Lecture Note 18

1 Efficient Implementation of Approximate Linear Programming

While the ALP may involve only a small number of variables, there is a potentially intractable number of constraints — one per state-action pair. As such, we cannot in general expect to solve the ALP exactly. The focus of this paper is on a tractable approximation to the ALP: the reduced linear program (RLP).

Generation of an RLP relies on three objects: (1) a constraint sample size m, (2) a probability measure ψ over the set of state-action pairs, and (3) a bounding set $\mathcal{N} \subseteq \Re^K$. The probability measure ψ represents a distribution from which we will sample constraints. In particular, we consider a set $\mathcal{X}$ of m state-action pairs, each independently sampled according to ψ. The set $\mathcal{N}$ is a parameter that restricts the magnitude of the RLP solution. This set should be chosen such that it contains $\Phi\tilde{r}$. The RLP is defined by

$$\begin{array}{ll} \text{maximize} & c^T \Phi r \\ \text{subject to} & g_a(x) + \alpha \sum_{y \in \mathcal{S}} P_a(x,y)(\Phi r)(y) \geq (\Phi r)(x), \quad \forall (x,a) \in \mathcal{X} \\ & r \in \mathcal{N}. \end{array} \tag{1}$$

Let $\tilde{r}$ be an optimal solution of the ALP and let $\hat{r}$ be an optimal solution of the RLP. In order for the solution of the RLP to be meaningful, we would like $\|J^* - \Phi\hat{r}\|_{1,c}$ to be close to $\|J^* - \Phi\tilde{r}\|_{1,c}$. To formalize this, we consider a requirement that

$$\Pr\left\{\left|\|J^* - \Phi\hat{r}\|_{1,c} - \|J^* - \Phi\tilde{r}\|_{1,c}\right| \leq \epsilon\right\} \geq 1 - \delta,$$

where $\epsilon > 0$ is an error tolerance parameter and $\delta > 0$ parameterizes a level of confidence $1 - \delta$. This paper focusses on understanding the sample size m needed in order to meet such a requirement.

1.1 Results of our Analysis

To apply the RLP, given a problem instance, one must select parameters m, ψ, and $\mathcal{N}$. In order for the RLP to be practically solvable, the sample size m must be tractable. Results of our analysis suggest that if ψ and $\mathcal{N}$ are well-chosen, an error tolerance of ϵ can be accommodated with confidence $1 - \delta$ given a sample size m that grows as a polynomial in K, $1/\epsilon$, and $\log 1/\delta$, and is independent of the total number of ALP constraints.

Our analysis is carried out in two parts:

1. **Sample complexity of near-feasibility.** The first part of our analysis applies to constraint sampling in general linear programs – not just the ALP. Suppose that we are given a set of linear constraints

   $$\gamma_z^T r + \kappa_z \geq 0, \ \forall z \in \mathcal{Z},$$

   on variables $r \in \Re^K$, a probability measure ψ on $\mathcal{Z}$, and a desired error tolerance ϵ and confidence $1 - \delta$. Let $z_1, z_2, \ldots$ be independent identically distributed samples drawn from $\mathcal{Z}$ according to ψ. We will establish that there is a sample size

   $$m = O\left(\frac{1}{\epsilon}\left(K \ln \frac{1}{\epsilon} + \ln \frac{1}{\delta}\right)\right)$$

such that, with probability at least $1-\delta$, there exists a subset $Z \subseteq \mathcal{Z}$ of measure $\psi(Z) \geq 1-\epsilon$ such that every vector r satisfying

$$\gamma_{z_i}^T r + \kappa_{z_i} \geq 0, \quad \forall i = 1, \ldots, m,$$

also satisfies

$$\gamma_{z_i}^T r + \kappa_{z_i} \geq 0, \quad \forall z \in Z.$$

We refer to the latter criterion as *near-feasibility* — nearly all the constraints are satisfied. The main point of this part of the analysis is that near-feasibility can be obtained with high-confidence through imposing a tractable number m of samples.

