Electronic Companion to Self-guided Approximate Linear Programs

Parshan Pakiman, Selvaprabu Nadarajah, Negar Soheili
College of Business Administration, University of Illinois at Chicago, 601 South Morgan Street, Chicago, IL60607, USA
{ppakim2@uic.edu, selvan@uic.edu, nazad@uic.edu}

Qihang Lin
Tippie College of Business, The University of Iowa, 21 East Market Street, Iowa City, IA 52242, USA
qihang-lin@uiowa.edu

Our proofs are reported in §EC.1. Sections §EC.2 and §EC.3 are addenda to §3.1 and §4.2, respectively. Additional details of our numerical study are reported in §EC.4. In §EC.4.1, we discuss our sampling approach to compute valid lower bounds for constraint-sampled ALPs, and we provide our numerical study for the generalized joint replenishment application in §EC.4.2.

EC.1. Proofs

We define a constant $\Gamma := (1+\gamma)/(1-\gamma)$ which we will use in various proofs.

EC.1.1. Additional Details of Assumption 1

Assumptions 1 and 2 will hold for all proofs in the electronic companions. In particular, Assumption 1 ensures the existence of an optimal policy solving program (1). There are known conditions in the literature that guarantee such existence. For the purposes of our proofs, we formalize Assumption 1 as follows.

Assumption EC.1. *It holds that (i) the MDP cost function is bounded over $\mathcal{S}\times\mathcal{A}_s$ and function $c(s,\cdot):\mathcal{A}_s\mapsto\mathbb{R}$ is lower semicontinuous for all $s\in\mathcal{S}$; (ii) for every bounded and measurable function $V:\mathcal{S}\mapsto\mathbb{R}$, the mapping $(s,a)\mapsto\int_{\mathcal{S}}V(s')P(\mathrm{d}s'|s,a)$ is bounded and continuous over $\mathcal{S}\times\mathcal{A}_s$; and (iii) there exists a finite-cost policy $\pi\in\Pi$ such that $\mathrm{PC}(s,\pi)<\infty$ for all $s\in\mathcal{S}$.*

Assumption EC.1 is adopted from assumptions 4.2.1 and 4.2.2 in Hernández-Lerma and Lasserre 1996, henceforth abbreviated as HL. Specifically, in Part (a) of Assumption 4.2.1 in HL, the cost function $c(s,\cdot)$ is assumed to be lower semi-continuous, non-negative, and inf-compact (defined in Condition 3.3.3 in HL) whereas, in our setting, non-negativity is replaced by boundedness and the inf-compactness is guaranteed by the virtue of $c(s,\cdot)$ being lower semi-continuous and its domain $\mathcal{A}_s$ being compact (please see the first paragraph of §2.1). Part (b) of Assumption 4.2.1 and Assumption 4.2.2 in HL are equivalent to parts (ii) and (iii) of Assumption EC.1, respectively. Under the aforementioned technical conditions, Part (b) of Theorem 4.2.3 in HL guarantees the existence of a deterministic and stationary policy $\pi^*\in\Pi$ that is "γ-discount optimal". In other words, $\pi^*\in\Pi$ solves (1) in our setting.

EC.1.2. Proofs of Statement in §2

Proof of Proposition 1. Part (i). Since the optimal value function $V^* \in \mathcal{C}$ is continuous (by Assumption 1) and the class of random basis function φ is universal (by Assumption 2), there is a finite constant $C \geq 0$ and $\bar{V} \in \mathcal{R}_C(\varphi,\rho)$ such that $\|V^* - \bar{V}\|_\infty \leq \varepsilon$. Since $\bar{V}$ belongs to $\mathcal{R}_C(\varphi,\rho)$, it can be written as $\bar{V}(s) = \bar{b}_0 + \langle \bar{\boldsymbol{b}}, \varphi(s)\rangle$ for some $(\bar{b}_0, \bar{\boldsymbol{b}})$ with $\|\bar{\boldsymbol{b}}\|_{\infty,\rho} \leq C$. Recall that $\Gamma = (1+\gamma)/(1-\gamma)$. We now show that $(b_0^\varepsilon, \boldsymbol{b}^\varepsilon) = \big(\bar{b}_0 - \Gamma\varepsilon, \bar{\boldsymbol{b}}\big)$ is the desired feasible FELP solution. This is because $\|\boldsymbol{b}^\varepsilon\|_{\infty,\rho} = \|\bar{\boldsymbol{b}}\|_{\infty,\rho} \leq C$ and for any $(s,a) \in \mathcal{S} \times \mathcal{A}_s$, we have

$$\begin{aligned}(1-\gamma)b_0^\varepsilon + \big\langle \boldsymbol{b}^\varepsilon, \varphi(s) - \gamma\mathbb{E}[\varphi(s')|s,a]\big\rangle &= (1-\gamma)\big(\bar{b}_0 - \Gamma\varepsilon\big) + \big\langle \bar{\boldsymbol{b}}, \varphi(s) - \gamma\mathbb{E}[\varphi(s')|s,a]\big\rangle \\ &= -(1+\gamma)\varepsilon + \bar{V}(s) - \gamma\mathbb{E}[\bar{V}(s')|s,a] \\ &\leq -(1+\gamma)\varepsilon + V^*(s) + \varepsilon - \gamma\mathbb{E}[V^*(s') - \varepsilon|s,a] \\ &= V^*(s) - \gamma\mathbb{E}[V^*(s')|s,a] \\ &\leq c(s,a),\end{aligned}$$

where the first inequality is valid since $\|V^* - \bar{V}\|_\infty \leq \varepsilon$, which ensures $\bar{V}(s) \leq V^*(s) + \varepsilon$ and $-\bar{V}(s) \leq -V^*(s) + \varepsilon$ for all $s \in \mathcal{S}$. Thus, $(b_0^\varepsilon, \boldsymbol{b}^\varepsilon)$ is feasible to FELP. In addition, the value function $V^\varepsilon(\cdot) := \bar{V}(\cdot) - \Gamma\varepsilon$ associated with the FELP feasible solution $(b_0^\varepsilon, \boldsymbol{b}^\varepsilon)$ belongs to $\mathcal{R}_C(\varphi,\rho)$ and $\|V^* - V^\varepsilon\|_\infty \leq \|V^* - \bar{V}\|_\infty + \Gamma\varepsilon \leq \varepsilon + \Gamma\varepsilon = 2\varepsilon/(1-\gamma)$, which completes the proof.

Part (ii). Consider an optimal solution $(b_0^{\text{FE}}, \boldsymbol{b}^{\text{FE}})$ to FELP and let $V^{\text{FE}}(\cdot) = b_0^{\text{FE}} + \big\langle \boldsymbol{b}^{\text{FE}}, \varphi(\cdot)\big\rangle$. Using the function V^ε defined in Part (i) of this proposition, we have $\|V^* - V^{\text{FE}}\|_{1,\nu} \leq \|V^* - V^\varepsilon\|_{1,\nu} \leq \|V^* - V^\varepsilon\|_\infty \leq 2\varepsilon/(1-\gamma)$ where the first inequality follows from (3) (which is based on Lemma 1 in De Farias and Van Roy 2003) since $(b_0^{\text{FE}}, \boldsymbol{b}^{\text{FE}})$ and $(b_0^\varepsilon, \boldsymbol{b}^\varepsilon)$ are optimal and feasible solutions, respectively, to FELP. □

EC.1.3. Proofs of Statements in §3

Lemma EC.1. *Any continuous function $V : \mathcal{S} \mapsto \mathbb{R}$ that is feasible to constraints* (2) *satisfies $V(s) \leq V^*(s)$ for all $s \in \mathcal{S}$.*

Proof. The proof follows from Part (b) of Lemma 4.2.7 in HL, which requires four assumptions to hold. We now show that these assumptions are true in our setting. First, since V is continuous, it is measurable. Second, the Bellman operator $\mathrm{T}V(s) := \min_{a\in\mathcal{A}_s}\{c(s,a) + \gamma\mathbb{E}[V(s')|s,a]\}$ is well defined since the minimum in its definition is attained via the compactness of $\mathcal{A}_s$ and the finiteness of the expectation $\mathbb{E}[V(s')|s,a] = \int_{\mathcal{S}} V(s')P(\mathrm{d}s'|s,a)$, which holds by Assumption EC.1. Third, since V is feasible to constraints (2), we have

$$V(s) \;\leq\; \min_{a\in\mathcal{A}_s}\{c(s,a) + \gamma\mathbb{E}[V(s')|s,a]\} \;=\; \mathrm{T}V(s), \qquad \forall s \in \mathcal{S}.$$

Fourth, the continuity of V and the compactness of $\mathcal{S}$ imply $\|V\|_\infty < \infty$ and

$$\lim_{n\to\infty} \gamma^n \mathbb{E}\left[\sum_{t=0}^{n} V(s_t^\pi)\Big| s_0 = s\right] \;\le\; \|V(s)\|_\infty \lim_{n\to\infty}(n+1)\gamma^n \;=\; 0, \qquad \forall s\in\mathcal{S}, \pi\in\Pi,$$

where expectation $\mathbb{E}$ and the notation s_t^π retain their definitions from §2.1. Hence, the function V fulfills the four assumptions of Part (b) of Lemma 4.2.7 in HL and thus $V(\cdot) \le V^*(\cdot)$. □

Definition EC.1. Fix an optimal solution $(b_0^{\text{FE}}, \boldsymbol{b}^{\text{FE}})$ to FELP. For N independent and identical sampled parameters $\{\theta_1, \theta_2, \ldots, \theta_N\}$ from ρ, we define the coordinates of $\boldsymbol{\beta}^{\text{FEAS}} \in \mathbb{R}^{N+1}$ as follows

$$\beta_i^{\text{FEAS}} := \begin{cases} b_0^{\text{FE}} & \text{if} \quad i = 0; \\ \dfrac{\boldsymbol{b}^{\text{FE}}(\theta_i)}{N\rho(\theta_i)} & \text{if} \quad i = 1, 2, \ldots, N. \end{cases}$$

Lemma EC.2. *Given $\varepsilon > 0$ and $\delta \in (0, 1]$, let*

$$N_\varepsilon := \left\lceil \varepsilon^{-2} \left\|\boldsymbol{b}^{\text{FE}}\right\|_{\infty,\rho}^2 (\Omega + \Delta_\delta)^2 \right\rceil. \tag{EC.1}$$

(i) If $N \ge N_\varepsilon$, it holds that $\left\|V^{\text{FE}} - V(\boldsymbol{\beta}^{\text{FEAS}})\right\|_\infty \le \varepsilon$ with a probability of at least $1-\delta$.

(ii) If $N \ge N_\varepsilon$, with a probability of at least $1-\delta$, the vector $(\beta_0^{\text{FEAS}} - \Gamma\varepsilon, \beta_1^{\text{FEAS}}, \ldots, \beta_N^{\text{FEAS}})$ is feasible to FALP$_{(\mathrm{N})}$ and

$$\left\|V^{\text{FE}} - \left(V(\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon\right)\right\|_\infty \le \frac{2\varepsilon}{(1-\gamma)}.$$

Proof. Part (i). Since $\beta_0^{\text{FEAS}} = b_0^{\text{FE}}$, we have for $N \ge N_\varepsilon$

$$\begin{aligned}
\left\|V^{\text{FE}} - V(\boldsymbol{\beta}^{\text{FEAS}})\right\|_\infty &= \left\| b_0^{\text{FE}} + \left\langle \boldsymbol{b}^{\text{FE}}, \varphi(s)\right\rangle - \left(b_0^{\text{FE}} + \sum_{i=1}^{N} \beta_i^{\text{FEAS}} \varphi(s;\theta_i)\right)\right\|_\infty \\
&= \left\| \left\langle \boldsymbol{b}^{\text{FE}}, \varphi(s)\right\rangle - \sum_{i=1}^{N} \beta_i^{\text{FEAS}} \varphi(s;\theta_i)\right\|_\infty \\
&\le \frac{\left\|\boldsymbol{b}^{\text{FE}}\right\|_{\infty,\rho}}{\sqrt{N}} \left(\sqrt{2\ln\left(\frac{1}{\delta}\right)} \;+\; 4(\operatorname{diam}(\mathcal{S})+1)\mathrm{L}_\varphi \sqrt{\mathbb{E}_\rho\left[\|\theta\|_2^2\right]}\right) \\
&\le \frac{\left\|\boldsymbol{b}^{\text{FE}}\right\|_{\infty,\rho}}{\sqrt{N_\varepsilon}} \left(\sqrt{2\ln\left(\frac{1}{\delta}\right)} \;+\; 4(\operatorname{diam}(\mathcal{S})+1)\mathrm{L}_\varphi \sqrt{\mathbb{E}_\rho\left[\|\theta\|_2^2\right]}\right),
\end{aligned} \tag{EC.2}$$

where the first inequality holds with a probability of at least $1-\delta$ by Theorem 3.2 of Rahimi and Recht (2008) after adjusting our notation to theirs. To help the reader, we discuss the notational differences in Remark EC.1 immediately following this proof. We can now use the definitions of Ω and Δ_δ (see §3.1) in N_ε to simplify the right hand side of (EC.2) to ε and get

$$\left\|V^{\text{FE}} - V(\boldsymbol{\beta}^{\text{FEAS}})\right\|_\infty \le \varepsilon$$

with a probability of at least $1-\delta$.

Part (ii). If $N \geq N_\varepsilon$, the vector $(\beta_0^{\text{FEAS}} - \Gamma\varepsilon, \beta_1^{\text{FEAS}}, \ldots, \beta_N^{\text{FEAS}})$ is feasible to $\text{FALP}_{(\text{N})}$ with a probability of at least $1-\delta$ since

$$\begin{aligned}(1-\gamma)\big(\beta_0^{\text{FEAS}} - \Gamma\varepsilon\big) &+ \sum_{i=1}^{N} \beta_i^{\text{FEAS}}\big(\varphi(s;\theta_i) - \gamma\mathbb{E}\big[\varphi(s';\theta_i)\big|s,a\big]\big)\\ &= V(s;\boldsymbol{\beta}^{\text{FEAS}}) - \varepsilon - \gamma\mathbb{E}\big[V(s';\boldsymbol{\beta}^{\text{FEAS}}) + \varepsilon\big|s,a\big]\\ &\leq V^{\text{FE}}(s) - \gamma\mathbb{E}[V^{\text{FE}}(s')|s,a]\\ &= (1-\gamma)b_0^{\text{FE}} + \big\langle \boldsymbol{b}^{\text{FE}}, \varphi(s) - \gamma\mathbb{E}[\varphi(s')|s,a]\big\rangle\\ &\leq c(s,a),\end{aligned} \tag{EC.3}$$

where the first equality comes from the definitions of $V(s;\boldsymbol{\beta}^{\text{FEAS}})$ and Γ; the first inequality holds because $|V^{\text{FE}}(s) - V(s;\boldsymbol{\beta}^{\text{FEAS}})| \leq \|V^{\text{FE}} - V(\boldsymbol{\beta}^{\text{FEAS}})\|_\infty \leq \varepsilon$ for all $s \in \mathcal{S}$ with a probability of at least $1-\delta$ by Part (i) of this lemma; the second equality results from using the definition of V^{FE}; and the second inequality holds because $(b_0^{\text{FE}}, \boldsymbol{b}^{\text{FE}})$ is an optimal (hence feasible) solution of FELP.

Moreover, if $N \geq N_\varepsilon$, by Part (i) of this lemma and the definition of Γ, we get

$$\big\|V^{\text{FE}} - \big(V(\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon\big)\big\|_\infty \leq \big\|V^{\text{FE}} - V(\boldsymbol{\beta}^{\text{FEAS}})\big\|_\infty + \Gamma\varepsilon \leq \varepsilon + \Gamma\varepsilon = \frac{2\varepsilon}{(1-\gamma)}$$

with a probability of at least $1-\delta$. □

Remark EC.1. We use the notations $(1,s)$, φ, $\boldsymbol{b}$, ρ, N, L_φ, and $\text{diam}(\mathcal{S})$ +1 in this paper instead of x, ϕ, $\boldsymbol{\alpha}$, p, K, L, and B, respectively, in Rahimi and Recht (2008). The additional 1 in the term $\text{diam}(\mathcal{S})+1$ is due to the notational differences between x and $(1,s)$ used in Rahimi and Recht (2008) and this paper, respectively. Moreover, the function class $\mathscr{F}$ defined in §III of Rahimi and Recht (2008) is the same as $\mathcal{R}_\infty(\varphi,\rho)$ and the functions $\big\langle \boldsymbol{b}^{\text{FE}}, \varphi(s)\big\rangle \in \mathcal{R}_\infty(\varphi,\rho)$ with $\|\boldsymbol{b}^{\text{FE}}\|_{\infty,\rho} < \infty$ and $\sum_{i=1}^{N} \beta_i^{\text{FEAS}}\varphi(s;\theta_i)$ satisfy the conditions of Theorem 3.2 in Rahimi and Recht (2008).

