The second question raises a conceptual issue and is more
difficult to answer. The problem one faces is one of
feature selection.

Basic forward selection
Assume that at step k, \( \hat{V}_k = \sum_{i=1}^{k} \sum_{x} \hat{V}(x) \) and let \( \hat{V}_k \) be the projection onto the orthogonal complement space (in \( \mathbb{R}^3 \)). Each step of the forward selection now performs the following operations:

1. Find index \( i \in \{k+1, \ldots, M\} \) which maximizes reduction of the Bellman residual in the current LSTD solution \( \alpha_i \) for \( \hat{V}_k \):

\[
\begin{align*}
\hat{V} \left( x, w \right) & = \sum_{i=1}^{N} \alpha_i \phi_i(x) \\
\text{orthogonal representation of unselected elements} & \\
\text{To efficiently determine the novel contribution for each un-} & \\
\text{selected basis vector } \alpha_i \text{ in Step 1, we store } \phi_i^* & \\
\text{The next best element to add is then simply} & \\
\text{Add: } & \\
\hat{V} \left( x, w \right) & = P_{\hat{V}} \left( \hat{V} + \sum_{i=1}^{N} \phi_i^* w_i \right) \\
\text{Orthogonal representation of unselected elements} & \\
\text{Whenever an unselected basis vector } \alpha_i \text{ is selected in Step } 2, \text{ the remaining } \phi_i^* \text{ need to be reorthogonalized with} & \\
\text{respect to the new } \hat{V}_{k+1} = \hat{V}_k + \phi_i^* \alpha_i. & \\
\text{Biorthogonal representation of selected elements} & \\
\text{Each selected basis vector } \alpha_i \text{ spanning } V \text{ is associated with a} & \\
\text{biorthogonal basis vector } \beta_i \text{ with the property} & \\
\left( \phi_i^*, \beta_i \right) = \delta_{ij} \text{ for } i = 1, \ldots, k. \text{ The } \beta_i \text{ span the same space } V \\
\text{and are chosen such they represent the projection} & \\
\text{onto } V \text{ in terms of the original (non-orthogonalized) basis} & \\
\text{vectors } \alpha_i \text{, i.e.,}& \\
\text{with a biorthogonal basis representation, } P_{\hat{V}} \text{ can be easily} & \\
\text{updated in both directions } P_{\hat{V}\alpha_i} \text{ (adding an element)} & \\
\text{and } P_{\hat{V}\alpha_i^*} \text{ (deleting an element)} \text{ [3].} & \\
\text{Methods compared} & \\
\text{• OLS-F(B(k)): select } \alpha_i \text{ basis vectors via forward selec-} & \\
\text{• OLS-FB2(k)): select } 2 \times \alpha_i \text{ basis vectors via forward} & \\
\text{• OLS-FB2(3)): add } \theta \text{ to regularization (3). At each} & \\
\text{step, either add or delete a basis vector until no further improve-} & \\
\text{ment is possible.} & \\
\text{• MP-F(\alpha_i^k)): implemented as described in [1].} & \\
\text{• LARS(\alpha_i^k)): implemented as described in [2].} & \\

\begin{align*}
\text{Mountain car} & \\
\text{Stats: } & \\
\text{2504 training samples, 7513 test samples, 1365 ba-} & \\
\text{sis function candidates (RBFS on a grid at various levels of} & \\
\text{coarseness)} & \\
\text{Related work} & \\
\text{[1] C. Dimaio-Wakefield and R. Parr. Graphy algorithms for sparse reinforce-} & \\
\text{[2] J. Zein and Ribeiro-Neira. Ratio and feature selection in least} & \\
\text{onal matching pursuit approach. IEEE Signal Processing Letter, 11:96-100, 2004.} & \\
\text{Biorthogonalization Techniques for Least Squares Temporal Difference Learning} & \\
Tobias Jung & \\
Montefiore Institute, University of Liège, Belgium & \\
etmail@ulg.ac.be &