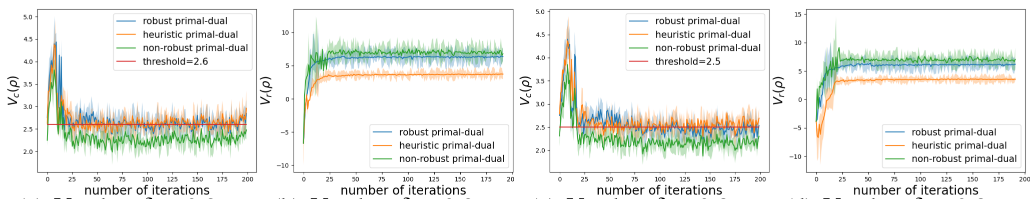
Appendix

A Additional Experiments

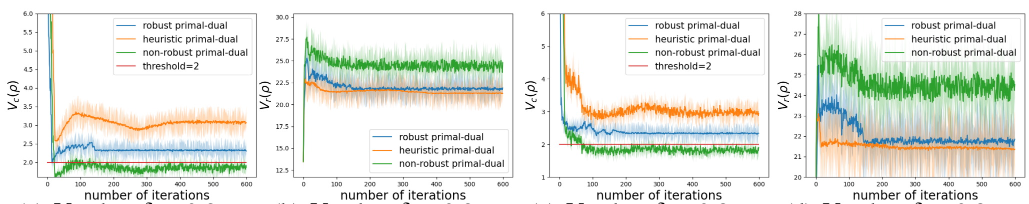
4×4 **Frozen Lake problem.** The 4×4 frozen lake is similar to the 8×8 one but with a smaller map. Similarly, we randomly generate the utility signal for each state-action pair. The results are shown in Fig.4.



(a) V_c when $\delta = 0.2$. (b) V_r when $\delta = 0.2$. (c) V_c when $\delta = 0.3$. (d) V_r when $\delta = 0.3$.

Figure 4: Comparison on 4×4 Frozen-Lake Problem.

N**-Chain problem.** We then compare three algorithms under the N-Chain problem environment. The N-chain problem involves a chain contains N nodes. The agent can either move to its left or right node. When it goes to left, it receives a reward-utility signal $(1, 0)$; When it goes right, it receives a reward-utility signal $(0, 2)$, and if the agent arrives the N-th node, it receives a bonus reward of 40. There is also a small probability that the agent slips to the different direction of its action. In this experiment, we set $N = 40$. The results are shown in Fig.5.



(a) V_c when $\delta = 0.2$. (b) V_r when $\delta = 0.2$. (c) V_c when $\delta = 0.3$. (d) V_r when $\delta = 0.3$.

Figure 5: Comparison on N-Chain Problem.

B Proof of two equivalent formulation of Lagrangian Functions

Consider the problem

$$\max_{\pi} \min_{\mathsf{P}} V^{\pi}_{\mathsf{P},r}, \tag{18}$$

$$s.t. \min_{\mathsf{P}} V^{\pi}_{\mathsf{P},c} \geq 0 \tag{19}$$

The Lagrangian function of it is given by $L(\pi, \lambda) = \min_{\mathsf{P}} V^{\pi}_{\mathsf{P},r} + \lambda \min_{\mathsf{P}} V^{\pi}_{\mathsf{P},c}$. Hence the original problem is equivalent to the problem $\max_{\pi} \min_{\lambda \geq 0} L(\pi, \lambda)$.

We also consider an alternative problem formulation $\max_{\pi} \min_{\lambda \geq 0} \hat{L}(\pi, \lambda)$, where $\hat{L}(\pi, \lambda) = \min_{\mathsf{P}} \{V^{\pi}_{\mathsf{P},r} + \lambda V^{\pi}_{\mathsf{P},c}\}$. These two problem is actually equivalent, i.e., $\min_{\lambda \geq 0} L(\pi, \lambda) = \min_{\lambda \geq 0} \hat{L}(\pi, \lambda)$ for any π.

Proof. For the feasible policies, i.e., $V^{\pi}_{\mathsf{P},c} \geq 0$ for any $\mathsf{P} \in \mathcal{P}$. Thus

$$\begin{aligned}\min_{\lambda \geq 0} \hat{L}(\pi, \lambda) &= \min_{\lambda \geq 0} \min_{\mathsf{P} \in \mathcal{P}} \{V^{\pi}_{\mathsf{P},r} + \lambda V^{\pi}_{\mathsf{P},c}\} \\ &= \min_{\mathsf{P} \in \mathcal{P}} \min_{\lambda \geq 0} \{V^{\pi}_{\mathsf{P},r} + \lambda V^{\pi}_{\mathsf{P},c}\}\end{aligned}$$

$$
\begin{aligned}
&= \min_{\mathsf{P}\in\mathcal{P}}\{\{V^{\pi}_{\mathsf{P},r} + \min_{\lambda\geq 0} \lambda V^{\pi}_{\mathsf{P},c}\}\} \\
&= \min_{\mathsf{P}\in\mathcal{P}} V^{\pi}_{\mathsf{P},r}. \qquad (20)
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
\min_{\lambda\geq 0} L(\pi,\lambda) &= \min_{\lambda\geq 0}\{\min_{\mathsf{P}\in\mathcal{P}} V^{\pi}_{\mathsf{P},r} + \min_{\mathsf{P}\in\mathcal{P}} \lambda V^{\pi}_{\mathsf{P},c}\} \\
&= \min_{\mathsf{P}\in\mathcal{P}} V^{\pi}_{\mathsf{P},r} + \min_{\lambda\geq 0}\min_{\mathsf{P}\in\mathcal{P}} \lambda V^{\pi}_{\mathsf{P},c} \\
&= \min_{\mathsf{P}\in\mathcal{P}} V^{\pi}_{\mathsf{P},r}. \qquad (21)
\end{aligned}
$$

For any infeasible π, i.e., $\exists \mathsf{P}' \in \mathcal{P}$, s.t. $\min_{\mathsf{P}} V^{\pi}_{\mathsf{P},c} < 0$. Clearly $\min_{\lambda\geq 0} L(\pi,\lambda) = -\infty$. And

$$
\min_{\lambda\geq 0} \hat{L}(\pi,\lambda) = \min_{\lambda\geq 0}\min_{\mathsf{P}\in\mathcal{P}}\{V^{\pi}_{\mathsf{P},r} + \lambda V^{\pi}_{\mathsf{P},c}\} \leq \min_{\lambda\geq 0}\{V^{\pi}_{\mathsf{P}',r} + \lambda V^{\pi}_{\mathsf{P}',c}\} = -\infty. \qquad (22)
$$

