OPERATIONS RESEARCH
http://dx.doi.org/10.1287/opre.1120.1044ec

e-companion

ONLY AVAILABLE IN ELECTRONIC FORM

Electronic Companion—"Approximate Dynamic Programming via a Smoothed Linear Program" by Vijay V. Desai, Vivek F. Farias, and Ciamac C. Moallemi, *Operations Research*, http://dx.doi.org/10.1287/opre.1120.1044.

Approximate Dynamic Programming via a Smoothed Linear Program (Electronic Companion)

Vijay V. Desai
Industrial Engineering and Operations Research
Columbia University
email: vvd2101@columbia.edu

Vivek F. Farias
Sloan School of Management
Massachusetts Institute of Technology
email: vivekf@mit.edu

Ciamac C. Moallemi
Graduate School of Business
Columbia University
email: ciamac@gsb.columbia.edu

November 22, 2011

A. Proofs for Sections 4.2–4.4

Lemma 1. *For any $r \in \mathbb{R}^K$ and $\theta \geq 0$:*

(i) $\ell(r,\theta)$ is a finite-valued, decreasing, piecewise linear, convex function of θ.

(ii)

$$\ell(r,\theta) \leq \frac{1+\alpha}{1-\alpha}\|J^* - \Phi r\|_\infty.$$

(iii) The right partial derivative of $\ell(r,\theta)$ with respect to θ satisfies

$$\frac{\partial^+}{\partial\theta^+}\ell(r,0) = -\left((1-\alpha)\sum_{x\in\Omega(r)} \pi_{\mu^*,\nu}(x)\right)^{-1},$$

where

$$\Omega(r) \triangleq \operatorname*{argmax}_{\{x\in\mathcal{X}\,:\,\pi_{\mu^*,\nu}(x)>0\}} \Phi r(x) - T\Phi r(x).$$

Proof. (i) Given any r, clearly $\gamma \triangleq \|\Phi r - T\Phi r\|_\infty$, $s \triangleq \mathbf{0}$ is a feasible point for (9), so $\ell(r,\theta)$ is feasible. To see that the LP is bounded, suppose (s,γ) is feasible. Then, for any $x \in \mathcal{X}$ with $\pi_{\mu^*,\nu}(x) > 0$,

$$\gamma \geq \Phi r(x) - T\Phi r(x) - s(x) \geq \Phi r(x) - T\Phi r(x) - \theta/\pi_{\mu^*,\nu}(x) > -\infty.$$

Letting (γ_1, s_1) and (γ_2, s_2) represent optimal solutions for the LP (9) with parameters (r, θ_1) and (r, θ_2) respectively, it is easy to see that $((\gamma_1 + \gamma_2)/2, (s_1 + s_2)/2)$ is feasible for the LP with parameters $(r, (\theta_1 + \theta_2)/2)$. It follows that $\ell(r, (\theta_1 + \theta_2)/2) \leq (\ell(r, \theta_1) + \ell(r, \theta_2))/2$. The remaining properties are simple to check.

(ii) Let $\epsilon \triangleq \|J^* - \Phi r\|_\infty$. Then, since T is an α-contraction under the $\|\cdot\|_\infty$ norm,

$$\|T\Phi r - \Phi r\|_\infty \leq \|J^* - T\Phi r\|_\infty + \|J^* - \Phi r\|_\infty \leq \alpha \|J^* - \Phi r\|_\infty + \epsilon = (1 + \alpha)\epsilon.$$

Since $\gamma \triangleq \|T\Phi r - \Phi r\|_\infty$, $s \triangleq \mathbf{0}$ is feasible for (9), the result follows.