2. **Sample complexity of a good approximation.** We would like the the error $\|J^* - \Phi\hat{r}\|_{1,c}$ of an optimal solution $\hat{r}$ to the RLP to be close to the error $\|J^* - \Phi\tilde{r}\|_{1,c}$ of an optimal solution to the ALP. In a generic linear program, near-feasibility is not sufficient to bound such an error metric. However, because of special structure associated with the RLP, given appropriate choices of ψ and $\mathcal{N}$, near-feasibility leads to such a bound. In particular, given a sample size

$$m = O\left(\frac{A\theta}{(1-\alpha)\epsilon}\left(K \ln \frac{A\theta}{(1-\alpha)\epsilon} + \ln \frac{1}{\delta}\right)\right),$$

where $A = \max_x |\mathcal{A}_x|$, with probability at least $1-\delta$ we have

$$\|J^* - \Phi\hat{r}\|_{1,c} \leq \|J^* - \Phi\tilde{r}\|_{1,c} + \epsilon\|J^*\|_{1,c}.$$

The parameter θ, which is to be defined precisely later, depends on the particular MDP problem instance, the choice of basis functions, and the set $\mathcal{N}$.

A major weakness of our error bound is that it relies on an idealized choice of ψ. In particular, the choice we will put forth assumes knowledge of an optimal policy. Alas, we typically do not know an optimal policy — that is what we are after in the first place. Nevertheless, the result provides guidance on what makes a desirable choice of distribution. The spirit here is analogous to one present in the importance sampling literature. In that context, the goal is to reduce variance in Monte Carlo simulation through intelligent choice of a sampling distribution and appropriate distortion of the function being integrated. Characterizations of idealized sampling distributions guide the design of heuristics that are ultimately implemented.

The set $\mathcal{N}$ also plays a critical role in the bound. It influences the value of θ, and an appropriate choice is necessary in order for this term to scale gracefully with problem size. Ideally, given a class of problems, there should be a mechanism for generating $\mathcal{N}$ such that θ grows no faster than a low-order polynomial function of the number of basis functions and the number of state variables. As we will later discuss through an example involving controlled queueing networks, we expect that it will be possible to design effective mechanisms for selecting $\mathcal{N}$ for practical classes of problems.

It is worth mentioning that our sample complexity bounds are loose. Our emphasis is on showing that the number of required samples can be independent of the total number of constraints and can scale gracefully with respect to the number of variables. Furthermore, our emphasis is on a general result that holds for a broad class of MDPs, and therefore we do not exploit special regularities associated with particular choices of basis functions or specific problems. In the presence of such special structure, one can sometimes provide much tighter bounds or even methods for exact solution of the ALP, and results of this nature can be found

in the literature, as discussed in the following literature review. The significance of our results is that they suggest viability of the linear programming approach to approximate dynamic programming even in the absence of such favorable special structure.

2 Sample Complexity of Near-Feasibility

Consider a set of linear constraints

$$\gamma_z^T r + \kappa_z \geq 0, \quad \forall z \in \mathcal{Z}$$

where $r \in \Re^K$ and $\mathcal{Z}$ is a set of constraint indices. We make the following assumption on the set of constraints:

Assumption 1 *There exists a vector $r \in \Re^K$ that satisfies the system of inequalities (2).*

We are interested in situations where there are relatively few variables and a possibly huge finite or infinite number of constraints, i.e., $K \ll |\mathcal{Z}|$. In such a situation, we expect that almost all the constraints will be irrelevant, either because they are always inactive or because they have a minor impact on the feasible region. Therefore one might speculate that the feasible region specified by all constraints can be well-approximated by a sampled subset of these constraints. In the sequel, we show that this is indeed the case, at least with respect to a certain criterion for a good approximation. We also show that the number of constraints necessary to guarantee a good approximation does not depend on the total number of constraints, but rather on the number of variables.

Our constraint sampling scheme relies on a probability measure ψ over $\mathcal{Z}$. The distribution ψ will have a dual role in our approximation scheme: on one hand, constraints will be sampled according to ψ; on the other hand, the same distribution will be involved in the criterion for assessing the quality of a particular set of sampled constraints.