Hence $\min_{\lambda\geq 0} L(\pi,\lambda) = \min_{\lambda\geq 0} \hat{L}(\pi,\lambda)$ for any π, and hence the two problems are equivalent. $\square$

C PROOF OF LEMMA 1

Denote by $\mathsf{P}^{\pi} = \{(p^{\pi})^a_s \in \Delta_{\mathcal{S}} : s \in \mathcal{S}, a \in \mathcal{A}\}$ the worst-case transition kernel corresponding to the policy π. We consider the δ-contamination uncertainty set defined in Section 4. We then show that under δ-contamination model, the set of visitation distributions is non-convex. The robust visitation distribution set can be written as follows:

$$
\left\{ d \in \Delta_{\mathcal{S}\times\mathcal{A}} : \exists \pi \in \Pi, \text{ s.t. } \forall (s,a), \begin{cases} d(s,a) = \pi(a|s)\sum_b d(s,b), \\ \gamma \sum_{s',a'} (p^{\pi})^{a'}_{s',s} d(s',a') + (1-\gamma)\rho(s) = \sum_a d(s,a). \end{cases} \right\}. \qquad (23)
$$

Under the δ-contamination model, P^{π} can be explicated as $(p^{\pi})^a_{s,s'} = (1-\delta)p^a_{s,s'} + \delta \mathbb{1}_{\{s'=\arg\min V^{\pi}\}}$. Hence the set in eq. (23) can be rewritten as

$$
\left\{ d \in \Delta_{\mathcal{S}\times\mathcal{A}} : \exists \pi, \text{ s.t. } \forall (s,a), \begin{cases} d(s,a) = \pi(a|s)\left(\sum_b d(s,b)\right), \\ \gamma(1-\delta)\sum_{s',a'} p^{a'}_{s',s} d(s',a') + \gamma\delta \mathbb{1}_{\{s=\arg\min V^{\pi}\}} \\ \qquad + (1-\gamma)\rho(s) = \sum_a d(s,a). \end{cases} \right\}. \qquad (24)
$$

Now consider any two pairs $(\pi_1, d_1), (\pi_2, d_2)$ of policy and their worst-case visitation distribution, to show that the set is convex, we need to find a pair (π', d') such that $\forall \lambda \in [0,1]$ and $\forall s, a$,

$$
\lambda d_1(s,a) + (1-\lambda) d_2(s,a) = d'(s,a), \qquad (25)
$$

$$
d'(s,a) = \pi'(a|s)\left(\sum_b d'(s,b)\right), \qquad (26)
$$

$$
\sum_{a'} d'(s,a') = \gamma(1-\delta)\sum_{s',a'} p^{a'}_{s',s} d'(s',a') + \gamma\delta \mathbb{1}_{\left\{s=\arg\min V^{\pi'}\right\}} + (1-\gamma)\rho(s). \qquad (27)
$$

eq. (27) firstly implies that $\forall s$,

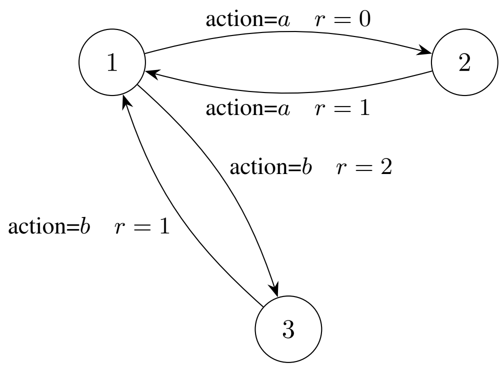
$$
\lambda \mathbb{1}_{\{s=\arg\min V^{\pi_1}\}} + (1-\lambda)\mathbb{1}_{\{s=\arg\min V^{\pi_2}\}} = \mathbb{1}_{\left\{s=\arg\min V^{\pi'}\right\}}, \qquad (28)
$$

where from eq. (25) and eq. (26), π' should be

$$
\pi'(a|s) = \frac{d'(s,a)}{\sum_b d'(s,b)} = \frac{\lambda d_1(s,a) + (1-\lambda) d_2(s,a)}{\sum_b (\lambda d_1(s,b) + (1-\lambda) d_2(s,b))}. \qquad (29)
$$

We then construct the following counterexample, which shows that there exists a robust MDP, two policy-distribution pairs $(\pi_1, d_1), (\pi_2, d_2)$, and $\lambda \in (0,1)$, such that $\lambda \mathbb{1}_{\{s=\arg\min V^{\pi_1}\}} + (1-\lambda)\mathbb{1}_{\{s=\arg\min V^{\pi_2}\}} \neq \mathbb{1}_{\left\{s=\arg\min V^{\pi'}\right\}}$, and therefore the set of robust visitation distribution is non-convex.

Consider the following Robust MDP. It has three states $1, 2, 3$ and two actions a, b. When the agent is at state 1, if it takes action a, the system will transit to state 2 and receive reward $r = 0$; if it takes action b, the system will transit to state 3 and receive reward $r = 2$. When the agent is at state $2/3$, it can only take action a/b, the system can only transits back to state 1 and the agent will receive reward $r = 1$. The initial distribution is $\mathbb{1}_{s=1}$.



Clearly all policy can be written as $\pi = (p, 1-p)$, where p is the probability of taking action a at state 1. We consider two policies, $\pi_1 = (1, 0)$ and $\pi_2 = (0, 1)$.