(iii) Fix $r \in \mathbb{R}^K$, and define

$$\Delta \triangleq \max_{\{x \in \mathcal{X} \,:\, \pi_{\mu^*,\nu}(x) > 0\}} \big(\Phi r(x) - T\Phi r(x)\big) - \max_{\{x \in \mathcal{X} \setminus \Omega(r) \,:\, \pi_{\mu^*,\nu}(x) > 0\}} \big(\Phi r(x) - T\Phi r(x)\big) > 0.$$

Consider the program for $\ell(r, \delta)$. It is easy to verify that for $\delta \geq 0$ and sufficiently small, viz. $\delta \leq \Delta \sum_{x \in \Omega(r)} \pi_{\mu^*,\nu}(x)$, $(\bar{s}_\delta, \bar{\gamma}_\delta)$ is an optimal solution to the program, where

$$\bar{s}_\delta(x) \triangleq \begin{cases} \frac{\delta}{\sum_{x \in \Omega(r)} \pi_{\mu^*,\nu}(x)} & \text{if } x \in \Omega(r), \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\bar{\gamma}_\delta \triangleq \gamma_0 - \frac{\delta}{\sum_{x \in \Omega(r)} \pi_{\mu^*,\nu}(x)},$$

so that

$$\ell(r, \delta) = \ell(r, 0) - \frac{\delta}{(1 - \alpha) \sum_{x \in \Omega(r)} \pi_{\mu^*,\nu}(x)}.$$

Thus,

$$\frac{\ell(r, \delta) - \ell(r, 0)}{\delta} = -\left((1 - \alpha) \sum_{x \in \Omega(r)} \pi_{\mu^*,\nu}(x)\right)^{-1}.$$

Taking a limit as $\delta \searrow 0$ yields the result. ■

Lemma 2. *Suppose that the vectors $J \in \mathbb{R}^{\mathcal{X}}$ and $s \in \mathbb{R}^{\mathcal{X}}$ satisfy*

$$J \leq T_{\mu^*} J + s.$$

Then,

$$J \leq J^* + \Delta^* s,$$

where

$$\Delta^* \triangleq \sum_{k=0}^{\infty} (\alpha P_{\mu^*})^k = (I - \alpha P_{\mu^*})^{-1},$$

and P_{μ^*} is the transition probability matrix corresponding to an optimal policy.

Proof. Note that the T_{μ^*}, the Bellman operator corresponding to the optimal policy μ^*, is monotonic and is a contraction. Then, repeatedly applying T_{μ^*} to the inequality $J \leq T_{\mu^*}J + s$ and using the fact that $T_{\mu^*}^k J \to J^*$, we obtain

$$J \leq J^* + \sum_{k=0}^{\infty} (\alpha P_{\mu^*})^k s = J^* + \Delta^* s.$$

■

Lemma 3. *For the autonomous queue with basis functions $\phi_1(x) \triangleq 1$ and $\phi_2(x) \triangleq x$, if N is sufficiently large, then*

$$\inf_{r,\psi\in\bar{\Psi}} \frac{2\nu^\top \psi}{1-\alpha\beta(\psi)} \|J^* - \Phi r\|_{\infty,1/\psi} \geq \frac{3\rho_2 q}{32(1-q)}(N-1).$$

Proof. We have:

$$\inf_{r,\psi\in\bar{\Psi}} \frac{2\nu^\top \psi}{1-\alpha\beta(\psi)} \|J^* - \Phi r\|_{\infty,1/\psi} \geq \inf_{\psi\in\bar{\Psi}} \frac{2\nu^\top \psi}{\|\psi\|_\infty} \inf_r \|J^* - \Phi r\|_\infty.$$

We will produce lower bounds on the two infima on the right-hand side above. Observe that

$$\begin{aligned}
\inf_r \|J^* - \Phi r\|_\infty &= \inf_r \max_x |\rho_2 x^2 + \rho_1 x + \rho_0 - r_1 x - r_0| \\
&\geq \inf_r \max\left(\max_x |\rho_2 x^2 + (\rho_1 - r_1)x| - |\rho_0 - r_0|, |\rho_0 - r_0|\right) \\
&= \inf_{r_0} \max\left(\inf_{r_1} \max_x |\rho_2 x^2 + (\rho_1 - r_1)x| - |\rho_0 - r_0|, |\rho_0 - r_0|\right),
\end{aligned}$$

which follows from the triangle inequality and the fact that

$$\max_x |\rho_2 x^2 + \rho_1 x + \rho_0 - r_1 x - r_0| \geq |\rho_0 - r_0|.$$

Routine algebra verifies that

$$\inf_{r_1} \max_x |\rho_2 x^2 + (\rho_1 - r_1)x| \geq \tfrac{3}{16}\rho_2(N-1)^2.$$