We consider a subset of constraints to be good if we can guarantee that, by satisfying this subset, the set of constraints that are not satisfied has small measure. In other words, given a tolerance parameter $\epsilon \in (0, 1)$, we want to have $\mathcal{W} \subseteq \mathcal{Z}$ satisfying

$$\sup_{\{r | \gamma_z^T r + \kappa_z \geq 0, \ \forall z \in \mathcal{W}\}} \psi\left(\left\{y : \gamma_y^T r + \kappa_y < 0\right\}\right) \leq \epsilon. \tag{2}$$

Whenever (2) holds for a subset $\mathcal{W}$, we say that $\mathcal{W}$ leads to *near-feasibility*.

The next theorem establishes a bound on the number m of (possibly repeated) sampled constraints necessary to ensure that the set $\mathcal{W}$ leads to near-feasibility with probability at least $1 - \delta$.

Theorem 1 *For any $\delta \in (0, 1)$ and $\epsilon \in (0, 1)$, and*

$$m \geq \frac{4}{\epsilon}\left(K \ln \frac{12}{\epsilon} + \ln \frac{2}{\delta}\right), \tag{3}$$

a set $\mathcal{W}$ of m i.i.d. random variables drawn from $\mathcal{Z}$ according to distribution ψ, satisfies

$$\sup_{\{r : \gamma_z^T r + \kappa_z \geq 0, \ \forall z \in \mathcal{W}\}} \psi\left(\left\{y : \gamma_y^T r + \kappa_y < 0\right\}\right) \leq \epsilon, \tag{4}$$

with probability at least $1 - \delta$.

This theorem implies that, even without any special knowledge about the constraints, we can ensure near-feasibility, with high probability, through imposing a tractable subset of constraints. The result follows immediately from Corollary 8.4.2 on page 95 of [1] and the fact that the collection of sets $\{\{(\gamma,\kappa)|\gamma^T r+\kappa \geq 0\}|r \in \Re^K\}$ has VC-dimension K, as established in [2]. The main ideas for the proof are as follows:

- We define, for each r, a function $f_r : \mathcal{Z} \mapsto \{0,1\}$, given by $r \to f_r(z) = \begin{cases} 1 & \text{if } \gamma_z^T r + k_z \geq 0 \\ 0 & \text{otherwise} \end{cases}$
- We are interested in finding r such that
$$\mathrm{E}_\varphi f_r(z) \approxeq 1,$$
or equivalently, r such that almost all constraints are satisfied.
- Let us consider the problem of estimating $\mathrm{E}_\varphi f_r(z)$ and generating a near-feasible r based on sampling. Suppose we have $\hat{\mathcal{Z}} = \{\mathcal{Z}_1, \ldots, \mathcal{Z}_m\}$,
$$\mathrm{E}_\varphi f_r(z) \approx \frac{1}{m}\sum_{i=1}^{m} f_r(z_i) = \hat{\mathrm{E}}_\varphi f_r(z)$$
Then if we could ensure that
$$|\mathrm{E}_\varphi f_r - \hat{\mathrm{E}}_\varphi f_r| \leq \epsilon, \tag{5}$$
then for all $r \in \mathcal{F}_{\hat{\mathcal{Z}}}$,
$$\hat{\mathrm{E}}_\varphi f_r = 1 \Rightarrow \mathrm{E}_\varphi \geq 1-\epsilon,\ \forall\, r \in \mathcal{F}_{\hat{\mathcal{Z}}}$$
- From the VC-dimension and supervised learning lecture, we know that there is a way of ensuring (5) holds simultaneously for all r if $\mathcal{C} = \left\{f_r : r \in \Re^P\right\}$ has a finite VC-dimension.
$$P(|\mathrm{E}_\varphi f - \hat{\mathrm{E}}_\varphi f| > \epsilon) = O(\exp(-k d_{VC}(\mathcal{C})))$$
if $\hat{f} \in \mathcal{C}$.
- The final part of the proof comes from verifying that $\mathcal{C}$ has VC-dimension less than or equal to p.

