An overview of ℓ_1-regularization for value function approximation

Journée GDR Robotique

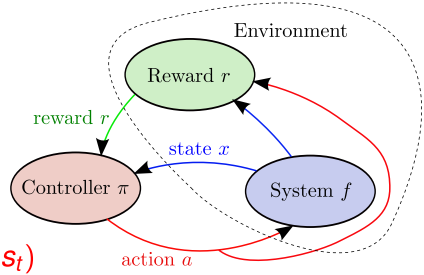
Matthieu Geist
(Supélec, IMS-MaLIS Research Group)
matthieu.geist@supelec.fr

September the 6th, 2012

Supélec

1. Introduction
   - MDP
   - Setting
2. Background
   - LSTD
   - ℓ_1-regularization
3. Algorithms
   - Lasso-TD
   - ℓ_1-PBR
   - ℓ_1-LSTD
   - Dantzig LSTD
4. Summary

(Bertsekas & Tsitsiklis, 1996; Sutton & Barto, 1998)



- Controller : $a_t = \pi(s_t)$
- Reward : $r_t = r(s_t, a_t)$
- System dynamics : $s_{t+1} \sim P(.|s_t, a_t)$

- **Goal :** Given a policy $\pi : S \to A$ compute its value

$$v^\pi(s) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, \pi(s_t)) \middle| s_0 = s, \pi\right] \quad (0 < \gamma < 1)$$

- The value function v satisfies the Bellman equation

$$v = r + \gamma P v \quad \Leftrightarrow \quad v = \mathcal{T} v$$

- We look for $\hat{v}(s) = \sum_{j=1}^{p} \theta_j \phi_j(s)$ or $\hat{v} = \Phi\theta$, where

$$\Phi = \begin{pmatrix} \phi(s_1)^\top \\ \vdots \\ \phi(s_{|S|})^\top \end{pmatrix} = \begin{pmatrix} \phi_1 & \dots & \phi_p \end{pmatrix} \text{ and } \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_p \end{pmatrix}$$

- $\mathcal{T}$ is only *known* through samples $(s_i, r_i, s'_i)_{i=1}^{n}$ where $s_i \sim \mu$:

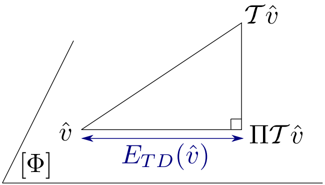
$$\tilde{\Phi} = \begin{pmatrix} \phi(s_1)^\top \\ \vdots \\ \phi(s_n)^\top \end{pmatrix}, \tilde{\Phi}' = \begin{pmatrix} \phi(s'_1)^\top \\ \vdots \\ \phi(s'_n)^\top \end{pmatrix}, \tilde{r} = \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix},$$

- We consider the situation where $n \ll p$

1. Introduction
   - MDP
   - Setting
2. Background
   - LSTD
   - ℓ_1-regularization
3. Algorithms
   - Lasso-TD
   - ℓ_1-PBR
   - ℓ_1-LSTD
   - Dantzig LSTD
4. Summary

LSTD (Bradtke & Barto, 1996)

One looks for $\hat{v}$ in the feature space satisfying $\hat{v} = \Pi_\mu \mathcal{T} \hat{v}$.



The solution $\hat{v} = \Phi\theta_0$ can be characterized as

$$\begin{cases} \omega_\theta = \arg\min_\omega \|r + \gamma P\Phi\theta - \Phi\omega\|_{\mu,2}^2 \\ \theta_0 = \arg\min_\theta \|\Phi\theta - \Phi\omega_\theta\|_{\mu,2}^2 \end{cases}$$

and approximated by its empirical counterpart :

$$\begin{cases} \omega_\theta = \arg\min_\omega \|\tilde{r} + \gamma\tilde{\Phi}'\theta - \tilde{\Phi}\omega\|_2^2 \\ \theta_0 = \arg\min_\theta \|\tilde{\Phi}\theta - \tilde{\Phi}\omega_\theta\|_2^2 \end{cases}$$