It thus follows that

$$\inf_r \|J^* - \Phi r\|_\infty \geq \inf_{r_0} \max\left(\tfrac{3}{16}\rho_2(N-1)^2 - |\rho_0 - r_0|, |\rho_0 - r_0|\right) \geq \tfrac{3}{32}\rho_2(N-1)^2.$$

We next note that any $\psi \in \tilde{\Psi}$ must satisfy $\psi \in \mathrm{span}(\Phi)$ and $\psi \geq \mathbf{1}$. Thus, $\psi \in \tilde{\Psi}$ must take the form $\psi(x) = \alpha_1 x + \alpha_0$ with $\alpha_0 \geq 1$ and $\alpha_1 \geq (1-\alpha_0)/(N-1)$. Thus, $\|\psi\|_\infty = \max(\alpha_1(N-1) +$

$\alpha_0, \alpha_0)$. Define $\kappa(N)$ to be the expected queue length under the distribution ν, i.e.,

$$\kappa(N) \triangleq \sum_{x=0}^{N-1} \nu(x)x = \frac{1-q}{1-q^N} \sum_{x=0}^{N-1} xq^x = \frac{q}{1-q} \left[\frac{1 - Nq^{N-1}(1-q) - q^N}{1-q^N} \right],$$

so that $\nu^\top \psi = \alpha_1 \kappa(N) + \alpha_0$, Thus,

$$\inf_{\psi \in \tilde{\Psi}} \frac{2\nu^\top \psi}{\|\psi\|_\infty} \inf_r \|J^* - \Phi r\|_\infty \geq \tfrac{3}{16}\rho_2 \inf_{\substack{\alpha_0 \geq 1 \\ \alpha_1 \geq \frac{1-\alpha_0}{N-1}}} \frac{\alpha_1 \kappa(N) + \alpha_0}{\max(\alpha_1(N-1) + \alpha_0, \alpha_0)} (N-1)^2$$

When $(1-\alpha_0)/(N-1) \leq \alpha_1 \leq 0$, we have

$$\begin{aligned} \frac{\alpha_1 \kappa(N) + \alpha_0}{\max(\alpha_1(N-1) + \alpha_0, \alpha_0)} (N-1)^2 &= \frac{\alpha_1 \kappa(N) + \alpha_0}{\alpha_0} (N-1)^2 \\ &\geq \frac{(1-\alpha_0)\kappa(N)/(N-1) + \alpha_0}{\alpha_0} (N-1)^2 \\ &\geq \left(1 - \frac{\kappa(N)}{N-1}\right)(N-1)^2. \end{aligned}$$

When $\alpha_1 > 0$, we have

$$\frac{\alpha_1 \kappa(N) + \alpha_0}{\max(\alpha_1(N-1) + \alpha_0, \alpha_0)} (N-1)^2 = \frac{\alpha_1 \kappa(N) + \alpha_0}{\alpha_1(N-1) + \alpha_0} (N-1)^2 \geq (N-1)\kappa(N),$$

where the inequality follows from the fact that $\kappa(N) \leq N-1$ and $\alpha_0 > 0$. It then follows that

$$\inf_{\psi \in \tilde{\Psi}} \frac{2\nu^\top \psi}{\|\psi\|_\infty} \inf_r \|J^* - \Phi r\|_\infty \geq \tfrac{3}{16}\rho_2 \min\left(\kappa(N)(N-1), \left(1 - \frac{\kappa(N)}{N-1}\right)(N-1)^2\right).$$