Theorem 1 may be perceived as a puzzling result: the number of sampled constraints necessary for a good approximation of a set of constraints indexed by $z \in \mathcal{Z}$ depends only on the number of variables involved in these constraints and not on the set $\mathcal{Z}$. Some geometric intuition can be derived as follows. The constraints are fully characterized by vectors $[\gamma_z^T \quad \kappa_z]$ of dimension equal to the number of variables plus one. Since near-feasibility involves only consideration of whether constraints are violated, and not the magnitude of violations, we may assume without loss of generality that $\|[\gamma_z^T \quad \kappa_z]\| = 1$, for an arbitrary norm. Hence constraints can be thought of as vectors in a low-dimensional unit sphere. After a large number of constraints is sampled, they are likely to form a *cover* for the original set of constraints — i.e., any other constraint is close to one of the already sampled ones, so that the sampled constraints *cover* the set of constraints. The number of sampled constraints necessary in order to have a cover for the original set of constraints is bounded above by the number of sampled vectors necessary to form a cover to the unit sphere, which naturally depends only on the dimension of the sphere, or alternatively, on the number of variables involved in the constraints.

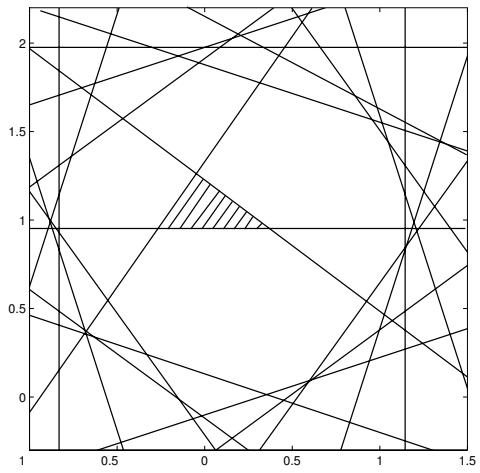

Figure 1: A feasible region defined by a large number of redundant constraints. Removing all but a random sample of constraints is likely to bring about a significant change the solution of the associated linear program.

3 Sample Complexity of a Good Approximation

In this section, we investigate the impact of using the RLP instead of the ALP on the error in the approximation of the cost-to-go function. We show in Theorem 2 that, by sampling a tractable number of constraints, the approximation error yielded by the RLP is comparable to the error yielded by the ALP.

The proof of Theorem 2 relies on special structure of the ALP. Indeed, it is easy to see that such a result cannot hold for general linear programs. For instance, consider a linear program with two variables, which are to be selected from the feasible region illustrated in Figure 1. If we remove all but a small random sample of the constraints, the new solution to the linear program is likely to be far from the solution to the original linear program. In fact, one can construct examples where the solution to a linear program is changed by an arbitrary amount by relaxing just one constraint.

Let us introduce certain constants and functions involved in our error bound. We first define a family of probability distributions on the state space $\mathcal{S}$, given by

$$\mu_u^T = (1-\alpha)c^T(I-\alpha P_u)^{-1}, \tag{6}$$

for each policy u. Note that, if c is a probability distribution, $\mu_u(x)/(1-\alpha)$ is the expected discounted number of visits to state x under policy u, if the initial state is distributed according to c. Furthermore, $\lim_{\alpha\uparrow 1}\mu_u(x)$ is a stationary distribution associated with policy u. We interpret μ_u as a measure of the relative importance of states under policy u.

We will make use of a *Lyapunov function* $V : \mathcal{S} \mapsto \Re^+$, which is defined as follows.

Definition 1 (Lyapunov function) *A function $V : \mathcal{S} \mapsto \Re^+$ is called a* Lyapunov function *if there is a scalar $\beta_V < 1$ and an optimal policy u^* such that*

$$\alpha P_{u^*} V \leq \beta_V V. \tag{7}$$

Our definition of a Lyapunov function is similar to that found in the previous lecture, with the difference that here the Lyapunov inequality (7) must hold only for an optimal policy, whereas in the previous lecture it must hold simultaneously for all policies.