LSTD (alternate writing)

We look for $\hat{v} = \Phi\theta_0$ that solves $\hat{v} = \Pi_\mu \mathcal{T}\hat{v}$. It is known that $\Pi_\mu = \Phi(\Phi^\top D_\mu \Phi)^{-1}\Phi^\top D_\mu$, and (after algebra) one can show that

$\theta_0 = A^{-1}b$ (Linear system of size p) $\Leftrightarrow$ $\theta_0 = \arg\min_\theta \|A\theta - b\|_2^2$

where

$$A = \Phi^\top D_\mu (I - \gamma P)\Phi$$
$$b = \Phi^\top D_\mu r$$

can be approximated through their empirical counterpart :

$$\tilde{A} = \frac{1}{n}\tilde{\Phi}^\top(\tilde{\Phi} - \gamma\tilde{\Phi}')$$
$$\tilde{b} = \frac{1}{n}\tilde{\Phi}^\top\tilde{r}$$

LSTD (another alternate writing)

Recall the first writing :

$$\begin{cases} \omega_\theta = \arg\min_\omega \|\tilde{r} + \gamma\tilde{\Phi}'\theta - \tilde{\Phi}\omega\|_2^2 \\ \theta_0 = \arg\min_\theta \|\tilde{\Phi}\theta - \tilde{\Phi}\omega_\theta\|_2^2 \end{cases}.$$

The second equation is minimized to zero for $\theta = \omega_\theta$ (the fixed point of $\Pi_\mu \mathcal{T}$ exists).
Therefore, the second equation rewrites

$$\theta_0 = \arg\min_\omega \|\tilde{r} + \gamma\tilde{\Phi}'\theta_0 - \tilde{\Phi}\omega\|_2^2$$

(fixed-point equation, θ_0 in both sides).

Least-squares regression and ℓ_2-penalization

- regression ($\gamma = 0$), find θ such that $\tilde{r} \approx \tilde{\Phi}\theta$;
- natural (LS) objective function :

$$\theta_0 = \arg\min_{\theta} \|\tilde{r} - \tilde{\Phi}\theta\|_2^2$$

  - analytical solution : $\theta_0 = (\tilde{\Phi}^\top \tilde{\Phi})^{-1} \tilde{\Phi} \tilde{r}$
  - this is an empirical projection : $\tilde{\Phi}\theta_0 = \tilde{\Pi}_\mu \tilde{r}$, (so $\tilde{\Pi}_\mu = \tilde{\Phi}(\tilde{\Phi}^\top \tilde{\Phi})^{-1} \tilde{\Phi}^\top$)
- ℓ_2-penalized LS :

$$\theta_\lambda = \arg\min_{\theta} \|\tilde{r} - \tilde{\Phi}\theta\|_2^2 + \lambda \|\theta\|_2^2$$

  - analytical solution : $\theta_\lambda = (\tilde{\Phi}^\top \tilde{\Phi} + \lambda \mathbf{I})^{-1} \tilde{\Phi} \tilde{r}$
  - this is a penalized empirical projection : $\tilde{\Phi}\theta_\lambda = \tilde{\Pi}_\mu^{(\lambda,2)} \tilde{r}$

ℓ_1-penalization

- Lasso (Tibshirani, 1996) (ℓ_1-penalized LS) :

$$\theta_\lambda = \arg\min_\theta \|\tilde{r} - \tilde{\Phi}\theta\|_2^2 + \lambda\|\theta\|_1$$

  - promotes sparsity
  - no analytical solution, but efficient computation through the regularization path (θ_λ piecewise linear) (Efron *et al.* , 2004)
  - this is a penalized empirical projection : $\tilde{\Phi}\theta_\lambda = \tilde{\Pi}_\mu^{(\lambda,1)}\tilde{r}$
- Dantzig Selector (Candes & Tao, 2007) :

$$\theta_\lambda = \arg\min_\theta \|\theta\|_1 \text{ subject to } \|\tilde{\Phi}^\top(\tilde{r} - \tilde{\Phi}\theta)\|_\infty \leq \lambda$$

  - promotes sparsity
  - this is actually a linear program
- many (Lasso) extensions, but lets keep things simple...