Proof of Theorem 1. Consider Part (ii) of Lemma EC.2 that ensures $(\beta_0^{\text{FEAS}} - \Gamma\varepsilon, \beta_1^{\text{FEAS}}, \ldots, \beta_N^{\text{FEAS}})$ is a feasible solution to $\text{FALP}_{(\text{N})}$ with a probability of at least $1-\delta$ if $N \geq N_\varepsilon$. Let $\{\theta_1,\ldots,\theta_N\}$ be any N independent and identical samples from ρ defining function $V(\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon$ corresponding to vector $(\beta_0^{\text{FEAS}} - \Gamma\varepsilon, \beta_1^{\text{FEAS}}, \ldots, \beta_N^{\text{FEAS}})$. Mapping $L : \Theta^N \mapsto \mathbb{R}$ defined as

$$L(\theta_1,\ldots,\theta_N) \;:=\; \mathbb{E}_\nu\big[V^{\text{FE}} - (V(\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon)\big] \;=\; \Gamma\varepsilon + \mathbb{E}_\nu\Bigg[\big\langle \boldsymbol{b}^{\text{FE}}, \varphi(s)\big\rangle - \sum_{i=1}^{N}\beta_i^{\text{FEAS}}\varphi(s;\theta_i)\Bigg], \tag{EC.4}$$

has important properties. First, it satisfies

$$\begin{aligned}\mathbb{E}_\rho[L(\theta_1,\ldots,\theta_N)] &= \Gamma\varepsilon + \mathbb{E}_\nu\Bigg[\big\langle \boldsymbol{b}^{\text{FE}}, \varphi(s)\big\rangle - \mathbb{E}_\rho\Bigg[\sum_{i=1}^{N}\beta_i^{\text{FEAS}}\varphi(s;\theta_i)\Bigg]\Bigg]\\ &= \Gamma\varepsilon + \mathbb{E}_\nu\Bigg[\big\langle \boldsymbol{b}^{\text{FE}}, \varphi(s)\big\rangle - \sum_{i=1}^{N}\mathbb{E}_\rho\Bigg[\frac{\boldsymbol{b}^{\text{FE}}(\theta_i)}{N\rho(\theta_i)}\varphi(s;\theta_i)\Bigg]\Bigg]\end{aligned}$$

$$
\begin{aligned}
&= \Gamma\varepsilon + \mathbb{E}_\nu\left[\left\langle \boldsymbol{b}^{\text{FE}}, \varphi(s)\right\rangle - \frac{1}{N}\sum_{i=1}^{N}\int_\Theta \frac{\boldsymbol{b}^{\text{FE}}(\theta)}{\rho(\theta)}\varphi(s;\theta)\rho(\theta)\,\mathrm{d}\,\theta\right]\Bigg]\\
&= \Gamma\varepsilon,
\end{aligned}
$$

where the second equality is obtained using the definition of $\boldsymbol{\beta}^{\text{FEAS}}$, the third one holds since the θ_i's are independent and identical samples, and the last one follows from the definition of $\left\langle \boldsymbol{b}^{\text{FE}}, \varphi(s)\right\rangle$. Second, for any $\ell \in \{1,2,\ldots,N\}$ and parameter $\hat{\theta}_\ell \in \Theta$, the function $L(\cdot)$ has the following property:

$$
\begin{aligned}
\sup_{\theta_1,\ldots,\theta_N,\hat{\theta}_\ell}\left|L(\theta_1,\ldots,\theta_\ell,\ldots,\theta_N) - L(\theta_1,\ldots,\hat{\theta}_\ell,\ldots,\theta_N)\right| &= \sup_{\theta_\ell,\hat{\theta}_\ell}\left|\beta_\ell^{\text{FEAS}}\mathbb{E}_\nu[\varphi(s;\theta_\ell)] - \beta_\ell^{\text{FEAS}}\mathbb{E}_\nu[\varphi(s;\hat{\theta}_\ell)]\right|\\
&\le 2\sup_{\theta_\ell}\left|\beta_\ell^{\text{FEAS}}\right|\\
&= 2\sup_{\theta}\left|\frac{\boldsymbol{b}^{\text{FE}}(\theta)}{N\rho(\theta)}\right|\\
&\le \frac{2}{N}\left\|\boldsymbol{b}^{\text{FE}}\right\|_{\infty,\rho}.
\end{aligned}
$$

The first equality follows from the fact that the two points $(\theta_1,\ldots,\theta_\ell,\ldots,\theta_N)$ and $(\theta_1,\ldots,\hat{\theta}_\ell,\ldots,\theta_N)$ only differ in their ℓ^{th} components. The first inequality is obtained using $\|\bar{\varphi}\|_\infty \le 1$ (please see Assumption 2), the second equality follows from the definition of β_ℓ^{FEAS} in Definition EC.1, and the last inequality is based on the definition of $\left\|\boldsymbol{b}^{\text{FE}}\right\|_{\infty,\rho}$. Given $\bar{\varepsilon} > 0$, these two properties of $L(\cdot)$ and an application of McDiarmid's inequality (see, e.g., Theorem D.3 in Mohri et al. 2012) to function $L(\cdot)$ give us

$$
\begin{aligned}
\Pr\left(L(\theta_1,\ldots,\theta_N) - \mathbb{E}_\rho[L(\theta_1,\ldots,\theta_N)] \ge \bar{\varepsilon}\right) &= \Pr\left(\Gamma\varepsilon + \mathbb{E}_\nu\left[\left\langle \boldsymbol{b}^{\text{FE}}, \varphi(s)\right\rangle - \sum_{i=1}^{N}\beta_i^{\text{FEAS}}\varphi(s;\theta_i)\right] - \Gamma\varepsilon \ge \bar{\varepsilon}\right)\\
&= \Pr\left(\mathbb{E}_\nu\left[\left\langle \boldsymbol{b}^{\text{FE}}, \varphi(s)\right\rangle - \sum_{i=1}^{N}\beta_i^{\text{FEAS}}\varphi(s;\theta_i)\right] \ge \bar{\varepsilon}\right)\\
&\le \exp\left(\frac{-N\bar{\varepsilon}^2}{2\left\|\boldsymbol{b}^{\text{FE}}\right\|_{\infty,\rho}^2}\right),
\end{aligned}
$$

where $\Pr(\cdot)$ denotes the probability over the samples $(\theta_1,\ldots,\theta_N)$ drawn from ρ. If we set the right-hand-side of the above inequality to δ and solve for $\bar{\varepsilon}$, then with a probability of at least $1-\delta$, it holds that

$$
\mathbb{E}_\nu\left[\left\langle \boldsymbol{b}^{\text{FE}}, \varphi(s)\right\rangle - \sum_{i=0}^{N}\beta_i^{\text{FEAS}}\varphi(s;\theta_i)\right] \quad\le\quad \left\|\boldsymbol{b}^{\text{FE}}\right\|_{\infty,\rho}\sqrt{2\ln\left(\frac{1}{\delta}\right)}\Big/\sqrt{N} \quad\le\quad \Delta_\delta\varepsilon\big/(\Omega+\Delta_\delta) \qquad \text{(EC.5)}
$$

where the second inequality follows from our choice of $N \ge N_\varepsilon$ and the definition of Δ_δ. Using (EC.4), (EC.5), and the definition of Γ, we obtain

$$
\mathbb{E}_\nu\left[V^{\text{FE}} - (V(\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon)\right] = \Gamma\varepsilon + \mathbb{E}_\nu\left[\left\langle \boldsymbol{b}^{\text{FE}}, \varphi(s)\right\rangle - \sum_{i=1}^{N}\beta_i^{\text{FEAS}}\varphi(s;\theta_i)\right]
$$

$$\begin{aligned}&\le \Gamma\varepsilon + \Delta_\delta\varepsilon/\left(\Omega+\Delta_\delta\right)\\ &\le \frac{2\varepsilon}{(1-\gamma)}\cdot\left(\frac{(1+\gamma)\Omega+2\Delta_\delta}{2(\Omega+\Delta_\delta)}\right),\end{aligned}\tag{EC.6}$$

which holds with a probability of at least $1-\delta$. Let $\lambda := ((1+\gamma)\Omega+2\Delta_\delta)/2(\Omega+\Delta_\delta)$. Then choosing ε as $\frac{\varepsilon}{\lambda}$ in Part (ii) of Lemma EC.2 indicates that for any

$$N \ge N_{\varepsilon/\lambda} = \left\lceil \varepsilon^{-2}\ \left\|\boldsymbol{b}^{\text{FE}}\right\|_{\infty,\rho}^{2}\ \left(\frac{(1+\gamma)}{2}\Omega\ +\ \Delta_\delta\right)^2\right\rceil,$$

the vector $(\beta_0^{\text{FEAS}}-\frac{\Gamma\varepsilon}{\lambda}, \beta_1^{\text{FEAS}},\dots,\beta_N^{\text{FEAS}})$ is feasible to $\text{FALP}_{(\text{N})}$ with a probability of at least $1-\delta$. In addition, following the exact same steps as in (EC.6) we get $\mathbb{E}_\nu\left[V^{\text{FE}}-(V(\boldsymbol{\beta}^{\text{FEAS}})-\frac{\Gamma\varepsilon}{\lambda})\right] \le \frac{2\varepsilon}{(1-\gamma)}$. Hence,

$$\begin{aligned}\left\|V^*-V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\right\|_{1,\nu} &= \mathbb{E}_\nu\left[V^*+V^{\text{FE}}-V^{\text{FE}}-V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\right]\\ &= \mathbb{E}_\nu\left[V^*-V^{\text{FE}}\right]+\mathbb{E}_\nu\left[V^{\text{FE}}-V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\right]\\ &\le \frac{2\varepsilon}{(1-\gamma)}+\mathbb{E}_\nu\left[V^{\text{FE}}-\left(V(\boldsymbol{\beta}^{\text{FEAS}})-\frac{\Gamma\varepsilon}{\lambda}\right)\right]\\ &\le \frac{4\varepsilon}{(1-\gamma)},\end{aligned}\tag{EC.7}$$

with a probability of at least $1-\delta$, where we used Lemma EC.1 to derive the first equality and Part (ii) of Proposition 1 and the optimality of to $\boldsymbol{\beta}_{\text{N}}^{\text{FA}}$ to $\text{FALP}_{(\text{N})}$ to derive the first inequality. □

Proof of Proposition 2. Given $\tau>0$, to prove this proposition, we show that τ^* becomes smaller than τ in a finite number of iterations with high probability. To do so, we bound the terms $\text{LB}(\boldsymbol{\beta}^{\text{LB}})$ and $\text{PC}(\boldsymbol{\beta}^{\text{UB}})$ used in the definition of τ^* from below and above, respectively, and show that the ratio $\text{LB}(\boldsymbol{\beta}^{\text{LB}})/\text{PC}(\boldsymbol{\beta}^{\text{UB}})$ can get arbitrarily close to one when N is sufficiently large.

Finding a lower bound on $\text{LB}(\boldsymbol{\beta}^{\text{LB}})$*:* Since we set $\mathcal{M}_{(\text{N})}$ to $\text{FALP}_{(\text{N})}$ in Algorithm 1, vector $\boldsymbol{\beta}_{\text{N}}$ in this algorithm equals $\boldsymbol{\beta}_{\text{N}}^{\text{FA}}$. For a given N, define $\mathcal{S}' := \{s\in\mathcal{S}:\nu(s)=0\}\subseteq\mathcal{S}$, $\mathcal{S}'' := \{s\in\mathcal{S}:|\mu_\chi(s;\boldsymbol{\beta}_{\text{N}})|=\infty\}\subseteq\mathcal{S}$, $\mathcal{S}^0 := \mathcal{S}'\cup\mathcal{S}''$, $W_1 := \sup_{s\in\mathcal{S}\setminus\mathcal{S}^0}\left\{\mu_\chi(s;\boldsymbol{\beta}_{\text{N}})/\nu(s)\right\}\in(0,\infty)$, and $W_2 := \sup_{s\in\mathcal{S}\setminus\mathcal{S}'}\left\{\chi(s)/\nu(s)\right\}\in(0,\infty)$. By our assumptions on ν and $\mu_\chi(\boldsymbol{\beta}_{\text{N}})$, the sets $\mathcal{S}'$, $\mathcal{S}''$, and $\mathcal{S}_0$ have zero measure. Then, when $N\ge\left\lceil\varepsilon^{-2}\ \left\|\boldsymbol{b}^{\text{FE}}\right\|_{\infty,\rho}^2\ \left(\frac{(1+\gamma)}{2}\Omega\ +\ \Delta_\delta\right)^2\right\rceil$, we can write

$$\begin{aligned}\mathbb{E}_\chi[V^*]-\mathbb{E}_\chi\left[V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\right] &= \int_{\mathcal{S}}\left(V^*(s)-V(s;\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\right)\chi(\mathrm{d}\,s)\\ &= \int_{\mathcal{S}\setminus\mathcal{S}'}\left(V^*(s)-V(s;\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\right)\frac{\chi(s)}{\nu(s)}\nu(\mathrm{d}\,s)\\ &\le W_2\left\|V^*-V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\right\|_{1,\nu}\\ &\le \frac{4W_2\varepsilon}{(1-\gamma)},\end{aligned}$$

where the second equality is valid since $\mathcal{S}'$ is a zero-measure set; the first inequality holds since V^* is a pointwise upper bound on $V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})$ by Lemma EC.1 given $V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})$ is continuous and feasible to constraints (2); and the second inequality holds for the choice of N with a probability of at least $1-\delta$ by Theorem 1. Using the above inequalities, with the same probability, it holds that

$$\text{LB}(\boldsymbol{\beta}^{\text{LB}}) \geq \text{LB}(\boldsymbol{\beta}_{\text{N}}) = \text{LB}(\boldsymbol{\beta}_{\text{N}}^{\text{FA}}) = \mathbb{E}_{\chi}\big[V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\big] \geq \mathbb{E}_{\chi}[V^*] - \frac{4W_2\varepsilon}{(1-\gamma)}, \tag{EC.8}$$

where the first inequality follows from Step (iv) of Algorithm 1 which indicates that $\text{LB}(\boldsymbol{\beta}^{\text{LB}})$ is always the largest lower bound.