It can be verified that $\arg\min V^{\pi_1} = 1$, and its robust visitation distribution, denoted by d_1, is

$$d_1(1, a) = \frac{1-\gamma}{1-\gamma^2}, \tag{30}$$
$$d_1(1, b) = 0, \tag{31}$$
$$d_1(2, a) = \frac{\gamma(1-\gamma)}{1-\gamma^2}, \tag{32}$$
$$d_1(2, b) = 0, \tag{33}$$
$$d_1(3, a) = 0, \tag{34}$$
$$d_1(3, b) = 0. \tag{35}$$

Similarly, $\arg\min V^{\pi_2} = 2$, and and its robust visitation distribution, denoted by d_2, is

$$d_2(1, a) = 0, \tag{36}$$
$$d_2(1, b) = \frac{1-\gamma}{1-\gamma^2}, \tag{37}$$
$$d_2(2, a) = 0, \tag{38}$$
$$d_2(2, b) = 0, \tag{39}$$
$$d_2(3, a) = 0, \tag{40}$$
$$d_2(3, b) = \frac{\gamma(1-\gamma)}{1-\gamma^2}. \tag{41}$$

Hence according to eq. (29), π' should be as follows:

$$\pi'(a|1) = \lambda, \pi'(b|1) = 1-\lambda, \pi'(a|2) = 1, \pi'(b|3) = 1. \tag{42}$$

We then show that there exists $\lambda \in [0,1]$, such that $\lambda \mathbb{1}_{\{s=1\}} + (1-\lambda)\mathbb{1}_{\{s=2\}} \neq \mathbb{1}_{\left\{\arg\min V^{\pi'}\right\}}$.

Clearly eq. (28) holds only if $V^{\pi'}(1) = V^{\pi'}(2) = \min_s V^{\pi'}(s)$. However, according to the Bellman equations for π', we have that

$$V^{\pi'}(1) = \lambda(\gamma(1-\delta)V^{\pi'}(2) + \gamma\delta \min V^{\pi'}) + (1-\lambda)(2 + \gamma(1-\delta)V^{\pi'}(3) + \gamma\delta \min V^{\pi'}), \tag{43}$$

$$V^{\pi'}(2) = 1 + \gamma(1-\delta)V^{\pi'}(1) + \gamma\delta \min V^{\pi'}, \tag{44}$$
$$V^{\pi'}(3) = 1 + \gamma(1-\delta)V^{\pi'}(1) + \gamma\delta \min V^{\pi'}. \tag{45}$$

If we set $\lambda = \frac{1}{3}$,

$$V^{\pi'}(1) = \frac{4}{3} + \gamma\delta \min V^{\pi'} + \gamma(1-\delta)V^{\pi'}(2), \tag{46}$$
$$V^{\pi'}(2) = 1 + \gamma\delta \min V^{\pi'} + \gamma(1-\delta)V^{\pi'}(1). \tag{47}$$

Clearly, $V^{\pi'}(1) \neq V^{\pi'}(2)$, and hence $\lambda \mathbb{1}_{\{\arg\min V^1\}} + (1-\lambda)\mathbb{1}_{\{\arg\min V^2\}} \neq \mathbb{1}_{\left\{\arg\min V^{\pi'}\right\}}$.

D PROOF OF LEMMAS 2 AND 5

Proof of Lemma 2

Proof. We first set $C = V_r^{\pi_{\theta^*}}(\rho) + \lambda^*(V_c^{\pi_{\theta^*}}(\rho) - b)$, clearly $\max_{\pi\in\Pi} V_r^\pi(\rho) + \lambda^*(V_c^\pi(\rho) - b) = C$, and hence

$$C = \max_{\pi\in\Pi} V_r^\pi(\rho) + \lambda^*(V_c^\pi(\rho) - b) \geq V_r^{\pi^\zeta}(\rho) + \lambda^*(V_c^{\pi^\zeta}(\rho) - b) \geq V_r^{\pi^\zeta}(\rho) + \lambda^*\zeta. \tag{48}$$

Thus we have that

$$\lambda^* \leq \frac{C - V_r^{\pi^\zeta}(\rho)}{\zeta}. \tag{49}$$

Note that

$$C = \min_{\lambda\geq 0}\max_{\pi\in\Pi} V_r^\pi(\rho) + \lambda(V_c^\pi(\rho) - b) \overset{(a)}{\leq} \max_{\pi\in\Pi} V_r^\pi(\rho) \leq \frac{1}{1-\gamma}, \tag{50}$$

where (a) is because $\min_{\lambda\geq 0}\max_{\pi\in\Pi} V_r^\pi(\rho) + \lambda(V_c^\pi(\rho) - b)$ is less than the optimal value of inner problem when $\lambda = 0$, i.e., $\max_{\pi\in\Pi} V_r^\pi(\rho)$, and $\frac{1}{1-\gamma}$ is the upper bound of robust value functions. Hence we have that

$$\lambda^* \leq \frac{1}{(1-\gamma)\zeta}, \tag{51}$$

which completes the proof. □

Proof of Lemma 5

Proof. Set $C = V_{\sigma,r}^{\theta^*}(\rho) + \lambda^*(V_{\sigma,c}^{\theta^*}(\rho) - b)$, then

$$C = \max_{\pi\in\Pi} V_{\sigma,r}^\pi(\rho) + \lambda^*(V_{\sigma,c}^\pi(\rho) - b) \geq V_{\sigma,r}^{\pi^{\zeta'}}(\rho) + \lambda^*(V_{\sigma,c}^{\pi^{\zeta'}}(\rho) - b) \geq V_{\sigma,r}^{\pi^{\zeta'}}(\rho) + \lambda^*\zeta'. \tag{52}$$

Thus we have that

$$C \geq V_{\sigma,r}^{\pi^\zeta}(\rho) + \lambda^*\zeta', \tag{53}$$

hence

$$\lambda^* \leq \frac{C - V_{\sigma,r}^{\pi^\zeta}(\rho)}{\zeta'}. \tag{54}$$

Note that

$$C = \min_{\lambda\geq 0}\max_{\pi\in\Pi} V_{\sigma,r}^\pi(\rho) + \lambda(V_{\sigma,c}^\pi(\rho) - b) \leq \max_{\pi\in\Pi} V_{\sigma,r}^\pi(\rho) \leq C_\sigma, \tag{55}$$

where C_σ is the upper bound of smoothed robust value functions (Wang & Zou, 2022): $C_\sigma = \frac{1}{1-\gamma}(1 + 2\gamma R \frac{\log|\mathcal{S}|}{\sigma})$. Hence we have that

$$\lambda^* \leq \frac{C_\sigma}{\zeta'}, \tag{56}$$

which completes the proof. □

E Proof of Lemma 6

Proof. For any λ, denote the optimal value of the inner problems $\max_{\pi\in\Pi_\Theta} V^\pi_{\sigma,r}(\rho)+\lambda(V^\pi_{\sigma,c}(\rho)-b)$ and $\max_{\pi\in\Pi_\Theta} V^\pi_{r}(\rho)+\lambda(V^\pi_{c}(\rho)-b)$ by $V^D(\lambda)$ and $V^D_\sigma(\lambda)$. It is then easy to verify that

$$|V^D(\lambda)-V^D_\sigma(\lambda)|\leq(1+\lambda)\epsilon\leq(1+\Lambda^*)\epsilon. \tag{57}$$

Denote the optimal solutions of $\min_{\lambda\in[0,\Lambda^*]} V^D(\lambda)$ and $\min_{\lambda\in[0,\Lambda^*]} V^D_\sigma(\lambda)$ by λ^D and λ^D_σ. We thus conclude that $|V^D_\sigma(\lambda^D_\sigma)-V^D(\lambda^D)|\leq(1+\Lambda^*)\,\epsilon$, and this thus completes the proof. □

F Proof of Theorem 1

We restate Theorem 1 with all the specific step sizes as follows.