1. Introduction
   - MDP
   - Setting
2. Background
   - LSTD
   - ℓ_1-regularization
3. Algorithms
   - Lasso-TD
   - ℓ_1-PBR
   - ℓ_1-LSTD
   - Dantzig LSTD
4. Summary

Algorithm

- LSTD (3^{rd} writing) :

$$\theta_0 = \arg\min_{\omega} \|\tilde{r} + \gamma\tilde{\Phi}'\theta_0 - \tilde{\Phi}\omega\|_2^2$$

- Lasso-TD (Kolter & Ng, 2009) :

$$\theta_\lambda = \arg\min_{\omega} \|\tilde{r} + \gamma\tilde{\Phi}'\theta_\lambda - \tilde{\Phi}\omega\|_2^2 + \lambda\|\omega\|_1$$

(not a standard lasso problem)

- alternative writing :

$$\begin{cases} \omega_\theta = \arg\min_\omega \|\tilde{r} + \gamma\tilde{\Phi}'\theta - \tilde{\Phi}\omega\|_2^2 + \lambda\|\omega\|_1 \\ \theta_\lambda = \arg\min_\theta \|\tilde{\Phi}\theta - \tilde{\Phi}\omega_\theta\|_2^2 \end{cases}$$

- another alternative writing : $\tilde{v}_\lambda = \tilde{\Pi}_\mu^{\lambda,1}\hat{\mathcal{T}}\tilde{v}_\lambda$ (with $\tilde{v}_\lambda = \tilde{\Phi}\theta_\lambda$)

Properties

- not a standard lasso $\Rightarrow$ originally requires an adhoc solver
- can be framed as an LCP (Linear Complementary Problem) (Johns *et al.* , 2010) : more generic solvers
- requires $\tilde{A}$ to be a P-matrix (not necessarily true in the off policy case : $\tilde{\Pi}_\mu^{\lambda,1}\hat{\mathcal{T}}$ may have zero or multiple fixed points)
- sparsity oracle inequality (Ghavamzadeh *et al.* , 2011) :

$$\inf_\lambda \|v - \tilde{v}_\lambda\|_n \leq \frac{1}{1-\gamma} \inf_\theta \left\{ \|v - \hat{v}_\theta\|_n + \mathcal{O}\left(\sqrt{\frac{\|\theta\|_0 \ln p}{n}}\right) \right\}$$

- remark : how to choose λ ? (no cross-validation !)

Algorithm (Geist & Scherrer, 2011; Hoffman *et al.* , 2011)

- lasso-td : $\begin{cases} \omega_\theta = \arg\min_\omega \|\tilde{r} + \gamma\tilde{\Phi}'\theta - \tilde{\Phi}\omega\|_2^2 + \lambda\|\omega\|_1 \\ \theta_\lambda = \arg\min_\theta \|\tilde{\Phi}\theta - \tilde{\Phi}\omega_\theta\|_2^2 \end{cases}$
- why not : $\begin{cases} \omega_\theta = \arg\min_\omega \|\tilde{r} + \gamma\tilde{\Phi}'\theta - \tilde{\Phi}\omega\|_2^2 \\ \theta_\lambda = \arg\min_\theta \|\tilde{\Phi}\theta - \tilde{\Phi}\omega_\theta\|_2^2 + \lambda\|\theta\|_1 \end{cases}$?