Finding an upper bound on $\text{UB}(\boldsymbol{\beta}^{\text{UB}})$: We can write for $N \geq \left\lceil \varepsilon^{-2} \; \left\|\boldsymbol{b}^{\text{FE}}\right\|_{\infty,\rho}^2 \; \left(\frac{(1+\gamma)}{2}\Omega \; + \; \Delta_\delta\right)^2 \right\rceil$ that

$$\begin{aligned}
\left\|V^* - V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\right\|_{\mu_\chi(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})} &= \int_{\mathcal{S}} \left|V^*(s) - V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\right| \mu_\chi(s;\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\,\mathrm{d}\,s \\
&= \int_{\mathcal{S}\setminus\mathcal{S}^0} \left|V^*(s) - V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\right| \frac{\mu_\chi(s;\boldsymbol{\beta}_{\text{N}}^{\text{FA}})}{\nu(s)} \nu(\mathrm{d}\,s) \\
&\leq \sup_{s\in\mathcal{S}\setminus\mathcal{S}^0} \left\{\frac{\mu_\chi(s;\boldsymbol{\beta}_{\text{N}}^{\text{FA}})}{\nu(s)}\right\} \left\|V^* - V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\right\|_{1,\nu} \\
&\leq \frac{4W_1\varepsilon}{(1-\gamma)},
\end{aligned}$$

where the second equality is valid since $\mathcal{S}_0$ is a zero-measure set and the second inequality holds for the chosen N with a probability of at least $1-\delta$ by Theorem 1. Utilizing Proposition 3 in §3.1 and the definition of $\text{PC}(\cdot)$ in §2.1, with the same probability, we obtain

$$\text{PC}(\boldsymbol{\beta}_{\text{N}}^{\text{FA}}) - \mathbb{E}_\chi[V^*] = \left\|\text{PC}(\cdot;\pi_g(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})) - V^*(\cdot)\right\|_{1,\chi} \;\leq\; \frac{\|V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}}) - V^*\|_{1,\mu_\chi(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})}}{1-\gamma} \leq \frac{4W_1\varepsilon}{(1-\gamma)^2}.$$

Therefore,

$$\text{PC}(\boldsymbol{\beta}^{\text{UB}}) \leq \text{PC}(\boldsymbol{\beta}_{\text{N}}) = \text{PC}(\boldsymbol{\beta}_{\text{N}}^{\text{FA}}) \leq \mathbb{E}_\chi\big[V^*\big] + \frac{4W_1\varepsilon}{(1-\gamma)^2}, \tag{EC.9}$$

where the first inequality follows from Step (v) in Algorithm 1 which guarantees that $\text{PC}(\boldsymbol{\beta}^{\text{UB}})$ is always the smallest upper bound. Using (EC.8) and (EC.9), we obtain

$$\tau^* = 1 - \frac{\text{LB}(\boldsymbol{\beta}^{\text{LB}})}{\text{PC}(\boldsymbol{\beta}^{\text{UB}})} \leq 1 - \left[\left(\mathbb{E}_\chi\left[V^*\right] - \frac{4W_2\varepsilon}{(1-\gamma)}\right)\Big/\left(\mathbb{E}_\chi\left[V^*\right] + \frac{4W_1\varepsilon}{(1-\gamma)^2}\right)\right] \leq \frac{4(W_1 + (1-\gamma)W_2)}{(1-\gamma)^2\mathbb{E}_\chi[V^*]}\varepsilon,$$

which holds with a probability of at least $1-\delta$. For $W_3 := (1-\gamma)^2\mathbb{E}_\chi[V^*]/4(W_1 + (1-\gamma)W_2)$, if we choose $\varepsilon < W_3\tau$, then for

$$N \;>\; N_\tau = \left\lceil \tau^{-2} W_3^{-2} \; \left\|\boldsymbol{b}^{\text{FE}}\right\|_{\infty,\rho}^2 \; \left(\frac{(1+\gamma)}{2}\Omega \; + \; \Delta_\delta\right)^2 \right\rceil,$$

we have $\tau^* < \tau$ and Algorithm 1 terminates with a probability of at least $1-\delta$ in $\lceil (N_\tau + 1)/B \rceil$ iterations. Notice that we used Theorem 1 to obtain N_τ by replacing ε with $W_3\tau$. □

EC.1.4. Proofs of Statements in §4

Proof of Proposition 4. Any VFA in the set $\{V(\cdot;\boldsymbol{\beta}^{\mathrm{FG}}_{\mathrm{mB}}) : m = 1,2,\dots,n\}$ is a continuous function because of the Lipschitz continuity of $\bar{\varphi}$ in Assumption 2. Moreover, each function $V(\cdot;\boldsymbol{\beta}^{\mathrm{FG}}_{\mathrm{mB}})$ is feasbile to the constraints (2) since each vector $\boldsymbol{\beta}^{\mathrm{FG}}_{\mathrm{mB}}$ is feasible to the constraints (6) of $\mathrm{FGLP}_{(\mathrm{mB})}$. As a result of these two observations, Lemma EC.1 guarantees $V(s;\boldsymbol{\beta}^{\mathrm{FG}}_{\mathrm{mB}}) \le V^*(s)$ for $m = 1,2,\dots,n$ and $s \in \mathcal{S}$. In addition, the constraints (7) in FGLP indicate that $V(\cdot;\boldsymbol{\beta}^{\mathrm{FG}}_{\mathrm{mB}}) \le V(\cdot;\boldsymbol{\beta}^{\mathrm{FG}}_{(\mathrm{m+1})\mathrm{B}})$ for $m = 1,2,\dots,n-1$. □

The rest of this section is devoted to our projection-based sampling bound for FGLP in Theorem 2. Our analysis is based on a known projection in the space of (2,ρ) integrable functions, which is formalized in Lemma EC.3.

Lemma EC.3 **(Example 4.5 in Rudin 1987)**. *Define the space of (2,ρ) integrable functions $\mathcal{B}$ with its associated inner product $\langle\cdot,\cdot\rangle_\rho$ and norm $\|\cdot\|^2_{2,\rho} := \langle\cdot,\cdot\rangle_\rho$ as*

$$\mathcal{B} := \left\{\boldsymbol{b} : \Theta \mapsto \mathbb{R} \middle| \|\boldsymbol{b}\|_{2,\rho} < \infty\right\} \quad \textit{and} \quad \langle \boldsymbol{b},\boldsymbol{b}'\rangle_\rho := \int_\Theta \frac{\boldsymbol{b}(\theta)\ \boldsymbol{b}'(\theta)}{\rho(\theta)}\,\mathrm{d}\theta \textit{ for } \boldsymbol{b},\boldsymbol{b}' \in \mathcal{B}.$$

Then, the space $\mathcal{B}$ equipped with inner product $\langle\cdot,\cdot\rangle_\rho$ form a Hilbert space.

Definition EC.2 below models coefficients of functions (excluding the intercept) in space $\mathcal{W}_\alpha(\Phi_N)$ and connects such a space to the Hilbert space $\mathcal{B}$. Lemma EC.4 uses this definition to set up our orthogonal projection.

Definition EC.2. Given N samples $\{\theta_1,\dots,\theta_N\}$ and $\alpha \in \left(0, \min_{i\neq j}\|\theta_i - \theta_j\|_2\right)$, we define

$$\mathcal{B}_{\alpha,N} \equiv \mathcal{B}_\alpha(\Phi_N) := \left\{\boldsymbol{b} \in \mathcal{B} \;\middle|\; \exists(\beta_1,\dots,\beta_N) \text{ and } \boldsymbol{b}(\theta) = \sum_{i=1}^N \beta_i \phi_{i,\alpha}(\theta)\right\},$$

and let $\overline{\mathcal{B}}_{\alpha,N}$ and $\overline{\mathcal{B}}^{\perp}_{\alpha,N}$ to be the closure of $\mathcal{B}_{\alpha,N}$ and the perpendicular complement of $\overline{\mathcal{B}}_{\alpha,N}$, respectively. In particular,

$$\overline{\mathcal{B}}^{\perp}_{\alpha,N} := \left\{\boldsymbol{b} \in \mathcal{B} \middle| \langle \boldsymbol{b},\boldsymbol{b}'\rangle_\rho = 0,\ \forall \boldsymbol{b}' \in \overline{\mathcal{B}}_{\alpha,N}\right\}.$$

Lemma EC.4. *Let $\alpha \in \left(0, \min_{i\neq j}\|\theta_i - \theta_j\|_2\right)$. For a given $\varepsilon > 0$, fix a feasible solution $(b_0^\varepsilon, \boldsymbol{b}^\varepsilon)$ to FELP satisfying $\|V^* - V^\varepsilon\|_\infty \le 2\varepsilon/(1-\gamma)$, where $V^\varepsilon(\cdot) = b_0^\varepsilon + \langle \boldsymbol{b}^\varepsilon, \varphi(\cdot)\rangle \in \mathcal{R}_C(\varphi,\rho)$. Then there exist $\boldsymbol{b}^\varepsilon_\alpha \in \overline{\mathcal{B}}_{\alpha,N}$ and $\overset{\perp}{\boldsymbol{b}}{}^\varepsilon_\alpha \in \overline{\mathcal{B}}^{\perp}_{\alpha,N}$ such that (i) the function $\boldsymbol{b}^\varepsilon$ admits the decomposition $\boldsymbol{b}^\varepsilon = \boldsymbol{b}^\varepsilon_\alpha + \overset{\perp}{\boldsymbol{b}}{}^\varepsilon_\alpha$; (ii) the Pythagorean identity $\|\boldsymbol{b}^\varepsilon\|_{2,\rho} = \|\boldsymbol{b}^\varepsilon_\alpha\|_{2,\rho} + \|\overset{\perp}{\boldsymbol{b}}{}^\varepsilon_\alpha\|_{2,\rho}$ holds; (iii) the norms $\|\boldsymbol{b}^\varepsilon_\alpha\|_{\infty,\rho}$ and $\|\overset{\perp}{\boldsymbol{b}}{}^\varepsilon_\alpha\|_{\infty,\rho}$ are finite; and (iv) $V^\varepsilon = V^\varepsilon_\alpha + \overset{\perp}{V}{}^\varepsilon_\alpha$ where*

$$V^\varepsilon_\alpha(\cdot) := b_0^\varepsilon + \langle \boldsymbol{b}^\varepsilon_\alpha, \varphi(\cdot)\rangle \quad \textit{and} \quad \overset{\perp}{V}{}^\varepsilon_\alpha(\cdot) := \langle \overset{\perp}{\boldsymbol{b}}{}^\varepsilon_\alpha, \varphi(\cdot)\rangle.$$

Proof. The required FELP feasible solution $(b_0^{\varepsilon}, \boldsymbol{b}^{\varepsilon})$ exists by Part (i) of Proposition 1. Clearly $\left\|\boldsymbol{b}^{\varepsilon}\right\|_{\infty,\rho} \leq C$ since $\boldsymbol{b}^{\varepsilon}$ is feasible to FELP. Thus, from the inequalities,

$$\left\|\boldsymbol{b}^{\varepsilon}\right\|_{2,\rho}^2 \;=\; \int_{\Theta} \frac{\boldsymbol{b}^{\varepsilon}(\theta)^2}{\rho(\theta)}\,\mathrm{d}\,\theta \;=\; \int_{\Theta}\left(\frac{\boldsymbol{b}^{\varepsilon}(\theta)}{\rho(\theta)}\right)^2 \rho(\mathrm{d}\,\theta) \;\leq\; \int_{\Theta}\left(\sup_{\theta}\left|\frac{\boldsymbol{b}^{\varepsilon}(\theta)}{\rho(\theta)}\right|\right)^2 \rho(\mathrm{d}\,\theta) \;\leq\; \left\|\boldsymbol{b}^{\varepsilon}\right\|_{\infty,\rho}^2,$$

we have $\left\|\boldsymbol{b}^{\varepsilon}\right\|_{2,\rho} \leq \left\|\boldsymbol{b}^{\varepsilon}\right\|_{\infty,\rho} < \infty$ which shows $\boldsymbol{b}^{\varepsilon} \in \mathcal{B}$. It is straightforward to see that $\overline{\mathcal{B}}_{\alpha,N}$ is a closed linear subspace of $\mathcal{B}$. Leveraging the orthogonal projection in Hilbert spaces (see, e.g., Theorem 5.24 in Folland 1999), we can decompose Hilbert space $\mathcal{B}$ into elements $\overline{\mathcal{B}}_{\alpha,N}$ and $\overline{\mathcal{B}}_{\alpha,N}^{\perp}$. Therefore, given $\boldsymbol{b}^{\varepsilon} \in \mathcal{B}$, there exist $\boldsymbol{b}_1 \in \overline{\mathcal{B}}_{\alpha,N}$ and $\boldsymbol{b}_2 \in \overline{\mathcal{B}}_{\alpha,N}^{\perp}$ such that $\boldsymbol{b}^{\varepsilon} = \boldsymbol{b}_1 + \boldsymbol{b}_2$. Since two components $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are orthogonal in Hilbert space $\mathcal{B}$ (see Lemma EC.3), they satisfy the Pythagorean identity $\left\|\boldsymbol{b}^{\varepsilon}\right\|_{2,\rho} = \left\|\boldsymbol{b}_1\right\|_{2,\rho} + \left\|\boldsymbol{b}_2\right\|_{2,\rho}$ (i.e., Theorem 5.23 in Folland 1999). We next show that weighting functions $\boldsymbol{b}_{\alpha}^{\varepsilon}$ and $\overset{\perp}{\boldsymbol{b}}{}_{\alpha}^{\varepsilon}$ with finite (∞,ρ)-norms can be constructed using $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$, respectively, such that $\boldsymbol{b}^{\varepsilon} = \boldsymbol{b}_{\alpha}^{\varepsilon} + \overset{\perp}{\boldsymbol{b}}{}_{\alpha}^{\varepsilon}$. Let $\Theta_1 := \left\{\theta \in \Theta : |\boldsymbol{b}_1(\theta)| = \infty\right\}$ and $\Theta_2 := \left\{\theta \in \Theta : |\boldsymbol{b}_2(\theta)| = \infty\right\}$. If these sets are empty, then we can set $\boldsymbol{b}_{\alpha}^{\varepsilon} = \boldsymbol{b}_1$ and $\overset{\perp}{\boldsymbol{b}}{}_{\alpha}^{\varepsilon} = \boldsymbol{b}_2$. Otherwise, we show they are both zero-measure sets. By contradiction, assume that at least one of them is not a zero-measure set. Then,

$$\begin{aligned}
\|\boldsymbol{b}_1\|_{2,\rho} + \|\boldsymbol{b}_2\|_{2,\rho} = &\int_{\Theta_1} \frac{(\boldsymbol{b}_1(\theta))^2}{\rho(\theta)}\,\mathrm{d}\,\theta + \int_{\Theta\backslash\Theta_1} \frac{(\boldsymbol{b}_1(\theta))^2}{\rho(\theta)}\,\mathrm{d}\,\theta + \int_{\Theta_2} \frac{(\boldsymbol{b}_2(\theta))^2}{\rho(\theta)}\,\mathrm{d}\,\theta + \int_{\Theta\backslash\Theta_2} \frac{(\boldsymbol{b}_2(\theta))^2}{\rho(\theta)}\,\mathrm{d}\,\theta \\
\geq &\int_{\Theta_1} \frac{(\boldsymbol{b}_1(\theta))^2}{\rho(\theta)}\,\mathrm{d}\,\theta + \int_{\Theta_2} \frac{(\boldsymbol{b}_2(\theta))^2}{\rho(\theta)}\,\mathrm{d}\,\theta = \infty
\end{aligned} \tag{EC.10}$$

which is a contradiction with $\|\boldsymbol{b}_1\|_{2,\rho} + \|\boldsymbol{b}_2\|_{2,\rho} = \|\boldsymbol{b}^{\varepsilon}\|_{2,\rho} \leq \|\boldsymbol{b}^{\varepsilon}\|_{\infty,\rho} \leq CU_{\rho} < \infty$. Further, Θ_1 must equal Θ_2 since otherwise for any $\hat{\theta} \in \Theta_1\backslash\Theta_2$ (or $\hat{\theta} \in \Theta_2\backslash\Theta_1$), we get $\boldsymbol{b}^{\varepsilon}(\hat{\theta}) = \boldsymbol{b}_1(\hat{\theta}) + \boldsymbol{b}_2(\hat{\theta}) = \infty$, which contradicts $\boldsymbol{b}^{\varepsilon}(\hat{\theta}) \leq CU_{\rho}$. To guarantee $\boldsymbol{b}_{\alpha}^{\varepsilon}$ and $\overset{\perp}{\boldsymbol{b}}{}_{\alpha}^{\varepsilon}$ have finite (∞,ρ)-norms, we construct them as follows:

$$\boldsymbol{b}_{\alpha}^{\varepsilon}(\theta) := \begin{cases} \boldsymbol{b}_1(\theta) & \text{if } \theta \in \Theta\backslash\Theta_1; \\ \boldsymbol{b}^{\varepsilon}(\theta)/2 & \text{if } \theta \in \Theta_1; \end{cases} \quad \text{and} \quad \overset{\perp}{\boldsymbol{b}}{}_{\alpha}^{\varepsilon}(\theta) := \begin{cases} \boldsymbol{b}_2(\theta) & \text{if } \theta \in \Theta\backslash\Theta_1; \\ \boldsymbol{b}^{\varepsilon}(\theta)/2 & \text{if } \theta \in \Theta_1. \end{cases} \tag{EC.11}$$

Since $\|\boldsymbol{b}^{\varepsilon}\|_{\rho,\infty}$ is finite, both $\left\|\boldsymbol{b}_{\alpha}^{\varepsilon}\right\|_{\rho,\infty}$ and $\left\|\overset{\perp}{\boldsymbol{b}}{}_{\alpha}^{\varepsilon}\right\|_{\rho,\infty}$ are finite by definition. In addition, it can be easily verified that the Pythagorean identity $\|\boldsymbol{b}_{\alpha}^{\varepsilon}\|_{2,\rho} + \|\overset{\perp}{\boldsymbol{b}}{}_{\alpha}^{\varepsilon}\|_{2,\rho} = \|\boldsymbol{b}^{\varepsilon}\|_{2,\rho}$ and the equation $\boldsymbol{b}^{\varepsilon}(\theta) = \boldsymbol{b}_{\alpha}^{\varepsilon}(\theta) + \overset{\perp}{\boldsymbol{b}}{}_{\alpha}^{\varepsilon}(\theta)$ hold. Replacing $\boldsymbol{b}^{\varepsilon}$ with $\boldsymbol{b}_{\alpha}^{\varepsilon} + \overset{\perp}{\boldsymbol{b}}{}_{\alpha}^{\varepsilon}$ in the definition of $V^{\varepsilon} = b_0^{\varepsilon} + \left\langle \boldsymbol{b}^{\varepsilon}, \varphi(s)\right\rangle$, we obtain the decomposition $V^{\varepsilon} = V_{\alpha}^{\varepsilon} + \overset{\perp}{V}{}_{\alpha}^{\varepsilon}$. □

Lemmas EC.5 and EC.6 show that the orthogonal functions V_{α}^{ε} and $\overset{\perp}{V}{}_{\alpha}^{\varepsilon}$ can be approximated using finite random samples θ from ρ.