Lemma 1 *Let V be a Lyapunov function for an optimal policy u^*. Then T_{u^*} is a contraction with respect to $\|\cdot\|_{\infty,1/V}$.*

Proof: Let J and $\bar{J}$ be two arbitrary vectors in $\Re^{|\mathcal{S}|}$. Then

$$T_{u^*}J - T_{u^*}\bar{J} = \alpha P_{u^*}(J - \bar{J}) \leq \|J - \bar{J}\|_{\infty,1/V}\alpha P_{u^*}V \leq \|J - \bar{J}\|_{\infty,1/V}\beta_V V.$$

□

For any Lyapunov function V, we also define another family of probability distributions on the state space $\mathcal{S}$, given by

$$\mu_{u,V}(x) = \frac{\mu_u(x)V(x)}{\mu_u^T V}. \tag{8}$$

We also define a distribution over state-action pairs

$$\psi_{u,V}(x,a) = \frac{\mu_{u,V}(x)}{|\mathcal{A}_x|},\ \forall a \in \mathcal{A}_x.$$

Finally, we define constants

$$A = \max_x |\mathcal{A}_x|$$

and

$$\theta = \frac{\mu_{u^*}^T V}{c^T J^*} \sup_{r \in \mathcal{N}} \|J^* - \Phi r\|_{\infty,1/V}. \tag{9}$$

We now present the main result of the paper — a bound on the approximation error introduced by constraint sampling.

Theorem 2 *Let ϵ and δ be scalars in $(0,1)$. Let u^* be an optimal policy and $\mathcal{X}$ be a (random) set of m state-action pairs sampled independently according to the distribution $\psi_{u^*,V}(x,a)$, for some Lyapunov function V, where*

$$m \geq \frac{16A\theta}{(1-\alpha)\epsilon}\left(K \ln \frac{48A\theta}{(1-\alpha)\epsilon} + \ln \frac{2}{\delta}\right), \tag{10}$$

Let $\tilde{r}$ be an optimal solution of the ALP that is in $\mathcal{N}$, and let $\hat{r}$ be an optimal solution of the corresponding RLP. If $\tilde{r} \in \mathcal{N}$ then, with probability at least $1-\delta$, we have

$$\|J^* - \Phi\hat{r}\|_{1,c} \leq \|J^* - \Phi\tilde{r}\|_{1,c} + \epsilon\|J^*\|_{1,c}. \tag{11}$$

Proof: From Theorem 1, given a sample size m, we have, with probability no less than $1-\delta$,

$$\begin{aligned}
\frac{(1-\alpha)\epsilon}{4A\theta} &\geq \psi_{u^*,V}\left(\{(x,a) : (T_a\Phi\hat{r})(x) < (\Phi\hat{r})(x)\}\right) \\
&= \sum_{x \in S} \frac{\mu_{u^*,V}(x)}{|\mathcal{A}_x|} \sum_{a \in \mathcal{A}_x} 1_{(T_a\Phi\hat{r})(x) < (\Phi\hat{r})(x)} \\
&\geq \frac{1}{A}\sum_{x \in S} \mu_{u^*,V}(x) 1_{(T_{u^*}\Phi\hat{r})(x) < (\Phi\hat{r})(x)}.
\end{aligned} \tag{12}$$