Now, observe that $\kappa(N)$ is increasing in N. Also, by assumption, $p < 1/2$, so $q < 1$ and thus $\kappa(N) \to q/(1-q)$ as $N \to \infty$. Then, for N sufficiently large, $\frac{1}{2}q/(1-q) \leq \kappa(N) \leq q/(1-q)$. Therefore, for N sufficiently large,

$$\inf_{\psi \in \tilde{\Psi}} \frac{2\nu^\top \psi}{\|\psi\|_\infty} \inf_r \|J^* - \Phi r\|_\infty \geq \frac{3\rho_2 q}{32(1-q)}(N-1),$$

as desired. ■

Lemma 4. *For every $\lambda \geq 0$, there exists a $\hat{\theta} \geq 0$ such that an optimal solution (r^*, s^*) to*

(A.1)
$$\begin{aligned} &\underset{r,s}{\text{maximize}} && \nu^\top \Phi r - \lambda \pi_{\mu^*,\nu}^\top s \\ &\text{subject to} && \Phi r \leq T\Phi r + s, \quad s \geq \mathbf{0}. \end{aligned}$$

is also an optimal solution the SALP (8) with $\theta = \hat{\theta}$.

Proof. Let $\hat{\theta} \triangleq \pi_{\mu^*,\nu}^\top s^*$. It is then clear that (r^*, s^*) is a feasible solution to (8) with $\theta = \hat{\theta}$. We claim that it is also an optimal solution. To see this, assume to the contrary that it is not an optimal solution, and let $(\tilde{r}, \tilde{s})$ be an optimal solution to (8). It must then be that $\pi_{\mu^*,\nu}^\top \tilde{s} \leq \hat{\theta} = \pi_{\mu^*,\nu}^\top s^*$ and moreover, $\nu^\top \Phi \tilde{r} > \nu^\top \Phi r^*$ so that

$$\nu^\top \Phi r^* - \lambda \pi_{\mu^*,\nu}^\top s^* < \nu^\top \Phi \tilde{r} - \lambda \pi_{\mu^*,\nu}^\top \tilde{s}.$$

This, in turn, contradicts the optimality of (r^*, s^*) for (A.1) and yields the result. ■

B. Proof of Theorem 4

Our proof of Theorem 4 is based on uniformly bounding the rate of convergence of sample averages of a certain class of functions. We begin with some definitions: consider a family $\mathcal{F}$ of functions from a set $\mathcal{S}$ to $\{0, 1\}$. Define the *Vapnik-Chervonenkis (VC) dimension* $\dim_{\mathrm{VC}}(\mathcal{F})$ to be the cardinality d of the largest set $\{x_1, x_2, \ldots, x_d\} \subset \mathcal{S}$ satisfying:

$$\forall\ e \in \{0,1\}^d,\ \exists f \in \mathcal{F} \text{ such that } \forall\ i,\ f(x_i) = 1 \text{ iff } e_i = 1.$$

Now, let $\mathcal{F}$ be some set of *real*-valued functions mapping $\mathcal{S}$ to $[0, B]$. The *pseudo-dimension* $\dim_P(\mathcal{F})$ is the following generalization of VC dimension: for each function $f \in \mathcal{F}$ and scalar $c \in \mathbb{R}$, define a function $g \colon \mathcal{S} \times \mathbb{R} \to \{0, 1\}$ according to:

$$g(x, c) \triangleq \mathbb{I}_{\{f(x) - c \geq 0\}}.$$

Let $\mathcal{G}$ denote the set of all such functions. Then, we define $\dim_P(\mathcal{F}) \triangleq \dim_{\mathrm{VC}}(\mathcal{G})$.

In order to prove Theorem 4, define the $\mathcal{F}$ to be the set of functions $f \colon \mathbb{R}^K \times \mathbb{R} \to [0, B]$, where, for all $x \in \mathbb{R}^K$ and $y \in \mathbb{R}$,

$$f(y, z) \triangleq \zeta\left(r^\top y + z\right).$$

Here, $\zeta(t) \triangleq \max\left(\min(t, B), 0\right)$, and $r \in \mathbb{R}^K$ is a vector that parameterizes f. We will show that $\dim_P(\mathcal{F}) \leq K + 2$. We will use the following standard result from convex geometry:

Lemma 5 (Radon's Lemma). *A set $A \subset \mathbb{R}^m$ of $m + 2$ points can be partitioned into two disjoint sets A_1 and A_2, such that the convex hulls of A_1 and A_2 intersect.*