LEMMA EC.5. *Given $\xi > 0$ and $\alpha \in \left(0, \min_{i\neq j}\|\theta_i - \theta_j\|_2\right)$, there is a function $V_N \in \mathcal{W}(\Phi_N)$ such that*

$$\left\|V_{\alpha}^{\varepsilon} - V_N\right\|_{\infty} \;\leq\; \sqrt{U_{\rho}}\xi + \alpha\sqrt{NU_{\rho}}\mathrm{L}_{\varphi}\left(\mathrm{diam}(\mathcal{S}) + 1\right)\left(\|\boldsymbol{b}^{\varepsilon}\|_{2,\rho} + \xi\right) \tag{EC.12}$$

where $V_{\alpha}^{\varepsilon}(\cdot) = b_0^{\varepsilon} + \left\langle \boldsymbol{b}_{\alpha}^{\varepsilon}, \varphi(\cdot)\right\rangle$ is defined in Lemma EC.4.

Proof. Since $\boldsymbol{b}^{\varepsilon}_{\alpha} \in \overline{\mathcal{B}}_{\alpha,N}$, there is a weighting function $\boldsymbol{b}_{\alpha} \in \mathcal{B}_{\alpha,N}$ such that $\|\boldsymbol{b}^{\varepsilon}_{\alpha} - \boldsymbol{b}_{\alpha}\|_{2,\rho} \le \xi$. Let $V_{\alpha}(\cdot) := b^{\varepsilon}_0 + \langle \boldsymbol{b}_{\alpha}, \varphi(\cdot)\rangle$. For all $s \in \mathcal{S}$, we have

$$\begin{aligned}
\Big(V_{\alpha}(s) - V^{\varepsilon}_{\alpha}(s)\Big)^2 &= \left(\int_{\Theta} \big(\boldsymbol{b}_{\alpha}(\theta) - \boldsymbol{b}^{\varepsilon}_{\alpha}(\theta)\big)\varphi(s;\theta)\,\mathrm{d}\,\theta\right)^2 \\
&\le \int_{\Theta} \big(\boldsymbol{b}_{\alpha}(\theta) - \boldsymbol{b}^{\varepsilon}_{\alpha}(\theta)\big)^2 (\varphi(s;\theta))^2\,\mathrm{d}\,\theta \\
&\le \int_{\Theta} \frac{U_{\rho}}{\rho(\theta)} \big(\boldsymbol{b}_{\alpha}(\theta) - \boldsymbol{b}^{\varepsilon}_{\alpha}(\theta)\big)^2\,\mathrm{d}\,\theta \\
&= U_{\rho}\|\boldsymbol{b}^{\varepsilon}_{\alpha} - \boldsymbol{b}_{\alpha}\|^2_{2,\rho},
\end{aligned}$$

where the first inequality holds by the Jensen's inequality and the second one follows from $\rho(\theta) \le U_{\rho}$ for all $\theta \in \Theta$ and the fact that $\|\bar{\varphi}\|_{\infty} \le 1$ by Assumption 2. Hence, we have

$$\big\|V_{\alpha} - V^{\varepsilon}_{\alpha}\big\|_{\infty} \;\le\; \sqrt{U_{\rho}}\|\boldsymbol{b}^{\varepsilon}_{\alpha} - \boldsymbol{b}_{\alpha}\|_{2,\rho} \;\le\; \sqrt{U_{\rho}}\xi. \tag{EC.13}$$

Since $\boldsymbol{b}_{\alpha} \in \mathcal{B}_{\alpha,N}$, it can be written as $\boldsymbol{b}_{\alpha}(\theta) = \sum_{i=1}^{N} \beta_i \phi_{i,\alpha}(\theta)$ for some real-valued coefficients $\beta_1, \ldots, \beta_N$ and the function $V_{\alpha}(\cdot) := b^{\varepsilon}_0 + \langle \boldsymbol{b}_{\alpha}, \varphi(\cdot)\rangle$ belongs to $\mathcal{W}_{\alpha}(\Phi_N)$. Define $V_N(\cdot) := b^{\varepsilon}_0 + \sum_{i=1}^{N} \beta_i \varphi(\cdot;\theta_i)$. We next show that $\|V_N - V_{\alpha}\|_{\infty}$ is bounded. Consider the inequalities

$$\begin{aligned}
|\varphi(s;\theta) - \varphi(s;\theta_i)| &= \big|\bar{\varphi}\big(\langle(1,s),\theta\rangle\big) - \bar{\varphi}\big(\langle(1,s),\theta_i\rangle\big)\big| \\
&\le \mathrm{L}_{\varphi}\|(1,s)\|_2\|\theta - \theta_i\|_2 \\
&\le \mathrm{L}_{\varphi}(\mathrm{diam}(\mathcal{S}) + 1)\|\theta - \theta_i\|_2,
\end{aligned}$$

where the equality follows from the definition of φ in Assumption 2 and the first inequality is obtained by the Lipschitz continuity of $\bar{\varphi}$ and the Hölder's inequality. Then, for all $i = 1, 2, \ldots, N$, $\theta \in \Theta$, and $s \in \mathcal{S}$ we have

$$\begin{aligned}
\frac{\beta_i}{z^i_{\alpha}}\rho(\theta)\mathbb{1}\big\{\|\theta - \theta_i\|_2 \le \alpha\big\}\varphi(s;\theta) \;\le\; &\frac{\beta_i}{z^i_{\alpha}}\rho(\theta)\mathbb{1}\big\{\|\theta - \theta_i\|_2 \le \alpha\big\}\varphi(s;\theta_i) \\
&+ \left|\frac{\beta_i}{z^i_{\alpha}}\right|\rho(\theta)\mathbb{1}\big\{\|\theta - \theta_i\|_2 \le \alpha\big\}\mathrm{L}_{\varphi}(\mathrm{diam}(\mathcal{S}) + 1)\|\theta - \theta_i\|_2.
\end{aligned}$$

Summing the above inequality over samples θ_i, integrating over θ, and adding intercept b^{ε}_0 leads to,

$$\begin{aligned}
V_{\alpha}(s) &= b^{\varepsilon}_0 + \int_{\Theta} \sum_{i=1}^{N} \frac{\beta_i}{z^i_{\alpha}}\rho(\theta)\mathbb{1}\big\{\|\theta - \theta_i\|_2 \le \alpha\big\}\varphi(s;\theta)\,\mathrm{d}\,\theta \\
&\le b^{\varepsilon}_0 + \sum_{i=1}^{N} \frac{\beta_i}{z^i_{\alpha}}\varphi(s;\theta_i)\int_{\Theta} \rho(\theta)\mathbb{1}\big\{\|\theta - \theta_i\|_2 \le \alpha\big\}\,\mathrm{d}\,\theta\, + \\
&\quad \mathrm{L}_{\varphi}(\mathrm{diam}(\mathcal{S}) + 1)\left(\sum_{i=1}^{N}\left|\frac{\beta_i}{z^i_{\alpha}}\right|\int_{\Theta} \rho(\theta)\mathbb{1}\big\{\|\theta - \theta_i\|_2 \le \alpha\big\}\|\theta - \theta_i\|_2\,\mathrm{d}\,\theta\right) \\
&\le b^{\varepsilon}_0 + \sum_{i=1}^{N} \beta_i\varphi(s;\theta_i) + \mathrm{L}_{\varphi}(\mathrm{diam}(\mathcal{S}) + 1)\alpha\sum_{i=1}^{N}\left|\frac{\beta_i}{z^i_{\alpha}}\right|\int_{\Theta} \rho(\theta)\mathbb{1}\big\{\|\theta - \theta_i\|_2 \le \alpha\big\}\,\mathrm{d}\,\theta \\
&= V_N(s) + \mathrm{L}_{\varphi}(\mathrm{diam}(\mathcal{S}) + 1)\alpha\sum_{i=1}^{N}|\beta_i|,
\end{aligned}$$

where the second inequality derived using the inequality $\|\theta-\theta_i\|_2 \leq \alpha$ that holds for all θ in the α-ball $\mathcal{U}_i(\alpha) \coloneqq \{\theta \in \Theta : \|\theta-\theta_i\|_2 \leq \alpha\}$ around sample θ_i with $i \in \{1,2,\ldots,N\}$, and the second equality follows from definition of z^i_α in §4.2. Similarly, one can bound V_α from below as $V_\alpha(s) \geq V_N(s) - \mathrm{L}_\varphi(\mathrm{diam}(\mathcal{S})+1)\alpha\sum_{i=1}^N |\beta_i|$ for all $s \in \mathcal{S}$. This indicates that

$$\big\|V_\alpha - V_N\big\|_\infty \;\leq\; \mathrm{L}_\varphi(\mathrm{diam}(\mathcal{S})+1)\alpha\sum_{i=1}^N |\beta_i|\,. \tag{EC.14}$$

We next find an α-independent upper bound on the term $\sum_{i=1}^N |\beta_i|$. Notice that for $\theta \in \mathcal{U} \coloneqq \bigcup_{i=1}^N \mathcal{U}_i(\alpha)$, there is exactly one index $i \in \{1,2,\ldots,N\}$ such that $\mathbb{1}\big\{\|\theta-\theta_i\|_2 \leq \alpha\big\} = 1$, and for all $\theta \in \Theta\backslash\mathcal{U}$, it holds that $\mathbb{1}\big\{\|\theta-\theta_i\|_2 \leq \alpha\big\} = 0$ for all $i \in \{1,2,\ldots,N\}$. Thus, we can write

$$\begin{aligned}
\big\|\boldsymbol{b}_\alpha\big\|^2_{2,\rho} &= \int_\Theta \frac{1}{\rho(\theta)}\bigg(\sum_{i=1}^N \frac{\beta_i}{z^i_\alpha}\rho(\theta)\mathbb{1}\big\{\|\theta-\theta_i\|_2 \leq \alpha\big\}\bigg)^2 \mathrm{d}\theta \\
&\geq \frac{1}{U_\rho}\int_\Theta \Big(\sum_{i=1}^N \frac{\beta_i}{z^i_\alpha}\rho(\theta)\mathbb{1}\big\{\|\theta-\theta_i\|_2 \leq \alpha\big\}\Big)^2 \mathrm{d}\theta \\
&= \frac{1}{U_\rho}\int_{\mathcal{U}} \Big(\sum_{i=1}^N \frac{\beta_i}{z^i_\alpha}\rho(\theta)\mathbb{1}\big\{\|\theta-\theta_i\|_2 \leq \alpha\big\}\Big)^2 \mathrm{d}\theta \\
&= \frac{1}{U_\rho}\sum_{k=1}^N \int_{\mathcal{U}_k(\alpha)} \bigg(\sum_{i=1}^N \frac{\beta^i}{z^i_\alpha}\rho(\theta)\mathbb{1}\big\{\|\theta-\theta_i\|_2 \leq \alpha\big\}\,\mathrm{d}\theta\bigg)^2 \qquad\qquad \text{(EC.15)} \\
&= \frac{1}{U_\rho}\sum_{k=1}^N \bigg(\frac{\beta_k}{z^k_\alpha}\bigg)^2 \int_{\mathcal{U}_k(\alpha)} \Big(\mathbb{1}\big\{\|\theta-\theta_k\|_2 \leq \alpha\big\}\rho(\mathrm{d}\theta)\Big)^2 \\
&\geq \frac{1}{U_\rho}\sum_{k=1}^N \bigg(\frac{\beta_k}{z^k_\alpha}\bigg)^2 \bigg(\int_{\mathcal{U}_k(\alpha)} \mathbb{1}\big\{\|\theta-\theta_k\|_2 \leq \alpha\big\}\rho(\mathrm{d}\theta)\bigg)^2 \\
&= \frac{1}{U_\rho}\sum_{k=1}^N \bigg(\frac{\beta_k}{z^k_\alpha}\bigg)^2 \bigg(\int_\Theta \mathbb{1}\big\{\|\theta-\theta_k\|_2 \leq \alpha\big\}\rho(\mathrm{d}\theta)\bigg)^2 \\
&= \frac{1}{U_\rho}\sum_{k=1}^N (\beta_k)^2,
\end{aligned}$$

where the first inequality above is valid since $\rho(\theta) \leq U_\rho$ for all $\theta \in \Theta$ by Assumption 2; the second equality is valid since all indicator functions take zero value for $\theta \in \Theta\backslash\mathcal{U}$; the third equality is followed by the definition of $\mathcal{U}$ and the fact that $\mathcal{U}_k$ are mutually exclusive; the fourth equality follows from $\mathbb{1}\big\{\|\theta-\theta_k\|_2 \leq \alpha\big\} = 0$ that holds for all $\theta \in \Theta\backslash\mathcal{U}_k$; the second inequality is obtained by the Jensen's inequality; the last equality is valid via the definition of z^k_α. Using $\|\boldsymbol{b}^\varepsilon_\alpha - \boldsymbol{b}_\alpha\|_{2,\rho} \leq \xi$, (EC.15), and $\|\boldsymbol{\beta}\|_1 \leq \sqrt{N}\|\boldsymbol{\beta}\|_2$ for $\boldsymbol{\beta} \in \mathbb{R}^N$, we have

$$\sum_{i=1}^N |\beta_i| \leq \sqrt{N}\Bigg[\sum_{i=1}^N (\beta_i)^2\Bigg]^{1/2} \leq \sqrt{NU_\rho}\|\boldsymbol{b}_\alpha\|_{2,\rho} \leq \sqrt{NU_\rho}\big(\|\boldsymbol{b}^\varepsilon_\alpha\|_{2,\rho} + \xi\big) \leq \sqrt{NU_\rho}\big(\|\boldsymbol{b}^\varepsilon\|_{2,\rho} + \xi\big), \tag{EC.16}$$

where the last inequality follows from the Pythagorean identity in Lemma EC.4. Using the inequalities (EC.13), (EC.14) and (EC.16), we obtain $\|V^{\varepsilon}_{\alpha} - V_N\|_{\infty} \le \big\|V^{\varepsilon}_{\alpha} - V_{\alpha}\big\|_{\infty} + \|V_{\alpha} - V_N\|_{\infty} \le \sqrt{U_{\rho}}\xi + \alpha\sqrt{NU_{\rho}}\mathrm{L}_{\varphi}(\mathrm{diam}(\mathcal{S})+1)(\|\boldsymbol{b}^{\varepsilon}\|_{2,\rho} + \xi)$. □

Lemma EC.6. *Given $\varepsilon > 0$ and $\alpha \in \big(0, \min_{i\neq j}\|\theta_i - \theta_j\|_2\big)$, for any $H \ge H_{\varepsilon} := \big\lceil \varepsilon^{-2}\big\|\boldsymbol{b}^{\varepsilon} - \boldsymbol{b}^{\varepsilon}_{\alpha}\big\|^2_{\infty,\rho}(\Omega + \Delta_{\delta})^2\big\rceil$, there exists a function $V_H \in \mathcal{W}(\Phi_H)$ such that*

$$\big\|\overset{\perp}{V}{}^{\varepsilon}_{\alpha} \ - \ V_H\big\|_{\infty} \le \varepsilon,$$

with a probability of at least $1-\delta$, where $\overset{\perp}{V}{}^{\varepsilon}_{\alpha}(\cdot) = \big\langle \overset{\perp}{\boldsymbol{b}}{}^{\varepsilon}_{\alpha}, \varphi(\cdot)\big\rangle$ is defined in Lemma EC.4.