  (this is ℓ_1-PBR)
- alternative writing :

$$\theta_\lambda = \arg\min_\theta \|\tilde{\Pi}_\mu(\tilde{v}_\theta - \hat{\mathcal{T}}\tilde{v}_\theta)\|_2^2 + \lambda\|\theta\|_1 \quad \text{(recall } \tilde{v}_\theta = \tilde{\Phi}\theta\text{)}$$

(ℓ_1-penalized Projected Bellman Residual)

Properties

- ℓ_1-PBR is a standard Lasso problem :

  $$\theta_\lambda = \arg\min_\theta \|\tilde{y} - \tilde{\Psi}\theta\|_2^2 + \lambda\|\theta\|_1, \quad \tilde{y} = \tilde{\Pi}_\mu \tilde{r}, \quad \tilde{\Psi} = \tilde{\Phi} - \gamma\tilde{\Pi}_\mu\tilde{\Phi}'$$

  (easy extension to other penalization)
- weaker conditions than Lasso-TD (off-policy is fine)
- but huge computational cost if $p \gg n$ (projection in $O(p^3)$)
- no finite sample analysis

Algorithm

- 2nd formulation of LSTD :

$$\tilde{A}\theta_0 = \tilde{b}, \quad \tilde{A} = \frac{1}{n}\tilde{\Phi}^\top(\tilde{\Phi} - \gamma\tilde{\Phi}'), \quad \tilde{b} = \frac{1}{n}\tilde{\Phi}^\top\tilde{r}$$

- equivalently :

$$\theta_0 = \arg\min_\theta \|\tilde{A}\theta - \tilde{b}\|_2^2$$

- ℓ_1-LSTD (Pires, 2011; Pires & Szepesvári, 2012) :

$$\theta_\lambda = \arg\min_\theta \|\tilde{A}\theta - \tilde{b}\|_2^2 + \lambda\|\theta\|_1$$

Properties

- "standard" Lasso problem (use any solver)
- no problem in the off-policy case
- built-in (theoretical) λ-selection scheme
- finite sample analysis :

$$\inf_\lambda \|A\theta_\lambda - b\|_2 \leq \mathcal{O}\left(\|\theta^*\|_1 \sqrt{\frac{p^2}{n} \ln \frac{1}{\delta}}\right) \text{ w.p. } 1 - \delta;$$

(recall $A = \lim_{n\to\infty} \tilde{A}$ and $b = \lim_{n\to\infty} \tilde{b}$, $A\theta^* = b$)

Algorithm

- 2nd formulation of LSTD :

$$\theta_0 = \arg\min_\theta \|\tilde{A}\theta - \tilde{b}\|_2^2$$

- Dantzig-LSTD (Geist *et al.* , 2012) :

$$\theta_\lambda = \arg\min_\theta \|\theta\|_1 \text{ subject to } \|\tilde{A}\theta - \tilde{b}\|_\infty \leq \lambda.$$

- this is a linear program :

$$\min_{u,\theta\in\mathbb{R}^p} \mathbf{1}^\top u \quad \text{subject to} \begin{cases} -u \leq \theta \leq u \\ -\lambda\mathbf{1} \leq \tilde{A}\theta - \tilde{b} \leq \lambda\mathbf{1} \end{cases}$$

Properties

- standard Linear Program
- no problem in the off-policy case
- connexion to Lasso-TD :

$$\|\tilde{A}\theta_\lambda^{\text{lassoTD}} - \tilde{b}\|_\infty \leq \lambda.$$

- heuristic model selection scheme
- finite sample analysis :

$$\inf_\lambda \|A\theta_\lambda - b\|_\infty \leq \mathcal{O}\left(\|\theta^*\|_1 \sqrt{\frac{1}{n} \ln \frac{p}{\delta}}\right) \text{ w.p. } 1-\delta;$$

(recall $A = \lim_{n\to\infty} \tilde{A}$ and $b = \lim_{n\to\infty} \tilde{b}$, $A\theta^* = b$)

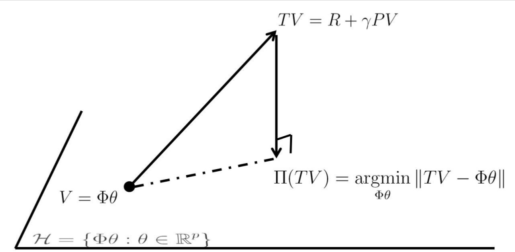
- empirically : Lasso-TD, D-LSTD $\geq$ ℓ_1-LSTD, ℓ_1-PBR