Set $b_t=\frac{19}{20\xi t^{0.25}}$, $\mu_t=\xi(C^V_\sigma)^2+\frac{16\tau(C^V_\sigma)^2}{\xi(b_{t+1})^2}-2\nu$, $\beta_t=\frac{1}{\xi}$, $\alpha_t=\nu+\mu_t$, where $\xi>\frac{2\nu+(1+\Lambda^*)L_\sigma}{(C^V_\sigma)^2}$, ν is any positive number and τ is any number greater than 2, then

$$\min_{1\leq t\leq T}\|G_t\|^2\leq 2\epsilon, \tag{58}$$

when

$$T=\max\left\{\frac{7(\Lambda^*)^4}{\xi^4\epsilon^4},\left(2+\frac{9\xi(\tau-2)(C^V_\sigma)^2uK}{\epsilon^2}\right)^2\right\}=\mathcal{O}(\epsilon^{-4}). \tag{59}$$

The definitions of u, K can be found in Section J.

Theorem 1 can be proved similarly as Theorem 2, and hence the proof is omitted here.

G Proof of Lemma 3

Proof. Recall that $V^L_\sigma(\theta,\lambda)=V^{\pi_\theta}_{\sigma,r}(\rho)+\lambda(V^{\pi_\theta}_{\sigma,c}(\rho)-b)$, hence we have that

$$\nabla_\lambda V^L_\sigma(\theta,\lambda)=V^{\pi_\theta}_{\sigma,c}(\rho)-b, \tag{60}$$
$$\nabla_\theta V^L_\sigma(\theta,\lambda)=\nabla_\theta V^{\pi_\theta}_{\sigma,r}(\rho)+\lambda\nabla_\theta V^{\pi_\theta}_{\sigma,c}(\rho). \tag{61}$$

Note that in (Wang & Zou, 2022), it has been shown that

$$\|V^{\pi_{\theta_1}}_{\sigma,r}-V^{\pi_{\theta_2}}_{\sigma,r}\|\leq C^V_\sigma\|\theta_1-\theta_2\|, \tag{62}$$
$$\|\nabla_\theta V^{\pi_{\theta_1}}_{\sigma,r}-\nabla_\theta V^{\pi_{\theta_2}}_{\sigma,r}\|\leq L_\sigma\|\theta_1-\theta_2\|, \tag{63}$$

where the definition of constants C^V_σ and L_σ can be found in Section J. Hence

$$\|\nabla_\lambda V^L_\sigma(\theta,\lambda)|_{\theta_1}-\nabla_\lambda V^L_\sigma(\theta,\lambda)|_{\theta_2}\|=\|V^{\pi_{\theta_1}}_{\sigma,c}(\rho)-V^{\pi_{\theta_2}}_{\sigma,c}(\rho)\|\leq C^V_\sigma\|\theta_1-\theta_2\|, \tag{64}$$
$$\|\nabla_\lambda V^L_\sigma(\theta,\lambda)|_{\lambda_1}-\nabla_\lambda V^L_\sigma(\theta,\lambda)|_{\lambda_2}\|=0. \tag{65}$$

Similarly, we have that

$$\|\nabla_\theta V^L_\sigma(\theta,\lambda)|_{\theta_1}-\nabla_\theta V^L_\sigma(\theta,\lambda)|_{\theta_2}\|\leq(1+\lambda)L_\sigma\|\theta_1-\theta_2\|\leq(1+\Lambda^*)L_\sigma\|\theta_1-\theta_2\|, \tag{66}$$
$$\|\nabla_\theta V^L_\sigma(\theta,\lambda)|_{\lambda_1}-\nabla_\theta V^L_\sigma(\theta,\lambda)|_{\lambda_2}\|\leq|(\lambda_1-\lambda_2)|\max_{\theta\in\Theta}\|\nabla_\theta V^{\pi_\theta}_{\sigma,c}(\rho)\|\leq C^V_\sigma|\lambda_1-\lambda_2|. \tag{67}$$

This completes the proof. □

H Proof of Proposition 1

Proof. The λ-entry of G_W is smaller than 2ϵ, i.e.,

$$|(G_W)_\lambda|=\left|\beta_W\left(\lambda_W-\prod\nolimits_{[0,\Lambda^*]}\left(\lambda_W-\frac{1}{\beta_W}\left(\nabla_\lambda V^L_\sigma(\theta_W,\lambda_W)\right)\right)\right)\right|<2\epsilon. \tag{68}$$

Denote $\lambda^+ \triangleq \prod_{[0,\Lambda^*]}\left(\lambda_W - \frac{1}{\beta_W}\left(\nabla_\lambda V_\sigma^L(\theta_W,\lambda_W)\right)\right)$. From Lemma 3 in (Ghadimi & Lan, 2016), $-\nabla_\lambda V_\sigma^L(\theta_W,\lambda^+)$ can be rewritten as the sum of two parts: $-\nabla_\lambda V_\sigma^L(\theta_W,\lambda^+) \in N_{[0,\Lambda^*]}(\lambda^+)+4\epsilon B$, where $N_K(x) \triangleq \left\{g \in \mathbb{R}^d : \langle g, y-x\rangle \leq 0 : \forall y \in K\right\}$ is the normal cone, and B is the unit ball.

This hence implies that for any $\lambda \in [0,\Lambda^*]$, $(\lambda-\lambda^+)(V_c^W - b) \geq -4(\lambda-\lambda^+)\epsilon$. By setting $\lambda = \Lambda^*$, we have $V_{c,\sigma}^W + 4\epsilon \geq b$, which means π_W is feasible with a 4ϵ-violation. □

I Proof of Theorem 2

We then prove Theorem 2. Our proof extends the one in (Xu et al., 2020) to the biased setting.

