Algorithms for Reinforcement Learning

Errata for the printed book

Csaba Szepesvári

August 7, 2010*

Contents

Page numbers refer to the printed copy. The online version (the "draft") is up-to-date. Thanks to my PhD student, Gabor Bartok and Sotetsu Koyamada who have found many of these errors.

- p. xi. Section 2) should be Section 2 (no closing parenthesis)
- p. 1. The dot should be in between the bars in the definition of the infinity norm, not on the top. That is, $\|\cdot\|_\infty$ is the intended form and not $||\dot{|}|_\infty$. Also, in "which, if θ, which" the part "which, if θ" should be deleted.
- p.2. The footnote from p.5 explaining the meaning of "almost surely" should be moved here.
- p.5. In the example on gambling the personal pronoun "his" should be replaced by "her".
- p.9. In Eq. (1.14) on the right-hand side of the equation $Q(y, \pi(x))$ should be $Q(y, \pi(y))$.
- p.12. In footnote 1, add "if" before "it".
- p. 19. line 25: "gives the the sum of rewards along the trajectory" should be "gives the difference between the return along the trajectory and the value estimate at the start state".

*Last update: June 25, 2018

- p. 21. line 1: "then" should be "than".
- p. 22. The text "goal is to approximate the value function V underlying $\mathcal{M}$" should be deleted.
- p. 23. Delete "be" from "is no longer be guaranteed". After $\theta^{(\lambda)}$ in the middle of page delete ".". The phrase "using V_θ" should be " using the chosen features φ".
- p. 24. line 19: "full control learning task" should be "full, control-learning task"
- p. 25. "some methods using which" should be "some methods that avoid" On the same page, in the figure caption, "true" should be "two". Also, on the figure the labels $L_n(\theta)$ and $L(\theta)$ are swapped.
- p. 29. "this corresponds to starting with a diagonal matrix in RLSTD" should be "this corresponds to starting with a positive diagonal matrix in RLSTD".
- p. 32. The word "complicate" (in the middle of page) should be "complicated".
- p. 34. "has full column rank" should be "has full row rank". On the same page, in Eq. (2.17): Replace $\|\theta_*^\top \varphi - r\|^2$ with $\|\theta_*^\top \varphi - V\|_\mu^2$ on the right-hand side (although note that $V = r$).
- p. 40. The text "Gittins (1989) has shown" should be "Gittins (1989) showed".
- p. 43. The 4th displayed equation and the text surrounding it should be deleted. This is the equation that says that $\mathbf{R}_T^{\mathrm{UCRL2}(\delta)} = O(D^2|\mathcal{X}|^2|\mathcal{A}|\log(T/\delta)/\varepsilon + \varepsilon T)$. This equation holds (under the cited conditions), but it does not lead itself to a logarithmic regret bound.
- p. 43, line -5: Replace ". This happens" with "as happens to be the case when".
- p. 45, Algorithm 10 (UCRL2): Instead of "repeat-until" the appropriate programming construct should have been "while true-end while" (the body of the loop is repeated indefinitely) [Hill Ma].
- p. 46, Algorithm 11 (OptSolve): In the repeat-until construction the conditions must be flipped. Two places [Hill Ma].
- p. 47. On line 4 of the 1st paragraph of Section 3.3., "optimalas" should be "optimal as". On the same page, on line 3, after Eq. (3.1): ", Algorithm 12 the pseudocode of Q-learning." should start with a full stop and is missing the word "shows". So, the text should be ". Algorithm 12 shows the pseudocode of Q-learning.".

- p. 48. Section 3.2 is mentioned twice in the same sentence (around the middle of the page). The second occurrence should be deleted.
- p. 56, Algorithm 16, line 7. The correct update equation is $b \leftarrow b + R_{t+1} \cdot z$ [Tom Schaul, Idsia].
- p. 57, Algorithm 16 (LSPI): Flip the condition in the until construct [Hill Ma].
- p. 58. The definition of regret should be $\mathbf{R}_T^{\mathcal{A}} = T\rho^* - \mathcal{R}_T^{\mathcal{A}}$ [Hamid Reza Maei, Stanford].
- p. 59. "likelihood ratio methods Glynn, 1990" should be "likelihood ratio methods (Glynn, 1990)."
- p. 61, Algorithm 18: Again, instead of repeat-until, one needs while(true)-endwhile [Hill Ma].
- p. 65. The definition of norm is missing the so-called homogeneity condition: For any $\lambda \in \mathbb{R}$, $v \in V$, $f(\lambda v) = |\lambda| \, f(v)$. One the same page, in the bottom, "ℓ^∞ norms" should be "ℓ^∞ norm".
- p. 66. "uniformly bounded" should be "bounded" (when mentioning a single function).
- p. 66. line 8 from the bottom, "$f_n(x) \to 0$ for each x" should be "Define f so that $f(x) = 0$ if $x \neq 0$ and $f(0) = 1$. Then $f_n(x) \to f(x)$ for each x." Replace "However, $\|f_n - f\|_\infty = \|f_n\|_\infty = 1 \not\to 0$." with "However, $\|f_n - f\|_\infty = 1 \not\to 0$."
- p. 67. "Polish mathematicians" should be singular: "Polish mathematician".
- p. 68. On the top of page, "Assume that T is a γ-contraction." should go into the next line.
- p. 68. In the first displayed equation the last inequality should be removed. In the next *two* displayed equations, on the right-hand side, replace $\|v_k\|$ with $\|v_k - v_0\|$. Finally, in the 4th displayed equation on the page, remove $\|v_0\|$ from the right-hand side.
- p. 69. In the line preceding the definition of $B(\mathcal{X})$, "uniformly bounded" should be "bounded".
- p. 69. "It is easy to see that V^π": At the end of the display following (in the bottom of the page) replace V with V^π.
- p. 72. "for the final policy π, we have $TV^\pi = T^\pi V^\pi = V^\pi$." should be "for the final policy π, we have $T^*V^\pi = T^\pi V^\pi = V^\pi$.".

For further information, visit `http://www.ualberta.ca/~szepesva/RLBook.html`.