1 Introduction
- MDP
- Setting

2 Background
- LSTD
- ℓ_1-regularization

3 Algorithms
- Lasso-TD
- ℓ_1-PBR
- ℓ_1-LSTD
- Dantzig LSTD

4 Summary

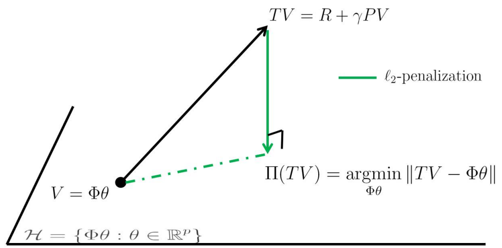


$$\omega_\theta = \arg\min_\omega \|\tilde{R} + \gamma\tilde{\Phi}'\theta - \tilde{\Phi}\omega\|_2^2$$

$$\theta_0 = \arg\min_\theta \|\tilde{\Phi}\theta - \tilde{\Phi}\omega_\theta\|_2$$

LSTD :

$$\hat{v}_0 = \tilde{\Pi}_\mu \hat{\mathcal{T}} \hat{v}_0$$

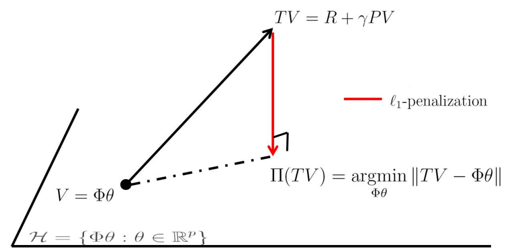


$$\omega_\theta = \arg\min_\omega \|\tilde{R} + \gamma\tilde{\Phi}'\theta - \tilde{\Phi}\omega\|_2^2 + \lambda_1\|\omega\|_2^2$$

$$\theta_{\lambda_1,\lambda_2} = \arg\min_\theta \|\tilde{\Phi}\theta - \tilde{\Phi}\omega_\theta\|_2 + \lambda_2\|\theta\|_2^2$$

$\ell_{2,2}$-LSTD (Farahmand *et al.* , 2008) :

$$\theta_{\lambda_{1,2}} = \arg\min \|\hat{v}_{\theta_{\lambda_{1,2}}} - \tilde{\Pi}_\mu^{\lambda_1,2}\hat{\mathcal{T}}\hat{v}_{\theta_{\lambda_{1,2}}}\|_2^2 + \lambda_2\|\theta\|_2^2$$

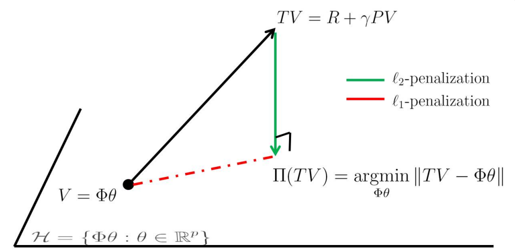


$$\omega_\theta = \arg\min_\omega \|\tilde{R} + \gamma\tilde{\Phi}'\theta - \tilde{\Phi}\omega\|_2^2 + \lambda_1\|\omega\|_1$$

$$\theta_{\lambda_1,\lambda_2} = \arg\min_\theta \|\tilde{\Phi}\theta - \tilde{\Phi}\omega_\theta\|_2$$

Lasso-TD :

$$v_{\theta_{\lambda_1}} = \tilde{\Pi}_\mu^{\lambda_1,1}\hat{\mathcal{T}} v_{\theta_{\lambda_1}}$$



$$\omega_\theta = \arg\min_\omega \|\tilde{R} + \gamma\tilde{\Phi}'\theta - \tilde{\Phi}\omega\|_2^2 + \lambda_1\|\omega\|_2^2$$

$$\theta_{\lambda_1,\lambda_2} = \arg\min_\theta \|\tilde{\Phi}\theta - \tilde{\Phi}\omega_\theta\|_2 + \lambda_2\|\theta\|_1$$