To simplify notations, we denote the updates in Algorithm 3 by $\hat{f}(\theta_t) \triangleq \hat{V}_{\sigma,c}^{\pi_{\theta_t}}(\rho) - b$, and $\hat{g}(\theta_t,\lambda_{t+1}) \triangleq \nabla_\theta \hat{V}_{\sigma,r}^{\pi_{\theta_t}}(\rho) + \lambda_{t+1}\nabla_\theta \hat{V}_{\sigma,c}^{\pi_{\theta_t}}(\rho)$, and denote the update functions in Algorithm 1 by $f(\theta_t) \triangleq V_{\sigma,c}^{\pi_{\theta_t}}(\rho) - b$, and $g(\theta_t,\lambda_{t+1}) \triangleq \nabla_\theta V_{\sigma,r}^{\pi_{\theta_t}}(\rho) + \lambda_{t+1}\nabla_\theta V_{\sigma,c}^{\pi_{\theta_t}}(\rho)$. Here $\hat{f}$ and $\hat{g}$ can be viewed as biased estimations of f and g.

In the following, we will first show several technical lemmas that will be useful in the proof of Theorem 2.

Lemma 7. *Recall that the step size $\alpha_t = \nu + \mu_t$. If $\mu_t > (1+\Lambda^*)L_\sigma$, $\forall t \geq 0$, then*

$$\begin{aligned} V_\sigma^L(\theta_{t+1},\lambda_{t+1}) - V_\sigma^L(\theta_t,\lambda_{t+1}) \geq \langle \theta_{t+1}-\theta_t, -\hat{g}(\theta_t,\lambda_{t+1}) + g(\theta_t,\lambda_{t+1})\rangle \\ + \left(\frac{\mu_t}{2}+\nu\right)\|\theta_{t+1}-\theta_t\|^2. \end{aligned} \tag{69}$$

Proof. Note that from the update of θ_t and proposition of projection, it implies that

$$\left\langle \theta_t + \frac{1}{\alpha_t}\hat{g}(\theta_t,\lambda_{t+1}) - \theta_{t+1}, \theta_t - \theta_{t+1}\right\rangle \leq 0. \tag{70}$$

Hence

$$\langle \hat{g}(\theta_t,\lambda_{t+1}) - \alpha_t(\theta_{t+1}-\theta_t), \theta_t-\theta_{t+1}\rangle \leq 0. \tag{71}$$

From Lemma 3, we have that

$$V_\sigma^L(\theta_{t+1},\lambda_{t+1}) - V_\sigma^L(\theta_t,\lambda_{t+1}) \geq \langle \theta_{t+1}-\theta_t, g(\theta_t,\lambda_{t+1})\rangle - \frac{(1+\Lambda^*)L_\sigma}{2}\|\theta_{t+1}-\theta_t\|^2. \tag{72}$$

Summing up the two inequalities implies

$$\begin{aligned} &V_\sigma^L(\theta_{t+1},\lambda_{t+1}) - V_\sigma^L(\theta_t,\lambda_{t+1}) \\ &\geq \langle \theta_{t+1}-\theta_t, -\hat{g}(\theta_t,\lambda_{t+1}) + g(\theta_t,\lambda_{t+1}) + \alpha_t(\theta_{t+1}-\theta_t)\rangle - \frac{(1+\Lambda^*)L_\sigma}{2}\|\theta_{t+1}-\theta_t\|^2 \\ &\geq \langle \theta_{t+1}-\theta_t, -\hat{g}(\theta_t,\lambda_{t+1}) + g(\theta_t,\lambda_{t+1})\rangle + \left(\alpha_t - \frac{L_\sigma(1+\Lambda^*)}{2}\right)\|\theta_{t+1}-\theta_t\|^2 \\ &\geq \langle \theta_{t+1}-\theta_t, -\hat{g}(\theta_t,\lambda_{t+1}) + g(\theta_t,\lambda_{t+1})\rangle + \left(\frac{\mu_t}{2}+\nu\right)\|\theta_{t+1}-\theta_t\|^2, \end{aligned} \tag{73}$$

and hence completes the proof. □

Lemma 8. *Recall that the step size $\beta_t = \frac{1}{\xi}$, and set $\xi \leq \frac{1}{b_0}$, then*

$$\begin{aligned} &V_\sigma^L(\theta_{t+1},\lambda_{t+1}) - V_\sigma^L(\theta_t,\lambda_t) \\ &\geq (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}))(\lambda_{t+1}-\lambda_t) + \langle \theta_{t+1}-\theta_t, -\hat{g}(\theta_t,\lambda_{t+1}) + g(\theta_t,\lambda_{t+1})\rangle - \frac{\xi(C_\sigma^V)^2}{2}\|\theta_t-\theta_{t-1}\|^2 \\ &\quad + \left(\frac{\mu_t}{2}+\nu\right)\|\theta_{t+1}-\theta_t\|^2 + \frac{b_{t-1}}{2}(\lambda_t^2 - \lambda_{t+1}^2) - \frac{1}{\xi}(\lambda_{t+1}-\lambda_t)^2 - \frac{1}{2\xi}(\lambda_t-\lambda_{t-1})^2. \end{aligned} \tag{74}$$

Proof. For any $t > 1$, define $\tilde{V}_t(\theta, \lambda) \triangleq V_\sigma^L(\theta, \lambda) + \frac{b_{t-1}}{2}\lambda^2$. Thus we have

$$|\nabla_\lambda \tilde{V}_t(\theta_t, \lambda_{t+1}) - \nabla_\lambda \tilde{V}_t(\theta_t, \lambda_t)| = b_{t-1}|\lambda_{t+1} - \lambda_t| \leq b_0|\lambda_{t+1} - \lambda_t|, \tag{75}$$

where that last inequality is due to $b_{t-1} \leq b_0$. Note that $\tilde{V}_t(\theta, \lambda)$ is b_{t-1}-strongly convex in λ, hence we have

$$\begin{aligned}
&(\nabla_\lambda \tilde{V}_t(\theta, \lambda_{t+1}) - \nabla_\lambda \tilde{V}_t(\theta, \lambda_t))(\lambda_{t+1} - \lambda_t) \\
&\geq b_{t-1}(\lambda_{t+1} - \lambda_t)^2 \\
&\geq b_{t-1}\left(\frac{b_{t-1} + b_0}{b_{t-1} + b_0}\right)(\lambda_{t+1} - \lambda_t)^2 \\
&= \frac{b_{t-1}b_0}{b_{t-1} + b_0}(\lambda_{t+1} - \lambda_t)^2 + \frac{b_{t-1}^2}{b_{t-1} + b_0}(\lambda_{t+1} - \lambda_t)^2 \\
&\geq \frac{b_{t-1}b_0}{b_{t-1} + b_0}(\lambda_{t+1} - \lambda_t)^2 + \frac{1}{b_{t-1} + b_0}(\nabla_\lambda \tilde{V}_t(\theta_t, \lambda_{t+1}) - \nabla_\lambda \tilde{V}_t(\theta_t, \lambda_t))^2,
\end{aligned} \tag{76}$$

where the last inequality is from eq. (75).