Proof. The proof of this lemma is similar to the proof of Lemma EC.2. In particular, consider $\overset{\perp}{V}{}^{\varepsilon}_{\alpha}(\cdot) = \big\langle \overset{\perp}{\boldsymbol{b}}{}^{\varepsilon}_{\alpha}, \varphi(\cdot)\big\rangle$ defined in Lemma EC.4 where $\big\|\overset{\perp}{\boldsymbol{b}}{}^{\varepsilon}_{\alpha}\big\|_{\infty,\rho} < \infty$. Theorem 3.2 in Rahimi and Recht (2008) ensures there exists a function $V_H \in \mathcal{W}(\Phi_H)$, such that

$$\big\|\overset{\perp}{V}{}^{\varepsilon}_{\alpha} - V_H\big\|_{\infty} \le \frac{\big\|\overset{\perp}{\boldsymbol{b}}{}^{\varepsilon}_{\alpha}\big\|_{\infty,\rho}}{\sqrt{H}}\left(\Delta_{\delta} + \Omega\right),$$

where Δ_{δ} and Ω are defined in §3.1. When $H \ge \left\lceil \varepsilon^{-2}\ \big\|\overset{\perp}{\boldsymbol{b}}{}^{\varepsilon}_{\alpha}\big\|^2_{\infty,\rho}\ \left(\Delta_{\delta} + \Omega\right)^2\right\rceil$, we can then guarantee $\big\|\overset{\perp}{V}{}^{\varepsilon}_{\alpha} - V_H\big\|_{\infty} \le \varepsilon$ with a probability of at least $1-\delta$. This inequality also holds for $H \ge H_{\varepsilon}$ since $\overset{\perp}{\boldsymbol{b}}{}^{\varepsilon}_{\alpha} = \boldsymbol{b}^{\varepsilon} - \boldsymbol{b}^{\varepsilon}_{\alpha}$. □

Remark EC.2. Given $\alpha \in \big(0, \min_{i\neq j}\|\theta_i - \theta_j\|_2\big)$, norm $\big\|\boldsymbol{b}^{\varepsilon} - \boldsymbol{b}^{\varepsilon}_{\alpha}\big\|_{\infty,\rho}$ used in H_{ε} is finite.

Now, we integrate our results in Lemmas EC.5 and EC.6 to prove Theorem 2.

Proof of Theorem 2. Consider $(b^{\varepsilon}_0, \boldsymbol{b}^{\varepsilon})$ given in Part (i) of Proposition 1 and its corresponding function V^{ε}. Using Lemma EC.4, we can decompose V^{ε} as $V^{\varepsilon} = V^{\varepsilon}_{\alpha} + \overset{\perp}{V}{}^{\varepsilon}_{\alpha}$. Let $V = V_N + V_H$ where $V_N \in \mathcal{W}(\Phi_N)$ and $V_H \in \mathcal{W}(\Phi_H)$ are defined in Lemmas EC.5 and EC.6, respectively. In addition, assume $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{N+H})$ are the coefficients defining V. We observe that $V \in \mathcal{W}(\Phi_N \cup \Phi_H)$. When $H \ge H_{\varepsilon/3}$ (we use $\varepsilon/3$ in lieu of ε in Lemma EC.6), with a probability of at least $1-\delta$, we have

$$\begin{aligned}
\big\|V^{\varepsilon} - V\big\|_{\infty} \ &= \ \big\|V^{\varepsilon}_{\alpha} + \overset{\perp}{V}{}^{\varepsilon}_{\alpha} - (V_N + V_H)\big\|_{\infty} \\
&\le \ \big\|V^{\varepsilon}_{\alpha} - V_N\big\|_{\infty} + \big\|\overset{\perp}{V}{}^{\varepsilon}_{\alpha} - V_H\big\|_{\infty} \\
&\le \ \sqrt{U_{\rho}}\xi + \alpha\sqrt{NU_{\rho}}\mathrm{L}_{\varphi}(\mathrm{diam}(\mathcal{S})+1)\big(\|\boldsymbol{b}^{\varepsilon}_{\alpha}\|_{2,\rho} + \xi\big) + \frac{\varepsilon}{3} \qquad \text{(EC.17)} \\
&\le \ \xi\sqrt{U_{\rho}}\Big(1 + \frac{\varepsilon}{\Omega'}\mathrm{L}_{\varphi}(\mathrm{diam}(\mathcal{S})+1)\Big) + \varepsilon\Big(\frac{1}{3} + \frac{\sqrt{U_{\rho}}}{\Omega'}\mathrm{L}_{\varphi}(\mathrm{diam}(\mathcal{S})+1)\|\boldsymbol{b}^{\varepsilon}_{\alpha}\|_{2,\rho}\Big) \\
&\le \varepsilon
\end{aligned}$$

where the third inequality follows from the choice of α in Theorem 2 and the last one from the definition of Ω' in §4.2 and choosing $\xi \in \left(0,\ \frac{\varepsilon}{3\sqrt{U_{\rho}}(1+\varepsilon\mathrm{L}_{\varphi}(\mathrm{diam}(\mathcal{S})+1)/\Omega')}\right]$. Consider the vector

$\boldsymbol{\beta}' = (\beta_0 - \Gamma\varepsilon, \beta_1, \ldots, \beta_{N+H})$. This vector is feasible to constraints (6) of $\text{FGLP}_{(\text{N+H})}$ with a probability of at least $1-\delta$ since it holds that

$$\begin{aligned}(1-\gamma)\big(\beta_0 - \Gamma\varepsilon\big) + \sum_{i=1}^{N+H}\beta_i\big(\varphi(s;\theta_i) - \gamma\mathbb{E}[\varphi(s';\theta_i)|s,a]\big) &= V(s) - \varepsilon - \gamma\mathbb{E}\big[V(s') + \varepsilon\big|s,a\big]\\ &\le V^{\varepsilon}(s) - \gamma\mathbb{E}[V^{\varepsilon}(s')|s,a]\\ &\le c(s,a),\end{aligned}$$

for all $(s,a)\in\mathcal{S}\times\mathcal{A}_s$ where the first inequality follows from (EC.17) and the last one from the feasibility of $(b_0^{\text{FE}}, \boldsymbol{b}^{\text{FE}})$ to FELP. Also, with a probability of at least $1-\delta$, the vector $\boldsymbol{\beta}'$ violates constraints (7) of $\text{FGLP}_{(\text{N+H})}$ by at most $\frac{4\varepsilon}{(1+\gamma)}$. In particular, with a probability of at least $1-\delta$, Part (i) of Proposition 1 and Lemma EC.1 guarantee that

$$V(s;\boldsymbol{\beta}_{\text{N}}^{\text{FG}}) \le V^*(s) \le V^{\varepsilon}(s) + \frac{2\varepsilon}{1-\gamma} \le (V(s) - \Gamma\varepsilon) + \varepsilon(1+\Gamma) + \frac{2\varepsilon}{(1-\gamma)} = V(s;\boldsymbol{\beta}') + \frac{4\varepsilon}{(1-\gamma)}. \quad \text{(EC.18)}$$

In (EC.18), we used inequality (EC.17) to derive $\|V^* - (V - \Gamma\varepsilon)\|_\infty \le \|V^* - V\|_\infty + \Gamma\varepsilon \le (1+\Gamma)\varepsilon$ and $V(s;\boldsymbol{\beta}') = V(s) - \Gamma\varepsilon$. In addition, (EC.18) together with the fact that V^* is a pointwise upper bound on $V(\boldsymbol{\beta}')$ (by Lemma EC.1) ensure that $\big\|V^* - V(\boldsymbol{\beta}')\big\|_\infty \le \frac{4\varepsilon}{(1-\gamma)}$ which holds with a probability of at least $1-\delta$ when $H \ge H_{\varepsilon/3} = \big\lceil 9\varepsilon^{-2}\big\|\boldsymbol{b}^{\varepsilon} - \boldsymbol{b}_{\alpha}^{\varepsilon}\big\|_{\infty,\rho}^2\,(\Omega + \Delta_\delta)^2\big\rceil$. □

EC.2. Addendum to §3.1: Constant factor in FALP sampling bound

In this section, we derive an FALP sampling bound without using the property that its VFA is a pointwise lower bound on V^*, which makes apparent the sharper constant that we obtain in Theorem 1 by using this FALP VFA property. Proposition EC.1 leverages Part (ii) of Lemma EC.2 alone to establish a sampling bound for FALP.

PROPOSITION EC.1. *Fix $\varepsilon > 0$ and $\delta \in (0,1]$. Then for any $N \ge N_\varepsilon$, where N_ε is defined in* (EC.1)*, and any optimal solution $\boldsymbol{\beta}_{\text{N}}^{\text{FA}}$ to $\text{FALP}_{(\text{N})}$, it holds that*

$$\big\|V^* - V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\big\|_{1,\nu} \;\le\; \frac{4\varepsilon}{(1-\gamma)}$$

with a probability of at least $1-\delta$.

Proof. Part (ii) of Lemma EC.2 ensures that vector $(\beta_0^{\text{FEAS}} - \Gamma\varepsilon, \beta_1^{\text{FEAS}}, \ldots, \beta_N^{\text{FEAS}})$ for $N \ge N_\varepsilon$, is feasible to $\text{FALP}_{(\text{N})}$ with a probability of at least $1-\delta$ and $\big\|V^{\text{FE}} - \big(V(\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon\big)\big\|_\infty \le {}^{2\varepsilon}/_{(1-\gamma)}$ with the same probability. Using Part (ii) of Proposition 1, it holds that

$$\begin{aligned}\big\|V^* - V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\big\|_{1,\nu} &\le \big\|V^* - V^{\text{FE}}\big\|_{1,\nu} + \big\|V^{\text{FE}} - V(\boldsymbol{\beta}_{\text{N}}^{\text{FA}})\big\|_{1,\nu}\\ &\le \frac{2\varepsilon}{(1-\gamma)} + \big\|V^{\text{FE}} - \big(V(\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon\big)\big\|_{1,\nu} \qquad \text{(EC.19)}\\ &\le \frac{4\varepsilon}{(1-\gamma)}\end{aligned}$$

with a probability of at least $1-\delta$, where the second inequality is valid since $(\beta_0^{\text{FEAS}} - \Gamma\varepsilon, \beta_1^{\text{FEAS}}, \ldots, \beta_N^{\text{FEAS}})$ and $\boldsymbol{\beta}_{\text{N}}^{\text{FA}}$ are feasible and optimal to FALP$_{(\text{N})}$, respectively, and the last inequality follows from $\left\|V^{\text{FE}} - \big(V(\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon\big)\right\|_{1,\nu} \le \left\|V^{\text{FE}} - \big(V(\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon\big)\right\|_{\infty} \le 2\varepsilon/(1-\gamma)$. □

The sampling bound in Proposition EC.1 minus the sampling bound in Theorem 1 equals

$$\varepsilon^{-2} \left\|\boldsymbol{b}^{\text{FE}}\right\|_{\infty,\rho}^2 \frac{(1-\gamma)\Omega}{2} \left(\frac{3+\gamma}{2}\Omega + 2\Delta_\delta\right).$$

In other words, the latter bound is tighter than the former one. This tightening is because we leveraged that V^* is a state-wise upper bound on any continuous function satisfying constraints (2) when obtaining the inequalities (EC.7) in the proof of Theorem 1. In contrast, this property is not used in the analogous inequalities (EC.19) in Proposition EC.1, where we directly employ the results in Part (ii) of Lemma EC.2.

EC.3. Addendum to §4.2: Applying FALP sampling analysis to FGLP

In this section, we show that the direct application of the FALP sampling bound analysis to FGLP leads to a sampling bound that is weak and does not account for the quality of the self-guiding constraints in an insightful manner. To directly apply the analysis used for FALP to FGLP, we require that $V(s;\boldsymbol{\beta}_{\text{N}}^{\text{FG}})$ is κ_{N} far from to $V^*(s)$, that is, $\min_{s\in\mathcal{S}} |V^*(s) - V(s;\boldsymbol{\beta}_{\text{N}}^{\text{FG}})| \ge \kappa_{\text{N}} > 0$. The positivity of κ_{N} may not be true when $V^*(\hat{s}) = V(\hat{s};\boldsymbol{\beta}_{\text{N}}^{\text{FG}})$ for a state $\hat{s}\in\mathcal{S}$. This is thus a restrictive assumption. Proposition EC.2 states a bound on the number of samples M that follows directly from Proposition EC.1 and is analogous to the number of samples $N+H$ in §4.2.

Proposition EC.2. *Suppose we have an optimal solution $\boldsymbol{\beta}_{\text{N}}^{\text{FG}}$ to FGLP$_{(\text{N})}$ such that $\kappa_{\text{N}} > 0$. Given $\varepsilon > 0$ and $\delta \in (0,1]$, if*

$$M \ge \left\lceil \min\{\varepsilon,\kappa_{\text{N}}\}^{-2} \left(\frac{4}{1-\gamma}\right)^2 \left\|\boldsymbol{b}^{\text{FE}}\right\|_{\infty,\rho}^2 (\Omega+\Delta_\delta)^2 \right\rceil,$$

then any optimal solution $\boldsymbol{\beta}_{\text{M}}^{\text{FG}}$ to FGLP$_{(\text{M})}$ satisfies

$$\left\|V^* - V(\boldsymbol{\beta}_{\text{M}}^{\text{FG}})\right\|_{1,\nu} \le \min\{\varepsilon,\kappa_{\text{N}}\},$$

with a probability of at least $1-\delta$.

Proof. Let $\varepsilon' = (1-\gamma)\min\{\varepsilon,\kappa_{\text{N}}\}/4$. Using Part (ii) of Lemma EC.2 with the choice of ε set to ε', we have that for any

$$M \ge N_{\varepsilon'} = \left\lceil \min\{\varepsilon,\kappa_{\text{N}}\}^{-2} \left(\frac{4}{1-\gamma}\right)^2 \left\|\boldsymbol{b}^{\text{FE}}\right\|_{\infty,\rho}^2 (\Omega+\Delta_\delta)^2 \right\rceil,$$

vector $(\beta_0^{\text{FEAS}} - \Gamma\varepsilon', \beta_1^{\text{FEAS}}, \ldots, \beta_M^{\text{FEAS}})$ is feasible to FALP$_{(\text{M})}$ and satisfies $\left\|V^{\text{FE}} - \big(V(\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon'\big)\right\|_{\infty} \le {}^{2\varepsilon'}\!/\!{}_{(1-\gamma)}$ with a probability of at least $1-\delta$. Next, by leveraging Part (ii) of Proposition 1 with ε chosen as ε', we can write

$$\left\|V^* - \left(V(\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon'\right)\right\|_\infty \le \frac{2\varepsilon'}{(1-\gamma)} + \left\|V^{\text{FE}} - \left(V(\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon'\right)\right\|_\infty \le \frac{4\varepsilon'}{(1-\gamma)} = \min\{\varepsilon, \kappa_{\text{N}}\} \le \kappa_{\text{N}},$$

which holds with a probability of at least $1-\delta$. Thus, for all $s \in \mathcal{S}$, we obtain $V(s;\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon' \ge V^*(s) - \kappa_{\text{N}}$, and from the definition of κ_{N}, we have $V^*(s) - \kappa_{\text{N}} \ge V(s;\boldsymbol{\beta}^{\text{FG}}_{\text{N}})$. Hence, it holds that

$$V(s;\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon' \;\ge\; V^*(s) - \kappa_{\text{N}} \;\ge\; V(s;\boldsymbol{\beta}^{\text{FG}}_{\text{N}}), \tag{EC.20}$$

for all $s \in \mathcal{S}$ with a probability of at least $1-\delta$. This shows that for $M \ge N_{\varepsilon'}$, the vector $(\beta^{\text{FEAS}}_0 - \Gamma\varepsilon', \beta^{\text{FEAS}}_1, \ldots, \beta^{\text{FEAS}}_M)$ is feasible to constraints (7) and (6) of $\text{FGLP}_{(\text{M})}$ and it satisfies $\left\|V^* - \left(V(\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon'\right)\right\|_\infty \le \min\{\varepsilon, \kappa_{\text{N}}\}$, where these statements hold with probability at least $1-\delta$. Therefore, with the same probability, an optimal FGLP solution $\boldsymbol{\beta}^{\text{FG}}_{\text{M}}$ has a smaller $(1,\nu)$ difference with respect to V^*, that is, we have

$$\left\|V^* - V(\boldsymbol{\beta}^{\text{FG}}_{\text{M}})\right\|_{1,\nu} \le \left\|V^* - (V(\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon)\right\|_{1,\nu} \le \left\|V^* - (V(\boldsymbol{\beta}^{\text{FEAS}}) - \Gamma\varepsilon)\right\|_\infty \le \min\{\kappa_{\text{N}}, \varepsilon\}.$$

□

The sampling lower bound in Proposition EC.2 is similar to the FALP bound but has two key differences: (i) it has an additional constant $(4/(1-\gamma))^2$ and (ii) ε is replaced by $\min\{\kappa_{\text{N}}, \varepsilon\}$. The additional constant $(4/(1-\gamma))^2$ stems from constructing in inequality (EC.20) a feasible solution to the self-guiding constraints. The intuition behind replacement of ε by $\min\{\kappa_{\text{N}}, \varepsilon\}$ is as follows. We assumed that $\min_{s\in\mathcal{S}} |V^*(s) - V(s;\boldsymbol{\beta}^{\text{FG}}_{\text{N}})| \ge \kappa_{\text{N}}$, that is, $V(s;\boldsymbol{\beta}^{\text{FG}}_{\text{N}})$ is below V^* by at least κ_N at all states. Therefore, a conservative approach to satisfy the self-guiding constraints is to sample sufficiently many random basis functions such that $V(s;\boldsymbol{\beta}^{\text{FG}}_{\text{M}})$ is within $\min\{\kappa_{\text{N}}, \varepsilon\}$ of $V^*(s)$ at all states.