Lemma 6. $\dim_P(\mathcal{F}) \leq K + 2$

Proof. Assume, for the sake of contradiction, that $\dim_P(\mathcal{F}) > K + 2$. It must be that there exists a 'shattered' set

$$\left\{\left(y^{(1)}, z^{(1)}, c^{(1)}\right), \left(y^{(2)}, z^{(2)}, c^{(2)}\right), \ldots, \left(y^{(K+3)}, z^{(K+3)}, c^{(K+3)}\right)\right\} \subset \mathbb{R}^K \times \mathbb{R} \times \mathbb{R},$$

such that, for all $e \in \{0,1\}^{K+3}$, there exists a vector $r_e \in \mathbb{R}^K$ with

$$\zeta\left(r_e^\top y^{(i)} + z^{(i)}\right) \geq c^{(i)} \text{ iff } e_i = 1, \quad \forall\ 1 \leq i \leq K+3.$$

Observe that we must have $c^{(i)} \in (0, B]$ for all i, since if $c^{(i)} \leq 0$ or $c^{(i)} > B$, then no such shattered set can be demonstrated. But if $c^{(i)} \in (0, B]$, for all $r \in \mathbb{R}^K$,

$$\zeta\left(r^\top y^{(i)} + z^{(i)}\right) \geq c^{(i)} \implies r_e^\top y^{(i)} \geq c^{(i)} - z^{(i)},$$

and

$$\zeta\left(r^\top y^{(i)} + z^{(i)}\right) < c^{(i)} \implies r_e^\top y^{(i)} < c^{(i)} - z^{(i)}.$$

For each $1 \leq i \leq K+3$, define $x^{(i)} \in \mathbb{R}^{K+1}$ component-wise according to

$$x_j^{(i)} \triangleq \begin{cases} y_j^{(i)} & \text{if } j < K+1, \\ c^{(i)} - z^{(i)} & \text{if } j = K+1. \end{cases}$$

Let $A = \{x^{(1)}, x^{(2)}, \ldots, x^{(K+3)}\} \subset \mathbb{R}^{K+1}$, and let A_1 and A_2 be subsets of A satisfying the conditions of Radon's lemma. Define a vector $\tilde{e} \in \{0,1\}^{K+3}$ component-wise according to

$$\tilde{e}_i \triangleq \mathbb{I}_{\left\{x^{(i)} \in A_1\right\}}.$$

Define the vector $\tilde{r} \triangleq r_{\tilde{e}}$. Then, we have

$$\sum_{j=1}^{K} \tilde{r}_j x_j \geq x_{K+1}, \quad \forall\ x \in A_1,$$

$$\sum_{j=1}^{K} \tilde{r}_j x_j < x_{K+1}, \quad \forall\ x \in A_2.$$

Now, let $\bar{x} \in \mathbb{R}^{K+1}$ be a point contained in both the convex hull of A_1 and the convex hull of A_2. Such a point must exist by Radon's lemma. By virtue of being contained in the convex hull of A_1, we must have

$$\sum_{j=1}^{K} \tilde{r}_j \bar{x}_j \geq \bar{x}_{K+1}.$$

Yet, by virtue of being contained in the convex hull of A_2, we must have

$$\sum_{j=1}^{K} \tilde{r}_j \bar{x}_j < \bar{x}_{K+1},$$

which is impossible. $\blacksquare$

With the above pseudo-dimension estimate, we can establish the following lemma, which provides a Chernoff bound for the *uniform* convergence of a certain class of functions:

Lemma 7. *Given a constant $B > 0$, define the function $\zeta\colon \mathbb{R} \to [0, B]$ by*

$$\zeta(t) \triangleq \max\left(\min(t, B), 0\right).$$

Consider a pair of random variables $(Y, Z) \in \mathbb{R}^K \times \mathbb{R}$. For each $i = 1, \ldots, n$, let the pair $\left(Y^{(i)}, Z^{(i)}\right)$ be an i.i.d. sample drawn according to the distribution of (Y, Z). Then, for all $\epsilon \in (0, B]$,

$$\mathsf{P}\left(\sup_{r\in\mathbb{R}^K}\left|\frac{1}{n}\sum_{i=1}^{n}\zeta\left(r^\top Y^{(i)} + Z^{(i)}\right) - \mathsf{E}\left[\zeta\left(r^\top Y + Z\right)\right]\right| > \epsilon\right) \\ \leq 8\left(\frac{32eB}{\epsilon}\log\frac{32eB}{\epsilon}\right)^{K+2}\exp\left(-\frac{\epsilon^2 n}{64B^2}\right).$$