Recall the update of λ_t in Algorithm 3 which can be rewritten as

$$\lambda_{t+1} = \prod\nolimits_{[0, \Lambda^*]}\left(\lambda_t - \frac{1}{\beta_t}\nabla_\lambda \tilde{V}_{t+1}(\theta_t, \lambda_t) + \frac{1}{\beta_t}(f(\theta_t) - \hat{f}(\theta_t))\right), \tag{77}$$

This further implies that $\forall \lambda \in [0, \Lambda^*]$:

$$(\beta_t(\lambda_{t+1} - \lambda_t) + \nabla_\lambda \tilde{V}_{t+1}(\theta_t, \lambda_t) - f(\theta_t) + \hat{f}(\theta_t))(\lambda - \lambda_{t+1}) \geq 0. \tag{78}$$

Hence setting $\lambda = \lambda_k$ implies that

$$(\beta_t(\lambda_{t+1} - \lambda_t) + \nabla_\lambda \tilde{V}_{t+1}(\theta_t, \lambda_t) - f(\theta_t) + \hat{f}(\theta_t))(\lambda_t - \lambda_{t+1}) \geq 0. \tag{79}$$

Similarly, we have that

$$(\beta_t(\lambda_t - \lambda_{t-1}) + \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}) - f(\theta_{t-1}) + \hat{f}(\theta_{t-1}))(\lambda_{t+1} - \lambda_t) \geq 0. \tag{80}$$

Note that $\tilde{V}_t$ is convex, we hence have that

$$\begin{aligned}
&\tilde{V}_t(\theta_t, \lambda_{t+1}) - \tilde{V}_t(\theta_t, \lambda_t) \\
&\geq (\nabla_\lambda \tilde{V}_t(\theta_t, \lambda_t))(\lambda_{t+1} - \lambda_t) \\
&= (\nabla_\lambda \tilde{V}_t(\theta_t, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))(\lambda_{t+1} - \lambda_t) + (\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))(\lambda_{t+1} - \lambda_t) \\
&\overset{(a)}{\geq} (\nabla_\lambda \tilde{V}_t(\theta_t, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1})(\lambda_{t-1}), \lambda_{t+1} - \lambda_t) \\
&\quad + (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}) - \beta_t(\lambda_t - \lambda_{t-1}))(\lambda_{t+1} - \lambda_t),
\end{aligned} \tag{81}$$

where (a) is from eq. (80). The first term in the RHS of eq. (81) can be further bounded as follows.

$$\begin{aligned}
&(\nabla_\lambda \tilde{V}_t(\theta_t, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))(\lambda_{t+1} - \lambda_t) \\
&= (\nabla_\lambda \tilde{V}_t(\theta_t, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t))(\lambda_{t+1} - \lambda_t) \\
&\quad + (\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))(\lambda_{t+1} - \lambda_t) \\
&= (\nabla_\lambda \tilde{V}_t(\theta_t, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t))(\lambda_{t+1} - \lambda_t) \\
&\quad + (\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))(\lambda_t - \lambda_{t-1}) \\
&\quad + m_{t+1}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1})),
\end{aligned} \tag{82}$$

where $m_{t+1} \triangleq (\lambda_{t+1} - \lambda_t) - (\lambda_t - \lambda_{t-1})$. Plug it in eq. (81) and we have that

$$\begin{aligned}
&\tilde{V}_t(\theta_t, \lambda_{t+1}) - \tilde{V}_t(\theta_t, \lambda_t) \\
&\geq (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}) - \beta_t(\lambda_t - \lambda_{t-1}))(\lambda_{t+1} - \lambda_t)
\end{aligned}$$

$$
\begin{aligned}
&+ \underbrace{(\nabla_\lambda \tilde{V}_t(\theta_t, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t))(\lambda_{t+1} - \lambda_t)}_{(a)} \\
&+ \underbrace{(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))(\lambda_t - \lambda_{t-1})}_{(b)} \\
&+ \underbrace{(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))m_{t+1}}_{(c)}. \qquad (83)
\end{aligned}
$$

We then provide bounds for each term in eq. (83) as follows.

Term (a) can be bounded as follows:

$$
\begin{aligned}
&(\nabla_\lambda \tilde{V}_t(\theta_t, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t))(\lambda_{t+1} - \lambda_t) \\
&= (\nabla_\lambda V_\sigma^L(\theta_t, \lambda_t) - \nabla_\lambda V_\sigma^L(\theta_{t-1}, \lambda_t))(\lambda_{t+1} - \lambda_t) \\
&\geq \frac{-(\lambda_{t+1} - \lambda_t)^2}{2\xi} - \frac{\xi}{2}(\nabla_\lambda V_\sigma^L(\theta_t, \lambda_t) - \nabla_\lambda V_\sigma^L(\theta_{t-1}, \lambda_t))^2 \\
&\geq \frac{-(\lambda_{t+1} - \lambda_t)^2}{2\xi} - \frac{\xi(C_\sigma^V)^2}{2}\|\theta_t - \theta_{t-1}\|^2, \qquad (84)
\end{aligned}
$$

which is from Cauchy–Schwarz inequality and C_σ^V-smoothness of $V_\sigma^L(\theta, \lambda)$.

Term (b) can be bounded as follows:

$$
\begin{aligned}
&(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))(\lambda_t - \lambda_{t-1}) \\
&\geq \frac{1}{b_{t-1} + b_0}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))^2, \qquad (85)
\end{aligned}
$$

which is from eq. (76).