EC.4. Addendum to Computational Study

In §EC.4.1, we discuss our sampling approach to compute valid lower bounds for constraint-sampled ALPs, that is used in §5 to compute bounds for constraint-sampled FALP and FGLP. We also provide our additional numerical study on generalized joint replenishment application in §EC.4.2.

EC.4.1. A Valid Lower Bound Estimate for Constraint-sampled ALPs

This material discusses how ideas in Lin et al. (2019) can be leveraged to estimate a valid lower bound while using a constraint-sampled version of a generic ALP, and in particular, the FALP and FGLP models in this paper. For any VFA $V(\boldsymbol{\beta})$, we define the function

$$y(s,a;\boldsymbol{\beta}) := \mathbb{E}_\chi[V(\boldsymbol{\beta})] + \frac{1}{1-\gamma}\Big(c(s,a) + \gamma\mathbb{E}\big[V(s';\boldsymbol{\beta}) \mid s,a\big] - V(s;\boldsymbol{\beta})\Big),$$

that encodes the violation of ALP constraints for a given $\boldsymbol{\beta}$ at a state-action pair (s,a). Suppose $\boldsymbol{\beta}^{\text{CS-FA}}_{\text{N}}$ is an optimal solution of a constraint-sampled $\text{FALP}_{(\text{N})}$. We observe that minimizing the function $y(s,a;\boldsymbol{\beta}^{\text{CS-FA}}_{\text{N}})$ over state-action pairs corresponds to finding the most violating constraint in the constraint-sampled $\text{FALP}_{(\text{N})}$ with the optimal solution $\boldsymbol{\beta}^{\text{CS-FA}}_{\text{N}}$ since term $\mathbb{E}_\chi[V(\boldsymbol{\beta})]$ is independent of the state and action and the term $(c(s,a) + \gamma\mathbb{E}[V(s';\boldsymbol{\beta})|s,a] - V(s;\boldsymbol{\beta}))/(1-\gamma)$ is the constraint slack. Thus, if the minimum value of function $y(s,a;\boldsymbol{\beta}_{\mathrm{N}}^{\text{CS-FA}})$ over state-action pairs is strictly less than $\mathbb{E}_{\chi}[V(\boldsymbol{\beta}_{\mathrm{N}}^{\text{CS-FA}})]$, then $\boldsymbol{\beta}_{\mathrm{N}}^{\text{CS-FA}}$ violates a constraint of $\mathrm{FALP}_{(\mathrm{N})}$. Otherwise, $\boldsymbol{\beta}_{\mathrm{N}}^{\text{CS-FA}}$ is feasible to $\mathrm{FALP}_{(\mathrm{N})}$. Under mild conditions, function y is Lipschitz with constant $\mathrm{L}_y > 0$.

Lemma EC.7 is directly based on Lemma EC.3 in Lin et al. 2019 and provides a lower bound on the optimal cost. For a given VFA $V(\boldsymbol{\beta})$ and $\lambda \in (0,1]$, we define a measure Y on $\mathcal{S}\times\mathcal{A}_s$ as $Y(s,a;\boldsymbol{\beta},\lambda) := \exp({}^{-y(s,a;\boldsymbol{\beta})}/_{\lambda})$.

Lemma EC.7 **(Lemma EC.3, Lin et al. 2019)**. *For all $\lambda \in (0,1]$ and $\boldsymbol{\beta}$, we have* $\mathrm{PC}(\pi^*) \geq \mathbb{E}_Y\big[y(s,a;\boldsymbol{\beta})\big] + \lambda(\Lambda + d_{\mathcal{S}\times\mathcal{A}}\ln(\lambda))$ *where*

$$\Lambda := -\ln\left[\Gamma\left(1+\frac{d_{\mathcal{S}\times\mathcal{A}}}{2}\right)\left(R_{\mathcal{S}\times\mathcal{A}_s}\sqrt{\pi}\right)^{-d_{\mathcal{S}\times\mathcal{A}}}\int_{\mathcal{S}\times\mathcal{A}}\mathrm{d}(s,a)\right] - \mathrm{L}_y(R_{\mathcal{S}\times\mathcal{A}_s} + \mathrm{diam}(\mathcal{S}\times\mathcal{A})),$$

and $d_{\mathcal{S}\times\mathcal{A}}$ is $d_{\mathcal{S}} + d_{\mathcal{A}}$. Function Γ is the standard gamma function, π is the Archimedes constant, $R_{\mathcal{S}\times\mathcal{A}_s} > 0$ is the radius of the largest ball contained in $\mathcal{S}\times\mathcal{A}$, and $\mathrm{diam}(\mathcal{S}\times\mathcal{A})$ is the diameter of $\mathcal{S}\times\mathcal{A}$.

Given a solution $\boldsymbol{\beta}_{\mathrm{N}}^{\text{CS-FA}}$, Lemma EC.7 suggests that a valid lower bound can be computed by estimating the expected value $\mathbb{E}_Y\big[y(s,a;\boldsymbol{\beta})\big]$ and a constant term. For our numerical experiments in §5, we estimate $\mathbb{E}_Y\big[f(\boldsymbol{\beta},s,a)\big]$ using the Metropolis-Hastings method with 4000 samples by generating 8 Markov Chains, each with length of 1500, where we burn the first 1000 samples and use the last 500. Parameter Λ can be easily evaluated for the instances in Table 1 since the perishable inventory control cost function is Lipschitz with constant $\mathrm{L}_c > 0$. In fact, it is easy to verify that $\mathrm{L}_c = 2(\gamma^L c_o\bar{a} + c_h\bar{a} + c_b\underline{s} + c_d\bar{a} + c_l\bar{a})$ and consequently, $\mathrm{L}_y = {}^{(4\|\boldsymbol{\beta}\|_1+\mathrm{L}_c)}/_{1-\gamma}$. We choose the other parameters defining Λ as follows: $d_{\mathcal{S}\times\mathcal{A}} = 4, R_{\mathcal{S}\times\mathcal{A}_s} = {}^{\bar{a}}/_{2}$, and $\mathrm{diam}(\mathcal{S}\times\mathcal{A}) = 3\bar{a}^2 + (\underline{s}-\bar{a})^2$. We set $\lambda = {}^{1}/_{(\Lambda+d_{\mathcal{S}\times\mathcal{A}})}$ but one can cross-validate this parameter to possibly obtain tighter bounds.

EC.4.2. Generalized Joint Replenishment

In this section, we test the effectiveness of ALPs with random basis functions for solving the generalized joint replenishment (GJR) problem considered in Adelman and Klabjan (2012, henceforth abbreviated AK). In §EC.4.2.1, we describe GJR and its average-cost semi-MDP formulation from AK. We provide the extension of the FALP sampling bound for average-cost semi-MDPs in §EC.4.2.2. In §EC.4.2.3, we summarize our methods, as well as, an adaptive basis function generation approach and a set of instances, both from AK. Then, in §EC.4.2.4, we elaborate on GJR constraint generation and greedy policy optimization. In §EC.4.2.5, we discuss our numerical findings.

EC.4.2.1. Average-cost Semi-MDP Formulation. The GJR problem involves the replenishment of a collection of products that are consumed at a fixed and deterministic rate and are coupled via a shared replenishment capacity (Adelman and Klabjan 2012). We describe the average-cost semi-MDP formulation for this problem from AK.

Consider managing the replenishment of inventories across J products over a continuous time horizon with index set $\{1,2,\ldots,J\}$. Each product j is consumed at a finite and deterministic rate $\lambda_j > 0$ and we denote by $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_J)$ the vector of these rates. A state vector $s = (s_1, s_2, \ldots, s_J)$ encodes the inventory levels of these items all measured in normalized units, where each component $s_j \geq 0$ is non-negative for all $j \in \{1,2,\ldots,J\}$. A zero value for the j-th state component signals that the j-th item is stocked out. Since the replenishment time can be postponed if no item is currently stocked out, it can be assumed that at least one item has zero inventory in the state. Thus, the state space is given by $\mathcal{S} := \{s : 0 \leq s \leq \bar{s},\ \ s_j = 0 \text{ for some } j \in \{1,2,\ldots,J\}\}$, where $\bar{s} \in (0,\infty)^J$ is a vector of maximum inventory levels. The replenishment decision is specified by $a \in \mathbb{R}_+^J$. This decision at a given state $s \in \mathcal{S}$ belongs to $\mathcal{A}_s := \left\{a \in \mathbb{R}_+^J : s + a \leq \bar{s},\ \sum_{j=1}^J a_j \leq \bar{a}\right\}$. Here $\bar{a} \in \mathbb{R}_+$ denotes a capacity constraint on the total amount of joint replenishment. The immediate cost $c(s,a)$ of an action a at state s has (i) a fixed component $c_{\mathrm{supp}(a)}$ that depends on the set of items replenished $\mathrm{supp}(a) := \{j \in \{1,\ldots,J\} | a_j > 0\}$, and (ii) a variable holding cost component $\sum_{j=1}^J (2s_j a_j + a_j^2) h_j / 2\lambda_j$, where h_j denotes the holding cost per unit per time. Since the usage rate is deterministic, the time till next replenishment is defined by $T(s,a) := \min_j \{(s_j + a_j)/\lambda_j\}$ and the system transitions to a new state $s' = s + a - T(s,a)\lambda$.

To find a deterministic and stationary policy $\pi : \mathcal{S} \mapsto \mathcal{A}_s$, one can in theory solve the semi-MDP optimality equations (see, e.g., Theorem 10.3.6 in Hernández-Lerma and Lasserre 1999)

$$u(s) = \inf_{a \in \mathcal{A}_s} \{c(s,a) - \eta T(s,a) + u(s')\}, \qquad \forall s \in \mathcal{S} \tag{EC.21}$$

where $\eta \in \mathbb{R}$ denotes the long-run optimal average cost and $u(\cdot)$ is a bias function that captures state-dependent transient costs. The action prescribed at a state $s \in \mathcal{S}$ by an optimal deterministic and stationary policy can be found by solving the infimum in the right hand side of (EC.21). Moreover, the optimality equations (EC.21) have the following linear programming representation

$$\sup_{(\eta',u') \in \mathbb{R} \times \mathcal{C}} \quad \eta' \tag{EC.22}$$

$$\eta' T(s,a) + u'(s) - u'(s') \leq c(s,a), \qquad \forall (s,a) \in \mathcal{S} \times \mathcal{A}_s. \tag{EC.23}$$

This infinite linear program is the average-cost analogue of ELP (see §2.1). However, note that there is no need to specify a state-relevance distribution such as ν in this case.

EC.4.2.2. FALP Sampling Bound for Average-cost Semi-MDPs. In this section, we develop an FALP sampling bound for an average-cost semi-MDP given by linear program (EC.22)-(EC.23). Suppose B denotes the sampling batch size. For a number of sampled random basis functions $N \geq 2B$, the average-cost analogue of $\mathrm{FGLP}_{(\mathrm{N})}$ is

$$
\begin{aligned}
\sup_{\eta',\boldsymbol{\beta}} \quad & \eta' \\
& \eta' T(s,a) + \sum_{j=1}^{J} \beta_{1,j}(s'_j - s_j) + \sum_{i=1}^{N} \beta_{2,i}\big(\varphi(s';\theta_i) - \varphi(s;\theta_i)\big) \ \le\ c(s,a), \qquad \forall (s,a)\in\mathcal{S}\times\mathcal{A}_s, \\
& \beta_0 - \sum_{j=1}^{J} \beta_{1,j} s_j - \sum_{i=1}^{N} \beta_{2,i}\varphi(s;\theta_j) \ \ge\ u(s;\boldsymbol{\beta}^{\text{FG-AC}}_{\text{N-B}}), \qquad \forall s\in\mathcal{S}.
\end{aligned}
$$

where vector $(\eta^{\text{FG-AC}},\boldsymbol{\beta}^{\text{FG-AC}}_{\text{N-B}})$ is an optimal solution to $\text{FGLP}_{(\text{N-B})}$ and $u(\cdot;\boldsymbol{\beta}^{\text{FG-AC}}_{\text{N-B}}) = -\infty$ for $N = B$. Indeed, the analogue of $\text{FALP}_{(\text{N})}$ can be obtained by removing the second set of self-guiding constraints from the above linear program. The average-cost $\text{FALP}_{(\text{N})}$ is thus

$$
\begin{aligned}
\sup_{\eta',\boldsymbol{\beta}} \quad & \eta' \\
& \eta' T(s,a) + \sum_{j=1}^{J} \beta_{1,j}(s'_j - s_j) + \sum_{i=1}^{N} \beta_{2,i}\big(\varphi(s';\theta_i) - \varphi(s;\theta_i)\big) \ \le\ c(s,a), \qquad \forall (s,a)\in\mathcal{S}\times\mathcal{A}_s.
\end{aligned}
$$

For the theoretical results in this section, we require the following assumptions to hold for the exact linear program and the average-cost FALP. Such assumptions are standard in the literature (see, e.g., Lemma 4.1 of Klabjan and Adelman 2007).

ASSUMPTION EC.2. *There is a feasible solution $(\eta^{\text{S}}, u^{\text{S}})$ to the linear program* (EC.22)-(EC.23) *such that*

$$\zeta := \inf_{(s,a)\in\mathcal{S}\times\mathcal{A}_s} \{c(s,a) - \eta^{\text{S}} T(s,a) - u^{\text{S}}(s) + u^{\text{S}}(s')\},$$

is strictly positive. Moreover, solution u^{S} can be obtained in set $\mathcal{R}_\infty(\varphi,\rho)$.

ASSUMPTION EC.3. *There is a feasible solution $(\eta^{\text{FA-S}},\boldsymbol{\beta}^{\text{FA-S}})$ to average-cost FALP such that*

$$\zeta^{\text{FA}} := \inf_{(s,a)\in\mathcal{S}\times\mathcal{A}_s} \left\{ c(s,a) - \eta^{\text{FA-S}} T(s,a) - \sum_{j=1}^{J} \beta^{\text{FA-S}}_{1,j}(s'_j - s_j) - \sum_{i=1}^{N} \beta^{\text{FA-S}}_{2,i}\big(\varphi(s';\theta_i) - \varphi(s;\theta_i)\big) \right\},$$

is strictly positive. Moreover, Assumption 2 holds.

Assumption EC.2 ensures that there is a Slater feasible point $(\eta^{\text{S}}, u^{\text{S}})$ to the average-cost exact linear program (EC.22)-(EC.23) such that all of its constraints are strictly satisfied. In addition, u^{S} belongs to set $\mathcal{R}_\infty(\varphi,\rho)$. Assumption EC.3 assumes a similar Slater condition for average-cost FALP as well as the standard assumptions on the random basis functions from the main text. Lemma EC.8 utilizes these Slater points to construct a feasible solution to program (EC.22)-(EC.23) starting from an ε-feasible solution.

LEMMA EC.8. *Given $\varepsilon > 0$, if solution (η,u) is ε-feasible to linear program* (EC.22)-(EC.23)*, then there is a feasible solution, denoted $(\hat\eta,\hat u)$, to this linear program such that*

$$|\eta - \hat\eta| \le \frac{\varepsilon}{(\zeta+\varepsilon)}|\eta - \eta^{\text{S}}| \quad \textit{and} \quad \|u - \hat u\|_\infty \le \frac{\varepsilon}{(\zeta+\varepsilon)}\|u - u^{\text{S}}\|_\infty. \tag{EC.24}$$

Proof. Let $R := \varepsilon/\zeta+\varepsilon \in (0,1)$ and define $R' := 1 - R = \zeta/\zeta+\varepsilon$. Since (η, u) is an ε-feasible to program (EC.22)-(EC.23), we have $\eta T(s,a) + u(s) - u(s') \leq c(s,a) + \varepsilon$ for all $(s,a) \in \mathcal{S} \times \mathcal{A}_s$. Also, from the definition of ζ, we have for all $(s,a) \in \mathcal{S} \times \mathcal{A}_s$ that $\eta^{\text{S}} T(s,a) + u^{\text{S}}(s) - u^{\text{S}}(s') \leq c(s,a) - \zeta$. Convex combination $(\hat{\eta}, \hat{u}) := \big(R\eta^{\text{S}} + R'\eta, Ru^{\text{S}} + R'u\big)$ fulfills

$$\begin{aligned}\big(R\eta^{\text{S}} + R'\eta\big) + \big(Ru^{\text{S}}(s) + R'u(s)\big) - \big(Ru^{\text{S}}(s') + R'u(s')\big) &\leq R\big(c(s,a) - \zeta\big) + R'\big(c(s,a) + \varepsilon\big) \\ &\leq (R + R')c(s,a) - R\zeta + R'\varepsilon \\ &\leq c(s,a)\end{aligned}$$

for all $(s,a) \in \mathcal{S} \times \mathcal{A}_s$. Thus, $(\hat{\eta}, \hat{u})$ is feasible to optimization (EC.22)-(EC.23) and satisfies (EC.24). □

Proposition EC.3 leverages Lemma EC.8 to tailor Part (i) of Proposition 1 to the deterministic and average-cost semi-MDPs. Let $(\eta^{\text{AC}}, u^{\text{AC}})$ be an optimal solution to the linear program (EC.22)-(EC.23).