For any vector J, we denote the positive and negative parts by

$$J^+ = \max(J, 0), \;\; J^- = \max(-J, 0),$$

where the maximization is carried out componentwise. Note that

$$\begin{aligned}
\|J^* - \Phi\hat{r}\|_{1,c} &= c^T \left|(I - \alpha P_{u^*})^{-1}(g_{u^*} - (I - \alpha P_{u^*})\Phi\hat{r})\right| \\
&\leq c^T (I - \alpha P_{u^*})^{-1} \left|g_{u^*} - (I - \alpha P_{u^*})\Phi\hat{r}\right| \\
&= c^T (I - \alpha P_{u^*})^{-1} \left[(g_{u^*} - (I - \alpha P_{u^*})\Phi\hat{r})^+ + (g_{u^*} - (I - \alpha P_{u^*})\Phi\hat{r})^-\right] \\
&= c^T (I - \alpha P_{u^*})^{-1} \left[(g_{u^*} - (I - \alpha P_{u^*})\Phi\hat{r})^+ - (g_{u^*} - (I - \alpha P_{u^*})\Phi\hat{r})^- + \right. \\
&\qquad \left. +2\left(g_{u^*} - (I - \alpha P_{u^*})\Phi\hat{r}\right)^-\right] \\
&= c^T (I - \alpha P_{u^*})^{-1} \left[g_{u^*} - (I - \alpha P_{u^*})\Phi\hat{r} + 2\left(T_{u^*}\Phi\hat{r} - \Phi\hat{r}\right)^-\right] \\
&= c^T (J^* - \Phi\hat{r}) + 2c^T (I - \alpha P_{u^*})^{-1} \left(T_{u^*}\Phi\hat{r} - \Phi\hat{r}\right)^- . \qquad (13)
\end{aligned}$$

The inequality comes from the fact that $c > 0$ and

$$(I - \alpha P_{u^*})^{-1} = \sum_{n=0}^{\infty} \alpha^n P_{u^*}^n \geq 0,$$

where the inequality is componentwise, so that

$$\begin{aligned}
\left|(I - \alpha P_{u^*})^{-1}(g_{u^*} - (I - \alpha P_{u^*})\Phi\hat{r})\right| &\leq \left|(I - \alpha P_{u^*})^{-1}\right| \left|(g_{u^*} - (I - \alpha P_{u^*})\Phi\hat{r})\right| \\
&= (I - \alpha P_{u^*})^{-1} \left|(g_{u^*} - (I - \alpha P_{u^*})\Phi\hat{r})\right| .
\end{aligned}$$

Now let $\tilde{r}$ be any optimal solution of the ALP[1]. Clearly, $\tilde{r}$ is feasible for the RLP. Since $\hat{r}$ is the optimal solution of the same problem, we have $c^T\Phi\hat{r} \geq c^T\Phi\tilde{r}$ and

$$\begin{aligned}
c^T (J^* - \Phi\hat{r}) &\leq c^T (J^* - \Phi\tilde{r}) \\
&= \|J^* - \Phi\tilde{r}\|_{1,c}, \qquad (14)
\end{aligned}$$

therefore we just need to show that the second term in (13) is small to guarantee that the performance of the RLP is not much worse than that of the ALP.

[1]Note that all optimal solutions of the ALP yield the same approximation error $\|J^* - \Phi r\|_{1,c}$, hence the error bound (11) is independent of the choice of $\tilde{r}$.

Now

$$\begin{aligned}
2c^T(I-\alpha P_{u^*})^{-1}\left(T_{u^*}\Phi\hat{r}-\Phi\hat{r}\right)^- &= \frac{2}{1-\alpha}\mu_{u^*}^T\left(T_{u^*}\Phi\hat{r}-\Phi\hat{r}\right)^- \\
&= \frac{2}{1-\alpha}\sum_{x\in S}\mu_{u^*}(x)\left((\Phi\hat{r})(x)-(T_{u^*}\Phi\hat{r})(x)\right)1_{(T_{u^*}\Phi\hat{r})(x)<(\Phi\hat{r})(x)} \\
&= \frac{2}{1-\alpha}\sum_{x\in S}\frac{(\Phi\hat{r})(x)-(T_{u^*}\Phi\hat{r})(x)}{V(x)}\mu_{u^*}(x)V(x)1_{(T_{u^*}\Phi\hat{r})(x)<(\Phi\hat{r})(x)} \\
&\leq \frac{2\mu_{u^*}^T V}{1-\alpha}\|T_{u^*}\Phi\hat{r}-\Phi\hat{r}\|_{\infty,1/V}\sum_{x\in S}\mu_{u^*,V}(x)1_{(T_{u^*}\Phi\hat{r})(x)<(\Phi\hat{r})(x)} \\
&\leq \frac{\epsilon}{2\theta}\mu_{u^*}^T V\|T_{u^*}\Phi\hat{r}-\Phi\hat{r}\|_{\infty,1/V} \\
&\leq \frac{\epsilon}{2\theta}\mu_{u^*}^T V(\|T_{u^*}\Phi\hat{r}-J^*\|_{\infty,1/V}+\|J^*-\Phi\hat{r}\|_{\infty,1/V}) \\
&\leq \frac{\epsilon}{2\theta}\mu_{u^*}^T V(1+\beta_V)\|J^*-\Phi\hat{r}\|_{\infty,1/V} \\
&\leq \epsilon\|J^*\|_{1,c},
\end{aligned}$$