Moreover, given $\delta \in (0, 1)$, if

$$n \geq \frac{64B^2}{\epsilon^2}\left(2(K+2)\log\frac{16eB}{\epsilon} + \log\frac{8}{\delta}\right),$$

then this probability is at most δ.

Proof. Given Lemma 6, this follows immediately from Corollary 2 of of Haussler (1992, Section 4). ∎

We are now ready to prove Theorem 4.

Theorem 4. *Under the conditions of Theorem 2, let r_{SALP} be an optimal solution to the SALP (14), and let $\hat{r}_{SALP}$ be an optimal solution to the sampled SALP (28). Assume that $r_{SALP} \in \mathcal{N}$. Further, given $\epsilon \in (0, B]$ and $\delta \in (0, 1/2]$, suppose that the number of sampled states S satisfies*

$$S \geq \frac{64B^2}{\epsilon^2}\left(2(K+2)\log\frac{16eB}{\epsilon} + \log\frac{8}{\delta}\right).$$

Then, with probability at least $1 - \delta - 2^{-383}\delta^{128}$,

$$\|J^* - \Phi\hat{r}_{SALP}\|_{1,\nu} \leq \inf_{\substack{r\in\mathcal{N}\\ \psi\in\Psi}} \|J^* - \Phi r\|_{\infty,\mathbf{1}/\psi}\left(\nu^\top\psi + \frac{2(\pi_{\mu^*,\nu}^\top\psi)(\alpha\beta(\psi)+1)}{1-\alpha}\right) + \frac{4\epsilon}{1-\alpha}.$$

Proof. Define the vectors

$$\hat{s}_{\mu^*} \triangleq \left(\Phi\hat{r}_{\text{SALP}} - T_{\mu^*}\Phi\hat{r}_{\text{SALP}}\right)^+, \quad \text{and} \quad \hat{s} \triangleq \left(\Phi\hat{r}_{\text{SALP}} - T\Phi\hat{r}_{\text{SALP}}\right)^+.$$

Note that $\hat{s}_{\mu^*} \leq \hat{s}$. One has, via Lemma 2, that

$$\Phi\hat{r}_{\text{SALP}} - J^* \leq \Delta^* \hat{s}_{\mu^*}$$

Thus, as in the last set of inequalities in the proof of Theorem 1, we have

$$\|J^* - \Phi\hat{r}_{\text{SALP}}\|_{1,\nu} \leq \nu^\top (J^* - \Phi\hat{r}_{\text{SALP}}) + \frac{2\pi_{\mu^*,\nu}^\top \hat{s}_{\mu^*}}{1-\alpha}. \tag{B.1}$$

Now, let $\hat{\pi}_{\mu^*,\nu}$ be the empirical measure induced by the collection of sampled states $\hat{\mathcal{X}}$. Given a state $x \in \mathcal{X}$, define a vector $Y(x) \in \mathbb{R}^K$ and a scalar $Z(x) \in \mathbb{R}$ according to

$$Y(x) \triangleq \Phi(x)^\top - \alpha P_{\mu^*}\Phi(x)^\top, \quad Z(x) \triangleq -g(x, \mu^*(x)),$$

so that, for any vector of weights $r \in \mathcal{N}$,

$$(\Phi r(x) - T_{\mu^*}\Phi r(x))^+ = \zeta\left(r^\top Y(x) + Z(x)\right).$$

Then,

$$\left|\hat{\pi}_{\mu^*,\nu}^\top \hat{s}_{\mu^*} - \pi_{\mu^*,\nu}^\top \hat{s}_{\mu^*}\right| \leq \sup_{r\in\mathcal{N}} \left| \frac{1}{S}\sum_{x\in\hat{\mathcal{X}}} \zeta\left(r^\top Y(x) + Z(x)\right) - \sum_{x\in\mathcal{X}} \pi_{\mu^*,\nu}(x)\zeta\left(r^\top Y(x) + Z(x)\right)\right|.$$