Proposition EC.3. *Given* $\varepsilon > 0$, *there is an intercept* $\eta^{\text{FE-AC}} \in \mathbb{R}$ *and a function* $u^{\text{FE-AC}}(\cdot) = b_0^{\text{FE-AC}} + \big\langle \boldsymbol{b}^{\text{FE-AC}}, \varphi(\cdot)\big\rangle \in \mathcal{R}_\infty(\varphi, \rho)$ *such that* $(\eta^{\text{FE-AC}}, u^{\text{FE-AC}})$ *is feasible to optimization* (EC.22)-(EC.23) *and*

$$|\eta^{\text{AC}} - \eta^{\text{FE-AC}}| \leq \frac{\varepsilon}{\varepsilon + \zeta}|\eta^{\text{FE-AC}} - \eta^{\text{S}}| \qquad \textit{and} \qquad \|u^{\text{AC}} - u^{\text{FE-AC}}\|_\infty \leq \frac{\varepsilon}{\varepsilon + \zeta}\Big(\varepsilon + \|u^{\text{S}} - u^{\text{AC}}\|_\infty\Big).$$

Proof. Since u^{AC} is a continuous function and random basis function φ is universal by our Assumption 2, then Definition 1 guarantees the existence of a finite $C \geq 0$ and a function $\hat{u} \in \mathcal{R}_C(\varphi, \rho)$ such that $\|u^{\text{AC}} - \hat{u}\|_\infty \leq \varepsilon$. Using the feasibility (optimality) of u^{AC} to linear program (EC.22)-(EC.23), for all $(s,a) \in \mathcal{S} \times \mathcal{A}_s$, we have

$$c(s,a) \;\geq\; \eta^{\text{AC}} T(s,a) + u^{\text{AC}}(s) - u^{\text{AC}}(s') \;\geq\; \eta^{\text{AC}} T(s,a) + \hat{u}(s) - \hat{u}(s') - 2\varepsilon.$$

Thus, $(\eta^{\text{AC}}, \hat{u})$ is 2ε-feasible to program (EC.22)-(EC.23). By applying Lemma EC.8 to this 2ε-feasible solution, we obtain a feasible solution, denoted $(\eta^{\text{FE-AC}}, u^{\text{FE-AC}})$, to program (EC.22)-(EC.23) that satisfies

$$|\eta^{\text{AC}} - \eta^{\text{FE-AC}}| \leq \frac{\varepsilon}{\varepsilon + \zeta}|\eta^{\text{AC}} - \eta^{\text{S}}|,$$

as well as

$$\|u - u^{\text{FE-AC}}\|_\infty \leq \frac{\varepsilon}{\varepsilon + \zeta}\|u - u^{\text{S}}\|_\infty \leq \frac{\varepsilon}{\varepsilon + \zeta}\Big(\varepsilon + \|u^{\text{S}} - u^{\text{AC}}\|_\infty\Big),$$

where the last inequality is obtained using the triangle inequality and the $\|u^{\text{AC}} - u\|_\infty \leq \varepsilon$. Moreover, it is straightforward to verify that $u^{\text{FE-AC}} \in \mathcal{R}_\infty(\varphi, \rho)$ given that $u^{\text{S}} \in \mathcal{R}_\infty(\varphi, \rho)$ and $\hat{u} \in \mathcal{R}_C(\varphi, \rho)$. □

Proposition EC.4 establishes a sampling bound for the average-cost FALP and extends our FALP sampling bound for the discounted-cost MDPs reported in Theorem 1.

Proposition EC.4. *Given $\varepsilon > 0$, suppose $u^{\text{FE-AC}}$ and $\boldsymbol{b}^{\text{FE-AC}}$ follow their definitions in Proposition EC.3 and satisfy the statement of this proposition. Then for $\delta \in (0,1]$ and $N \geq N_{\varepsilon}^{\text{AC}} := \left\lceil \varepsilon^{-2} \left\| \boldsymbol{b}^{\text{FE-AC}} \right\|_{\infty,\rho}^{2} (\Omega + \Delta_\delta)^2 \right\rceil$, there is a feasible solution $(\eta^{\text{FA-AC}}, \boldsymbol{\beta}_{\text{N}}^{\text{FA-AC}})$ to the average-cost FALP$_{(\text{N})}$ such that*

$$|\eta^{\text{FE-AC}} - \eta^{\text{FA-AC}}| \leq \frac{\varepsilon}{(\zeta^{\text{FA}} + \varepsilon)} |\eta^{\text{FE-AC}} - \eta^{\text{FA-S}}|$$

and

$$\left\| u^{\text{FE-AC}} - u(\boldsymbol{\beta}^{\text{FA-AC}}) \right\|_\infty \leq \varepsilon + \frac{\varepsilon}{(\zeta^{\text{FA}} + \varepsilon)} \Big(\varepsilon + \| u^{\text{FE-AC}} - u(\boldsymbol{\beta}^{\text{FA-S}}) \|_\infty \Big)$$

with a probability of $1 - \delta$.

Proof. Similar to Part (i) of Lemma EC.2, we employ Theorem 3.2 of Rahimi and Recht (2008) to approximate the inner product $\left\langle \boldsymbol{b}^{\text{FE-AC}}, \varphi(\cdot) \right\rangle$ by sampling N random basis functions. This theorem ensures there is a function of the form $\sum_{i=1}^{N} \beta_i'' \varphi(s;\theta_i)$ (e.g., equation (5) in Rahimi and Recht 2008) such that

$$\left\| \left\langle \boldsymbol{b}^{\text{FE-AC}}, \varphi(s) \right\rangle - \sum_{i=1}^{N} \beta_i'' \varphi(s;\theta_i) \right\|_\infty \;\leq\; \frac{\|\boldsymbol{b}^{\text{FE-AC}}\|_{\infty,\rho}}{\sqrt{N}} (\Omega + \Delta_\delta)$$

with a probability of at least $1 - \delta$. Equivalently, for $N \geq N_\varepsilon^{\text{AC}}$, if we let $u''(s) = b_0^{\text{FE-AC}} + \sum_{i=1}^{N} \beta_i'' \varphi(s;\theta_i)$, then it holds that $\left\| u^{\text{FE-AC}} - u'' \right\|_\infty \leq \varepsilon$ with a probability of at least $1 - \delta$. With the same probability, since $(\eta^{\text{FE-AC}}, u^{\text{FE-AC}})$ is feasible to optimization (EC.22)-(EC.23), we can write

$$c(s,a) \;\geq\; \eta^{\text{FE-AC}} T(s,a) + u^{\text{FE-AC}}(s) - u^{\text{FE-AC}}(s') \geq\; \eta^{\text{FE-AC}} T(s,a) + u''(s) - u''(s') - 2\varepsilon,$$

which shows that $(\eta^{\text{FE-AC}}, u'')$ is 2ε-feasible to optimization (EC.22)-(EC.23). In addition, this shows that vector $(\eta^{\text{FE-AC}}, \beta_1'', \ldots, \beta_N'')$ is feasible to FALP$_{(\text{N})}$ with a probability of at least $1 - \delta$. Since we assume there is a Slater point $(\eta^{\text{FA-S}}, \boldsymbol{\beta}^{\text{FA-S}})$ for FALP$_{(\text{N})}$, Lemma EC.8 can be adapted to find a feasible solution $(\eta^{\text{FA-AC}}, \boldsymbol{\beta}^{\text{FA-AC}})$ to FALP$_{(\text{N})}$ (as a convex combination of the feasible solutions $(\eta^{\text{FA-S}}, \boldsymbol{\beta}^{\text{FA-S}})$ and $(\eta^{\text{FE-AC}}, \beta_1'', \ldots, \beta_N'')$) such that

$$|\eta^{\text{FE-AC}} - \eta^{\text{FA-AC}}| \leq \frac{\varepsilon}{(\zeta^{\text{FA}} + \varepsilon)} |\eta^{\text{FE-AC}} - \eta^{\text{FA-S}}|,$$

and

$$\|u'' - u(\boldsymbol{\beta}^{\text{FA-AC}})\|_\infty \leq \frac{\varepsilon}{(\zeta^{\text{FA}} + \varepsilon)} \|u'' - u(\boldsymbol{\beta}^{\text{FA-S}})\|_\infty \leq \frac{\varepsilon}{(\zeta^{\text{FA}} + \varepsilon)} \Big(\varepsilon + \|u^{\text{FE-AC}} - u(\boldsymbol{\beta}^{\text{FA-S}})\|_\infty \Big),$$

where the above approximation bound holds with a probability of at least $1 - \delta$. Thus, with the same probability, if we use the error bound $\left\| u^{\text{FE-AC}} - u'' \right\|_\infty \leq \varepsilon$ and the above approximation gap, we obtain

$$\left\| u^{\text{FE-AC}} - u(\boldsymbol{\beta}^{\text{FA-AC}}) \right\|_\infty \leq \left\| u^{\text{FE-AC}} - u'' \right\|_\infty + \left\| u'' - u(\boldsymbol{\beta}^{\text{FA-AC}}) \right\|_\infty \leq \varepsilon + \tfrac{\varepsilon}{(\zeta^{\text{FA}} + \varepsilon)} \Big(\varepsilon + \|u^{\text{FE-AC}} - u(\boldsymbol{\beta}^{\text{FA-S}})\|_\infty \Big).$$

□

It is possible to combine our results in Proposition EC.4 with the orthogonal projection ideas in §4 to construct an ε-feasible and ε-optimal solution to the average-cost FGLP. We omit a formal statement and proof of these results for brevity.

EC.4.2.3. Methods and Instances. Solving the infinite linear program (EC.22)-(EC.23) is intractable for the same reasons as ELP. AK thus replace the bias function $u(s)$ by an approximation to obtain an ALP. Their approximation has a (static) affine component $\beta_0 - \sum_{j=1}^{J} \beta_{1,j} s_j$ and an adaptive component $\sum_{i=1}^{I} \beta_{2,i} f^i(r^i s)$ with I terms, where $f^i : \mathbb{R} \mapsto \mathbb{R}$ is a piecewise linear ridge function and $r^i \in \mathbb{R}^J$ is a ridge vector. Putting these two components together gives the bias function approximation

$$u(s;\boldsymbol{\beta}) := \beta_0 - \sum_{j=1}^{J} \beta_{1,j} s_j - \sum_{i=1}^{I} \beta_{2,i} f^i(r^i s).$$

They also approximate η in (EC.22)-(EC.23), which is not needed for tractability, but facilitates managerial interpretation. This approximation is $\eta(\boldsymbol{\lambda}) = \hat{\eta} + \sum_{j=1}^{J} \beta_{1,j} \lambda_j$, where $\hat{\eta}$ is an intercept, $\beta_{1,j}$ can be interpreted here as marginal values associated with each item, and $\boldsymbol{\lambda} := (\lambda_1, \lambda_2, \ldots, \lambda_J)$. We refer to the resulting approximation of (EC.22)-(EC.23) as the ridge linear program (RLP). AK approach the solution of RLP using constraint generation, which involves solving mixed integer linear programs. In addition, they dynamically generate the ridge basis functions in the bias approximation via an approximation algorithm that that exploits the policy structure in the GJR application. We implemented RLP as a benchmark following the details in AK.

To study the effectiveness of random basis functions in this context, we derive an average-cost FGLP analogue starting from the exact linear program (EC.22)-(EC.23). To be consistent with AK, we use the same approximation $\eta(\boldsymbol{\lambda})$ for η and replace the bias function $u(s)$ by

$$u(s;\boldsymbol{\beta}) := \beta_0 - \sum_{j=1}^{J} \beta_{1,j} s_j - \sum_{i=1}^{N} \beta_{2,i} \varphi(s;\theta_i), \tag{EC.25}$$

where the adaptive basis function component in the RLP bias function approximation has been substituted with random basis functions. We select $\varphi(s;\theta)$ to be random stumps defined using the $\operatorname{sgn}(\cdot)$ function which returns -1, 0, and 1, respectively, if its argument is negative, zero, and positive. Specifically, $\varphi(s;\theta) = \operatorname{sgn}(s_q - \omega)$ where $\theta = (q, \omega)$, q is a random index uniformly distributed over the set $\{1, 2, \ldots, J\}$, and ω is uniformly distributed in the interval $[-\sigma, \sigma]$. We uniformly randomize the choice of σ over the interval $[1, \max_j \bar{s}_j]$. For this class of random basis functions, we show in EC.4.2.4 that FGLP can be solved using constraint generation where the separation problem to identify a violated constraint is a mixed integer linear program.

Our setup of RLP and FGLP thus differ mainly in how adaptive basis functions are generated. In the former approach, ridge basis functions are generated via a application-specific approximation algorithm whereas in the latter case we sample random stump basis functions. The difficulty of generating lower bounds and policy costs using the approximations from RLP or FGLP is similar. Since RLP and FGLP are solved via constraint generation, the approximation $\eta(\boldsymbol{\lambda})$ can be shown to

Table EC.1 Parameters of the GJR instances.

AK Instance Index	J	$\bar{s}$	z	AK Instance Index	J	$\bar{s}$	z	AK Instance Index	J	$\bar{s}$	z
2	4	Random	100	6	4	Discrete	100	9	6	Random	100
14	6	Discrete	67	15	6	Discrete	100	18	8	Random	75
19	8	Random	100	22	8	Constant	75	23	8	Constant	100
25	8	Discrete	50	26	8	Discrete	75	27	8	Discrete	100
32	10	Random	100	35	10	Constant	60	36	10	Constant	80
37	10	Constant	100	41	10	Discrete	80	42	10	Discrete	100

provide a lower bound on the optimal policy cost. To obtain a policy cost estimate, we can simulate the policy whose action at state s is obtained by (i) replacing η and $u(\cdot)$ in the right hand side of (EC.21) by $\eta(\boldsymbol{\lambda})$ and $u(\cdot;\boldsymbol{\beta})$, respectively, and (ii) solving the resulting optimization problem.

For testing, we compare the optimality gaps from RLP and FGLP on the GJR instances in AK, which contains instances with and without holding costs. AK find that the instances without holding costs are the ones where adding basis functions adaptively on top of the affine bias function approximation has significant impact. We thus focus on these instances and in particular look at a subset of 18 instances where the lower bound improves by at least 2% as a result of ridge basis function generation in RLP. In Table EC.1, we summarize the considered GJR instances, also indicating the index of the instance used in Table 2 of AK. The number of items (J) in these instances is $4,6,8,$ and 10. The usage rate λ_j is distributed uniformly in the interval $[0,10]$. The vector of maximum inventory levels $\bar{s}$ is chosen based on two random variables u_j and α_j associated with each item $j\in\{1,2,\dots,J\}$ that are distributed uniformly over $[0,1]$ and $\{2,4,8\}$, respectively. These random variables are independent across items. The j-th bound $\bar{s}_j$ on the inventory level is defined in three ways, labeled "random", "constant", and "discrete", as $10\lambda_j u_j+\lambda_j$, $\bar{s}_j=\sum_{k=1}^{J}\lambda_k(u_k+{}^1\!/_J)$, and $\bar{s}_j=\alpha_j\sum_{k=1}^{J}\lambda_k(u_k+{}^1\!/_J)$, respectively. The joint replenishment capacity $\bar{a}$ is set equal to the summation of the first z% of the smallest storage limits $\bar{s}_j$, $j=1,2,\dots,J$, where z varies in set $\{50,60,67,75,80,100\}$ across instances. The immediate cost form is $c(s,a)=c_{\mathrm{supp}(a)}=c'+\sum_{j\in\mathrm{supp}(a)}c''_j$, where $c'\geq 0$ and $c''_j\geq 0$ are constant and item-specific fixed costs, respectively. AK set $c'=100$ and sample c''_j from a uniform distribution over the range $[0,60]$.