Term (c) can be bounded as follows by Cauchy–Schwarz inequality:

$$
\begin{aligned}
&m_{t+1}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1})) \\
&\geq -\frac{\xi}{2}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))^2 - \frac{1}{2\xi}m_{t+1}^2 \qquad (86)
\end{aligned}
$$

Moreover, it can be shown that

$$
\frac{1}{\xi}(\lambda_{t+1} - \lambda_t)(\lambda_t - \lambda_{t-1}) = \frac{1}{2\xi}(\lambda_{t+1} - \lambda_t)^2 + \frac{1}{2\xi}(\lambda_t - \lambda_{t-1})^2 - \frac{1}{2\xi}m_{t+1}^2. \qquad (87)
$$

Plug eq. (84) to eq. (87) in 83, and we have that

$$
\begin{aligned}
&\tilde{V}_t(\theta_t, \lambda_{t+1}) - \tilde{V}_t(\theta_t, \lambda_t) \\
&\geq (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}))(\lambda_{t+1} - \lambda_t) - \beta_t(\lambda_t - \lambda_{t-1})(\lambda_{t+1} - \lambda_t) \\
&\quad + (\nabla_\lambda \tilde{V}_t(\theta_t, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t))(\lambda_{t+1} - \lambda_t) + (\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))(\lambda_t - \lambda_{t-1}) \\
&\quad + m_{t+1}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1})) \\
&\geq (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}))(\lambda_{t+1} - \lambda_t) - \frac{1}{2\xi}(\lambda_{t+1} - \lambda_t)^2 - \frac{1}{2\xi}(\lambda_t - \lambda_{t-1})^2 + \frac{1}{2\xi}m_{t+1}^2 \\
&\quad - \frac{(\lambda_{t+1} - \lambda_t)^2}{2\xi} - \frac{\xi(C_\sigma^V)^2}{2}\|\theta_t - \theta_{t-1}\|^2 + \frac{1}{b_{t-1} + b_0}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))^2 \\
&\quad - \frac{\xi}{2}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))^2 - \frac{1}{2\xi}m_{t+1}^2 \\
&\geq (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}))(\lambda_{t+1} - \lambda_t) - \frac{1}{\xi}(\lambda_{t+1} - \lambda_t)^2 - \frac{1}{2\xi}(\lambda_t - \lambda_{t-1})^2 - \frac{\xi(C_\sigma^V)^2}{2}\|\theta_t - \theta_{t-1}\|^2. \qquad (88)
\end{aligned}
$$

From the definition of $\tilde{V}_t$, we have that

$$
\tilde{V}_t(\theta_t, \lambda_{t+1}) - \tilde{V}_t(\theta_t, \lambda_t)
$$

$$= V_\sigma^L(\theta_t, \lambda_{t+1}) + \frac{b_{t-1}}{2}\lambda_{t+1}^2 - V_\sigma^L(\theta_t, \lambda_t) - \frac{b_{t-1}}{2}\lambda_t^2. \tag{89}$$

Then we have that

$$\begin{aligned}
&V_\sigma^L(\theta_t, \lambda_{t+1}) - V_\sigma^L(\theta_t, \lambda_t) \\
&\geq \frac{b_{t-1}}{2}(\lambda_t^2 - \lambda_{t+1}^2) + (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}))(\lambda_{t+1} - \lambda_t) \\
&\quad - \frac{1}{\xi}(\lambda_{t+1} - \lambda_t)^2 - \frac{1}{2\xi}(\lambda_t - \lambda_{t-1})^2 - \frac{\xi(C_\sigma^V)^2}{2}\|\theta_t - \theta_{t-1}\|^2.
\end{aligned} \tag{90}$$

Combining with Lemma 7, if $\forall t$, $\mu_t > (1 + \Lambda^*)L_\sigma$, we then have that

$$\begin{aligned}
&V_\sigma^L(\theta_{t+1}, \lambda_{t+1}) - V_\sigma^L(\theta_t, \lambda_t) \\
&\geq (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}))(\lambda_{t+1} - \lambda_t) + \langle \theta_{t+1} - \theta_t, -\hat{g}(\theta_t, \lambda_{t+1}) + g(\theta_t, \lambda_{t+1})\rangle - \frac{\xi(C_\sigma^V)^2}{2}\|\theta_t - \theta_{t-1}\|^2 \\
&\quad + \left(\frac{\mu_t}{2} + \nu\right)\|\theta_{t+1} - \theta_t\|^2 + \frac{b_{t-1}}{2}(\lambda_t^2 - \lambda_{t+1}^2) - \frac{1}{\xi}(\lambda_{t+1} - \lambda_t)^2 - \frac{1}{2\xi}(\lambda_t - \lambda_{t-1})^2.
\end{aligned} \tag{91}$$

□

Lemma 9. *Define*

$$\begin{aligned}
F_{t+1} \triangleq& -\frac{8}{\xi^2 b_{t+1}}(\lambda_t - \lambda_{t+1})^2 - \frac{8}{\xi}\left(1 - \frac{b_t}{b_{t+1}}\right)\lambda_{t+1}^2 + V_\sigma^L(\theta_{t+1}, \lambda_{t+1}) + \frac{b_t}{2}\lambda_{t+1}^2 \\
&+ \left(-\frac{16(C_\sigma^V)^2}{\xi b_{t+1}^2} - \frac{\xi(C_\sigma^V)^2}{2}\right)\|\theta_{t+1} - \theta_t\|^2 + \left(\frac{8}{\xi} - \frac{1}{2\xi}\right)(\lambda_{t+1} - \lambda_t)^2,
\end{aligned} \tag{92}$$

and if $\frac{1}{b_{t+1}} - \frac{1}{b_t} \leq \frac{\xi}{5}$, *then*

$$\begin{aligned}
&F_{t+1} - F_t \\
&\geq S_t + \left(\frac{\mu_t}{2} + \nu - \frac{16(C_\sigma^V)^2}{\xi b_{t+1}^2} - \frac{\xi(C_\sigma^V)^2}{2}\right)\|\theta_{t+1} - \theta_t\|^2 + \frac{b_t - b_{t-1}}{2}\lambda_{t+1}^2 \\
&\quad + \frac{9}{10\xi}(\lambda_{t+1} - \lambda_t)^2 + \frac{8}{\xi}\left(\frac{b_t}{b_{t+1}} - \frac{b_{t-1}}{b_t}\right)\lambda_{t+1}^2,
\end{aligned} \tag{93}$$

where $S_t \triangleq \frac{16}{b_t \xi}(f(\theta_{t-1}) - \hat{f}(\theta_{t-1}) - f(\theta_t) + \hat{f}(\theta_t))(-\lambda_t + \lambda_{t+1}) + (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}))(\lambda_{t+1} - \lambda_t) + \langle \theta_{t+1} - \theta_t, -\hat{g}(\theta_t, \lambda_{t+1}) + g(\theta_t, \lambda_{t+1})\rangle$.