with probability greater than or equal to $1-\delta$, where second inequality follows from (12) and the fourth inequality follows from Lemma 1. The error bound (11) then follows from (13) and (14). □

Three aspects of Theorem 2 deserve further consideration. The first of them is the dependence of the number of sampled constraints (10) on θ. Two parameters of the RLP influence the behavior of θ: the Lyapunov function V and the bounding set $\mathcal{N}$. Graceful scaling of the sample complexity bound depends on the ability to make appropriate choices for these parameters.

The number of sampled constraints also grows polynomially with the maximum number of actions available per state A, which makes the proposed approach inapplicable to problems with a large number of actions per state. It can be shown that complexity in the action space can be exchanged for complexity in the state space, so that such problems can be recast in a format that is amenable to our approach.

Finally, a major weakness of Theorem 2 is that it relies on sampling constraints according to the distribution $\psi_{u^*,V}$. In general, $\psi_{u^*,V}$ is not known, and constraints must be sampled according to an alternative distribution $\overline{\psi}$. Suppose that $\overline{\psi}(x,a)=\overline{\mu}(x)/|\mathcal{A}_x|$ for some state distribution $\overline{\mu}$. If $\overline{\mu}$ is "similar" to $\mu_{u^*,V}$, one might hope that the error bound (11) holds with a number of samples m close to the number suggested in the theorem. We discuss two possible motivations for this:

1. It is conceivable that sampling constraints according to $\overline{\psi}$ leads to a small value of
$$\mu_{u^*,V}\left(\{x:(\Phi\hat{r})(x)\geq(T_{u^*}\Phi\hat{r})(x)\}\right)\leq(1-\alpha)\epsilon/2,$$
with high probability, even though $\mu_{u^*,V}$ is not identical to $\overline{\mu}$. This would lead to a graceful sample complexity bound, along the lines of (10). Establishing such a guarantee is closely related to the problem of computational learning when the training and testing distributions differ.

2. If
$$\mu_{u^*}^T(T_{u^*}\Phi r-\Phi r)^-\leq C\tilde{\mu}^T(T_{u^*}\Phi r-\Phi r)^-,$$
for some scalar C and all r, where
$$\tilde{\mu}(x)=\frac{\overline{\mu}(x)/V(x)}{\sum_{y\in S}\overline{\mu}(y)/V(y)},$$

then the error bound (11) holds with probability $1-\delta$ given

$$m \geq \frac{16A\theta C}{(1-\alpha)\epsilon}\left(K \ln \frac{48A\theta C}{(1-\alpha)\epsilon} + \ln \frac{2}{\delta}\right),$$

samples. It is conceivable that this will be true for a reasonably small value of C in relevant contexts.

How to choose $\overline{\mu}$ is an open question, and most likely to be addressed adequately having in mind the particular application at hand. As a simple heuristic, noting that $\mu_{u^*}(x) \to c(x)$ as $\alpha \to 0$, one might choose $\overline{\mu}(x) = c(x)V(x)$.

References

[1] D. Anthony and N. Biggs. *Computational Learning Theory.* Cambridge University Press, 1992.

[2] R.M. Dudley. Central limit theorems for empirical measures. *Annals of Probability*, 6(6):899–928, 1978.