Applying Lemma 7, we have that

$$\mathsf{P}\left(\left|\hat{\pi}_{\mu^*,\nu}^\top \hat{s}_{\mu^*} - \pi_{\mu^*,\nu}^\top \hat{s}_{\mu^*}\right| > \epsilon\right) \leq \delta. \tag{B.2}$$

Next, suppose $(r_{\text{SALP}}, \bar{s})$ is an optimal solution to the SALP (14). Then, with probability at least $1-\delta$,

$$\begin{aligned}
\nu^\top (J^* - \Phi\hat{r}_{\text{SALP}}) + \frac{2\pi_{\mu^*,\nu}^\top \hat{s}_{\mu^*}}{1-\alpha} &\leq \nu^\top (J^* - \Phi\hat{r}_{\text{SALP}}) + \frac{2\hat{\pi}_{\mu^*,\nu}^\top \hat{s}_{\mu^*}}{1-\alpha} + \frac{2\epsilon}{1-\alpha} \\
&\leq \nu^\top (J^* - \Phi\hat{r}_{\text{SALP}}) + \frac{2\hat{\pi}_{\mu^*,\nu}^\top \hat{s}}{1-\alpha} + \frac{2\epsilon}{1-\alpha} \\
&\leq \nu^\top (J^* - \Phi r_{\text{SALP}}) + \frac{2\hat{\pi}_{\mu^*,\nu}^\top \bar{s}}{1-\alpha} + \frac{2\epsilon}{1-\alpha},
\end{aligned} \tag{B.3}$$

where the first inequality follows from (B.2), and the final inequality follows from the optimality of $(\hat{r}_{\text{SALP}}, \hat{s})$ for the sampled SALP (28).

Notice that, without loss of generality, we can assume that $\bar{s}(x) = (\Phi r_{\text{SALP}}(x) - T\Phi r_{\text{SALP}}(x))^+$, for each $x \in \mathcal{X}$. Thus, $0 \leq \bar{s}(x) \leq B$. Further,

$$\hat{\pi}_{\mu^*,\nu}^\top \bar{s} - \pi_{\mu^*,\nu}^\top \bar{s} = \frac{1}{S} \sum_{x \in \hat{\mathcal{X}}} \left(\bar{s}(x) - \pi_{\mu^*,\nu}^\top \bar{s} \right),$$

where the right-hand-side is of a sum of zero-mean bounded i.i.d. random variables. Applying Hoeffding's inequality,

$$\mathsf{P}\left(\left| \hat{\pi}_{\mu^*,\nu}^\top \bar{s} - \pi_{\mu^*,\nu}^\top \bar{s} \right| \geq \epsilon \right) \leq 2 \exp\left(-\frac{2S\epsilon^2}{B^2} \right) < 2^{-383}\delta^{128},$$

where final inequality follows from our choice of S. Combining this with (B.1) and (B.3), with probability at least $1 - \delta - 2^{-383}\delta^{128}$, we have

$$\begin{aligned} \|J^* - \Phi \hat{r}_{\text{SALP}}\|_{1,\nu} &\leq \nu^\top (J^* - \Phi r_{\text{SALP}}) + \frac{2\hat{\pi}_{\mu^*,\nu}^\top \bar{s}}{1-\alpha} + \frac{2\epsilon}{1-\alpha} \\ &\leq \nu^\top (J^* - \Phi r_{\text{SALP}}) + \frac{2\pi_{\mu^*,\nu}^\top \bar{s}}{1-\alpha} + \frac{4\epsilon}{1-\alpha}. \end{aligned}$$

The result then follows from (17)–(19) in the proof of Theorem 2. ■

References

D. Haussler. Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Information and Computation*, 100:78–150, 1992.