EC.4.2.4. Constraint Generation and Greedy Policy Optimization. We employed constraint generation to solve the average-cost variants of FALP and FGLP with random stump basis functions. Consider the FALP formulation in §EC.4.2.2 and recall the decomposition $\hat{\eta}+\sum_{j=1}^{J}\beta_{1,j}\lambda_j$ for the long-run optimal average cost $\eta(\boldsymbol{\lambda})$. Let $(\hat{\eta}_{\mathrm{N}}^{\text{INI}},\boldsymbol{\beta}_{\mathrm{N}}^{\text{INI}})$ be a solution to a version of $\text{FALP}_{(\mathrm{N})}$ with constraints enforced for state-action pairs $(s,a)\in\hat{\mathcal{S}}\times\hat{\mathcal{A}}_s$ alone, where $\hat{\mathcal{S}}\times\hat{\mathcal{A}}_s$ is a sampled subset of $\mathcal{S}\times\mathcal{A}_s$. Given solution $(\hat{\eta}_{\mathrm{N}}^{\text{INI}},\boldsymbol{\beta}_{\mathrm{N}}^{\text{INI}})$, the following separation problem can be solved to find a state-action pair, if any, that violates the FALP constraints corresponding to $\mathcal{S}\times\mathcal{A}_s$:

$$\Psi(\hat{\eta}_{\mathrm{N}}^{\text{INI}},\boldsymbol{\beta}_{\mathrm{N}}^{\text{INI}}):=\min_{(s,a)\in\mathcal{S}\times\mathcal{A}_s}\left\{c(s,a)-\hat{\eta}_{\mathrm{N}}^{\text{INI}}\,T(s,a)-\sum_{j=1}^{J}\beta_{1,j}^{\text{INI}}a_j-\sum_{i=1}^{N}\beta_{2,i}^{\text{INI}}\big(\varphi(s';\theta_i)-\varphi(s;\theta_i)\big)\right\},$$

where we use the definition of GJR transition function, that is, $s' = s + a - \lambda T(s,a)$, to derive this program. The above separation problem is based on the average-cost FALP constraints shown in §EC.4.2.2 and can also be found in (11) of Adelman and Klabjan (2012). Motivated by the mixed integer linear programming reformulation of the separation problem (with no holding cost) in §3.1 of Adelman and Klabjan (2012), we discuss the analogous mixed integer linear programming formulation $\Psi(\hat{\eta}^{\text{INI}}_{\text{N}}, \boldsymbol{\beta}^{\text{INI}}_{\text{N}})$ for the FALP separation problem when using random stump basis functions. Recalling the transition time $T(s,a) = \min_j\{{}^{s_j+a_j}/_{\lambda_j}\}$ and the bias function approximation (EC.25), this formulation is

$$\Psi(\hat{\eta}^{\text{INI}}_{\text{N}}, \boldsymbol{\beta}^{\text{INI}}_{\text{N}}) \equiv$$

$$\min_{(G,Q,Q',s,a,t,s',Z,Z')} \left(c' + \sum_{j=1}^{J} c''_j G_j\right) - \left(\hat{\eta} t + \sum_{j=1}^{J} \beta_{1,j} a_j \;\; + \sum_{i=1}^{N} \beta_{2,i}\big(Z'_i - Z_i\big)\right)$$

$$\begin{array}{llll}
\displaystyle\sum_{j=1}^{J} G_j \geq 1; & & a_j \leq \bar{s}_j G_j, & j = 1,2,\dots J; \\
s'_j = s_j + a_j - \lambda_j t, & j = 1,2,\dots J; & s_j + a_j \leq \bar{s}_j, & j = 1,2,\dots J; \\
\displaystyle\sum_{j=1}^{J} a_j \leq \bar{a}; & & s_j \leq \bar{s}_j(1 - Q_j), & j = 1,2,\dots J; \\
\displaystyle\sum_{j=1}^{J} Q_j \geq 1; & & s'_i \leq \bar{s}_j(1 - Q'_j), & j = 1,2,\dots J; \\
\displaystyle\sum_{j=1}^{J} Q'_j \geq 1; & & Q_j \leq G_j, & j = 1,2,\dots J; \\
Z_i = \operatorname{sgn}(s'_{q_i} - \omega_i), & i = 1,\dots,N; & Z'_i = \operatorname{sgn}(s_{q_i} - \omega_i), & i = 1,\dots,N; \\
G, Q, Q' \text{ binary}; & & Z, Z' \text{ integer}; & \\
s, a, t, s' \text{ nonnegative}. & & &
\end{array}$$

In the above mixed integer linear program, the variable G_j is one if item j is replenished and zero otherwise. Constraint $\sum_{j=1}^{J} G_j \geq 1$ ensures that at least one item is replenished. If $G_j = 1$ for some $j \in \{1, 2\dots, J\}$, then the constraint $a_j \leq \bar{s}_j$ ensures that the replenishment decision a_j can take any feasible value, and if $G_j = 0$, we have $a_j = 0$. Constraints $s'_j = s_j + a_j - \lambda_j t$ model the MDP transition function. Constraints $s_j + a_j \leq \bar{s}_j$ and $\sum_{j=1}^{J} a_j \leq \bar{a}$ check that the state-action pair (s,a) adheres to the inventory and replenishment capacities, respectively. For $j \in \{1, 2\dots, J\}$, if binary variable Q_j is one, then item j is stocked out at the current decision time, i.e. $s_j = 0$, and if Q'_j is one, then this item will be stocked out in the next decision epoch, i.e. $s'_j = 0$. Constraints $\sum_{j=1}^{J} Q_j \geq 1$ and $\sum_{j=1}^{J} Q'_j \geq 1$ ensure at least an item at the current and the next decision epochs is stocked out. If $G_j = 0$ for some item j, then it should not be stocked out, and thus $Q_j = 0$ via constraint $Q_j \leq G_j$; otherwise, $Q_j \in$

$\{0,1\}$, that is, we can either replenish a stocked-out item or an item with a non-zero inventory level. Integer variables $Z_i \in \{-1,0,1\}$ and $Z'_i \in \{-1,0,1\}$ model the value of random basis function $\varphi(s;\theta_i)$ and $\varphi(s';\theta_i)$. A sign function can be implemented in a solver as a piecewise constant function using a big-M formulation or approximately as a piecewise linear function. After encountering numerical issues with the first option, we used the function `setPWLObj` in Gurobi (Gurobi Optimization 2019) to model sign functions. Specifically, we implemented the following piecewise linear approximation:

$$\operatorname{sgn}(x) \approx \begin{cases} 1 & x \geq \epsilon; \\ \frac{x}{\epsilon} & x \in [-\epsilon, \epsilon]; \\ -1 & x \leq -\epsilon. \end{cases}$$

To be consistent, we also use the above approximation to construct our VFAs.

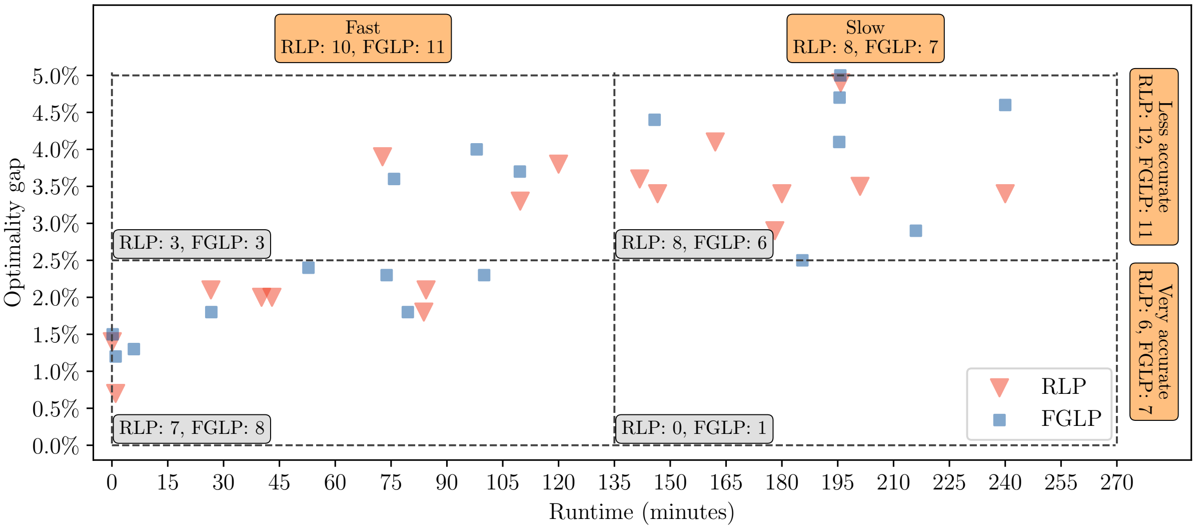
The separation problem is key to constraint generation. Given solution $(\hat{\eta}^{\text{INI}}_{\text{N}}, \boldsymbol{\beta}^{\text{INI}}_{\text{N}})$ that is obtained by solving $\text{FALP}_{(\text{N})}$ with constraints in $\hat{\mathcal{S}} \times \hat{\mathcal{A}}_s$, if the optimal objective value $\Psi(\hat{\eta}^{\text{INI}}_{\text{N}}, \boldsymbol{\beta}^{\text{INI}}_{\text{N}})$ is positive, the current solution $(\hat{\eta}^{\text{INI}}_{\text{N}}, \boldsymbol{\beta}^{\text{INI}}_{\text{N}})$ is feasible to the continuum of $\text{FALP}_{(\text{N})}$ constraints. Otherwise, the state-action component of an optimal solution, $(s^{\text{SEP}}, a^{\text{SEP}})$, computed by $\Psi(\hat{\eta}_{\text{N}}, \boldsymbol{\beta}_{\text{N}})$ corresponds to an $\text{FALP}_{(\text{N})}$ constraint violated by $(\hat{\eta}^{\text{INI}}_{\text{N}}, \boldsymbol{\beta}^{\text{INI}}_{\text{N}})$. In this case, we update $\hat{\mathcal{S}} \times \hat{\mathcal{A}}_s$ to $\hat{\mathcal{S}} \times \hat{\mathcal{A}}_s \cup \{(s^{\text{SEP}}, a^{\text{SEP}})\}$ and re-solve $\text{FALP}_{(\text{N})}$ with the new set of constraints to find a solution. We repeat this procedure until the violation becomes negligible. The optimal value of $\text{FALP}_{(\text{N})}$ in the last iteration of this process is a lower bound on the optimal cost.

To estimate policy cost associated with a bias function approximation, we follow Algorithm 1 in Adelman and Klabjan (2012). The core of this algorithm is to solve the greedy policy optimization (4) via a mixed-integer linear program similar to the separation problem. We thus solve a modification of greedy policy optimization, that is known as K-step greedy policy optimization. Given bias function approximation $u(s;\boldsymbol{\beta}) = \beta_0 - \sum_{j=1}^{J} \beta_{1,j} s_j - \sum_{i=1}^{N} \beta_{2,i} \varphi(s;\theta_i)$ and $\eta = \hat{\eta} + \sum_{j=1}^{J} \beta_{1,j} \lambda_j$ computed by the $\text{FALP}_{(\text{N})}$ in §EC.4.2.2, the action taken by the K-step greedy policy $\pi_{g,K}(s_t;\hat{\eta},\boldsymbol{\beta})$ at the current stage t and state $\hat{s}_t$ is defined by the a_t component of the optimal solution to

$$\begin{aligned} \min_{(a_t, s_t, \ldots, a_{t+K-1}, s_{t+K-1}, s_{t+K})} \quad & \sum_{t'=t}^{t+K-1} \big(c(s_{t'}, a_{t'}) - \eta T(s_{t'}, a_{t'}) \big) \; + \; u(s_{t+K}) \\ \text{s.t.} \quad & s_t = \hat{s}_t, \\ & s_{t'+1} = s_{t'} + a_{t'} - \lambda_{t'} T(s_{t'}, a_{t'}), && \forall t' = t, \ldots, K-1, \\ & a_{t'} \in \mathcal{A}_{s_{t'}}, && \forall t' = t, \ldots, K-1. \end{aligned}$$

Due to our choice of random stump bases, we can efficiently solve the K-step greedy optimization by casting it as a mixed integer linear program similar to the optimization problem (PD) in Adelman and Klabjan (2012). We do not repeat this program here as it is analogous to math program $\Psi(\hat{\eta}^{\text{INI}}_{\text{N}}, \boldsymbol{\beta}^{\text{INI}}_{\text{N}})$.

Figure EC.1 Comparison of RLP and FGLP on the GJR instances.



For implementation, we use $K = 4$ and we set the number of stages for simulating policy in Algorithm 1 of Adelman and Klabjan (2012), e.g., N, to 4000. We use $\epsilon = 0.01$ in the approximation to the sign function. We follow a similar constraint separation strategy for FGLP applied only to constraints (6), that is we do not separate the self-guiding constraints (7) as their exact feasibility is not needed to obtain a valid lower bound. Instead, we enforce constraints (7) only on a set of sampled which include 5,000 initially sampled states plus those encountered during the constraint separation process applied to constraints (6).

EC.4.2.5. Results. We implemented adaptive basis function generation in RLP following the procedure described in AK and did this for FGLP in the framework of Algorithm 1 with ten new basis functions added every iteration (i.e., $B = 10$). As a termination criteria for Algorithm 1, we set an optimality gap tolerance of 2% (i.e., $\tau = 0.02$) and chose run time limits of 1, 2, 3, and 4 hours for instances with J equal to 4, 6, 8, and 10 items, respectively. For each instance specification in Table EC.1, we generated five realizations of the corresponding random variables and computed the (average) optimality gap and (average) run time.

We find that RLP and FGLP are able to obtain policies with optimality gaps less than 5% across all 18 instances. The lower bounds from RLP and FGLP improve on average the lower bounds based on an affine bias function approximation (i.e., no adaptive basis function generation) by 4.7% and 4.5%, respectively. Their corresponding maximum lower bound improvements are 13.7% and 12.1% (across the eighteen instances and the five realizations for each instance). In contrast, the policy costs from an affine approximation on these instances do not improve significantly as a result of adaptive basis function generation in both RLP and FGLP. These observations are consistent with those reported in AK.

To understand the relative performance of RLP and FGLP, we represent the instances in Figure EC.1 in terms of their run time (x-axis) and optimality gap (y-axis). There are thus 18 points for each method. We use squares and triangles to represent results from FGLP and RLP, respectively. We also divide each axis into two halves which leads to four quadrants. The lower-left quadrant contains the instances that are solved the fastest (less than 135 minutes) and have the least optimality gap (less than 2.5%), while the upper-right quadrant include the ones that take the most time and have the highest optimality gaps. The counts of the number of instances in each quadrant is also shown and is largely the same for both RLP and FGLP, with one notable exception in the upper right quadrant where FGLP has two fewer instances than RLP. Overall, these results show that FGLP is competitive with RLP on the GJR instances. This is encouraging because the random basis function generation approach used in FGLP does not exploit any application-specific structure.

References

Adelman D, Klabjan D (2012) Computing near-optimal policies in generalized joint replenishment. *INFORMS Journal on Computing* 24(1):148–164.

De Farias DP, Van Roy B (2003) The linear programming approach to approximate dynamic programming. *Operations Research* 51(6):850–865.

Folland GB (1999) *Real Analysis: Modern Techniques and Their Applications* (New York, NY: John Wiley & Sons).

Gurobi Optimization L (2019) Gurobi optimizer reference manual. URL `http://www.gurobi.com`.

Hernández-Lerma O, Lasserre JB (1996) *Discrete-time Markov Control Processes: Basic Optimality Criteria*, volume 30 (New York, NY: Springer Science & Business Media).

Hernández-Lerma O, Lasserre JB (1999) *Further Topics on Discrete-time Markov Control Processes*, volume 42 (New York, NY: Springer Science & Business Media).

Klabjan D, Adelman D (2007) An infinite-dimensional linear programming algorithm for deterministic semi-Markov decision processes on Borel spaces. *Mathematics of Operations Research* 32(3):528–550.

Lin Q, Nadarajah S, Soheili N (2019) Revisiting approximate linear programming: Constraint-violation learning with applications to inventory control and energy storage. *Management Science* (forthcoming).

Mohri M, Rostamizadeh A, Talwalkar A (2012) *Foundations of Machine Learning* (Cambridge, MA: MIT press), first edition.

Rahimi A, Recht B (2008) Uniform approximation of functions with random bases. *2008 46th Annual Allerton Conference on Communication, Control, and Computing*, 555–561.

Rudin W (1987) *Real and Complex Analysis* (Singapore: McGraw-Hill).