ℓ_1-PBR :

$$\theta_{\lambda_{1,2}} = \arg\min \|\hat{v}_{\theta_{\lambda_{1,2}}} - \tilde{\Pi}_\mu^{\lambda_1,2}\hat{\mathcal{T}}\hat{v}_{\theta_{\lambda_{1,2}}}\|_2^2 + \lambda_2\|\theta\|_1$$

based on the residual $\tilde{A}\theta - \tilde{b}$

- ℓ_1-LSTD :

$$\theta_\lambda = \operatorname*{arg\,min}_\theta \|\theta\|_1 \text{ subject to } \|\tilde{A}\theta - \tilde{b}\|_2^2 \leq \lambda$$

- Dantzig-LSTD :

$$\theta_\lambda = \operatorname*{arg\,min}_\theta \|\theta\|_1 \text{ subject to } \|\tilde{A}\theta - \tilde{b}\|_\infty \leq \lambda$$

Thank you for your attention !

Questions ?

Final Remark. This talk focused on projected-fixed-point-based regularization, but it is also possible to consider residual approaches, with a potential bias problem. The first (as far as I know) algorithm to use ℓ_1-penalization for value function estimation is actually a (biased) residual approach (Loth *et al.* , 2007).

References I

Bertsekas, Dimitri P., & Tsitsiklis, John N. 1996.
Neuro-Dynamic Programming.
Athena Scientific.

Bradtke, S. J., & Barto, A. G. 1996.
Linear Least-Squares algorithms for temporal difference learning.
Machine Learning, **22**, 33–57.

Candes, E., & Tao, T. 2007.
The Dantzig selector : statistical estimation when p is much larger than n.
Annals of Statistics, **35**(6), 2313–2351.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. 2004.
Least Angle Regression.
Annals of Statistics, **32**(2), 407–499.

Farahmand, A., Ghavamzadeh, M., Szepesvári, C., & Mannor, S. 2008.
Regularized Policy Iteration.
In : Proc. of NIPS 21.

Geist, M., & Scherrer, B. 2011.
ℓ_1-penalized projected Bellman residual.
In : Proc. of EWRL 9.

Geist, Matthieu, Scherrer, Bruno, Lazaric, Alessandro, & Ghavamzadeh, Mohammad. 2012.
A Dantzig Selector Approach to Temporal Difference Learning.
In : International Conference on Machine Learning (ICML).

References II

Ghavamzadeh, M., Lazaric, A., Munos, R., & Hoffman, M. 2011.
Finite-Sample Analysis of Lasso-TD.
In : Proc. of ICML.

Hoffman, M. W., Lazaric, A., Ghavamzadeh, M., & Munos, R. 2011.
Regularized Least Squares Temporal Difference learning with nested ℓ_2 and ℓ_1 penalization.
In : Proc. of EWRL 9.

Johns, J., Painter-Wakefield, C., & Parr, R. 2010.
Linear Complementarity for Regularized Policy Evaluation and Improvement.
In : Proc. of NIPS 23.

Kolter, J. Z., & Ng, A. Y. 2009.
Regularization and Feature Selection in Least-Squares Temporal Difference Learning.
In : Proc. of ICML.

Loth, Manuel, Davy, Manuel, & Preux, Philippe. 2007.
Sparse Temporal Difference Learning using LASSO.
In : IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning.

Pires, B. A. 2011.
Statistical analysis of ℓ_1-penalized linear estimation with applications.
M.Phil. thesis, University of Alberta.

Pires, Bernardo A., & Szepesvári, Csaba. 2012.
Statistical linear estimation with penalized estimators : an application to reinforcement learning.
In : Proceedings of the 29th International Conference on Machine Learning (ICML 2012).

References III

Sutton, R. S., & Barto, A. G. 1998.
Reinforcement Learning : an Introduction.
The MIT Press.

Tibshirani, R. 1996.
Regression Shrinkage and Selection via the Lasso.
Journal of the Royal Statistical Society, **58**(1), 267–288.