Proof. From eq. (79) and eq. (80), we have that

$$\begin{aligned}
\beta_t m_{t+1}(\lambda_t - \lambda_{t+1}) \geq& (\nabla_\lambda \tilde{V}_{t+1}(\theta_t, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))(-\lambda_t + \lambda_{t+1}) \\
&+ (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}) - f(\theta_t) + \hat{f}(\theta_t))(-\lambda_t + \lambda_{t+1}).
\end{aligned} \tag{94}$$

The first term can be rewritten as

$$\begin{aligned}
&(\nabla_\lambda \tilde{V}_{t+1}(\theta_t, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))(\lambda_{t+1} - \lambda_t) \\
&= (\nabla_\lambda \tilde{V}_{t+1}(\theta_t, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t))(\lambda_{t+1} - \lambda_t) + (\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))(\lambda_{t+1} - \lambda_t) \\
&= (\nabla_\lambda \tilde{V}_{t+1}(\theta_t, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t))(\lambda_{t+1} - \lambda_t) + (\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))(\lambda_t - \lambda_{t-1}) \\
&\quad + m_{t+1}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1})).
\end{aligned} \tag{95}$$

The first term in eq. (95) can be bounded as

$$\begin{aligned}
&(\nabla_\lambda \tilde{V}_{t+1}(\theta_t, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t))(\lambda_{t+1} - \lambda_t) \\
&= (\nabla_\lambda V_\sigma^L(\theta_t, \lambda_t) - \nabla_\lambda V_\sigma^L(\theta_{t-1}, \lambda_t))(\lambda_{t+1} - \lambda_t) + (b_t\lambda_t - b_{t-1}\lambda_t)(\lambda_{t+1} - \lambda_t) \\
&\overset{(a)}{\geq} -\frac{1}{2h}(\nabla_\lambda V_\sigma^L(\theta_t, \lambda_t) - \nabla_\lambda V_\sigma^L(\theta_{t-1}, \lambda_t))^2 - \frac{h}{2}(\lambda_{t+1} - \lambda_t)^2
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{(b_t - b_{t-1})}{2}(\lambda_{t+1}^2 - \lambda_t^2) - \frac{(b_t - b_{t-1})}{2}(\lambda_{t+1} - \lambda_t)^2 \\
&\overset{(b)}{\geq} -\frac{(C_\sigma^V)^2}{2h}\|\theta_t - \theta_{t-1}\|^2 - \frac{h}{2}(\lambda_{t+1} - \lambda_t)^2 \\
&+ \frac{(b_t - b_{t-1})}{2}(\lambda_{t+1}^2 - \lambda_t^2) - \frac{(b_t - b_{t-1})}{2}(\lambda_{t+1} - \lambda_t)^2,
\end{aligned} \tag{96}
$$

where (a) is from the Cauchy–Schwarz inequality and (b) is from the C_σ^V-smoothness of V_σ^L, for any $h > 0$.

Similar to eq. (76), the second term in eq. (95) can be bounded as

$$
\begin{aligned}
&(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))(\lambda_t - \lambda_{t-1}) \\
&\geq \frac{b_{t-1}b_0}{b_{t-1} + b_0}(\lambda_t - \lambda_{t-1})^2 + \frac{1}{b_{t-1} + b_0}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))^2.
\end{aligned} \tag{97}
$$

The third term in eq. (95) can be bounded as

$$
\begin{aligned}
&m_{t+1}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1})) \\
&\geq -\frac{\xi}{2}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))^2 - \frac{1}{2\xi}m_{t+1}^2.
\end{aligned} \tag{98}
$$

Hence combine eq. (96) to eq. (97) and plug in eq. (95), we have that

$$
\begin{aligned}
&(\nabla_\lambda \tilde{V}_{t+1}(\theta_t, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))(\lambda_{t+1} - \lambda_t) \\
&\geq -\frac{(C_\sigma^V)^2}{2h}\|\theta_t - \theta_{t-1}\|^2 - \frac{h}{2}(\lambda_{t+1} - \lambda_t)^2 \\
&+ \frac{(b_t - b_{t-1})}{2}(\lambda_{t+1}^2 - \lambda_t^2) - \frac{(b_t - b_{t-1})}{2}(\lambda_{t+1} - \lambda_t)^2 \\
&+ \frac{b_{t-1}b_0}{b_{t-1} + b_0}(\lambda_t - \lambda_{t-1})^2 + \frac{1}{b_{t-1} + b_0}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))^2 \\
&- \frac{\xi}{2}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))^2 - \frac{1}{2\xi}m_{t+1}^2.
\end{aligned} \tag{99}
$$

Hence eq. (94) can be further bounded as

$$
\begin{aligned}
&(\beta_t m_{t+1})(\lambda_t - \lambda_{t+1}) \\
&\geq (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}) - f(\theta_t) + \hat{f}(\theta_t))(-\lambda_t + \lambda_{t+1}) \\
&- \frac{(C_\sigma^V)^2}{2h}\|\theta_t - \theta_{t-1}\|^2 - \frac{h}{2}(\lambda_{t+1} - \lambda_t)^2 \\
&+ \frac{(b_t - b_{t-1})}{2}(\lambda_{t+1}^2 - \lambda_t^2) - \frac{(b_t - b_{t-1})}{2}(\lambda_{t+1} - \lambda_t)^2 \\
&+ \frac{b_{t-1}b_0}{b_{t-1} + b_0}(\lambda_t - \lambda_{t-1})^2 + \frac{1}{b_{t-1} + b_0}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))^2 \\
&- \frac{\xi}{2}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))^2 - \frac{1}{2\xi}m_{t+1}^2.
\end{aligned} \tag{100}
$$

It can be directly verified that

$$
m_{t+1}(\lambda_t - \lambda_{t+1}) = \frac{1}{2}(\lambda_t - \lambda_{t-1})^2 - \frac{1}{2}(\lambda_t - \lambda_{t+1})^2 - \frac{m_{t+1}^2}{2}. \tag{101}
$$

Recall that $\beta_t = \frac{1}{\xi}$, hence

$$
\begin{aligned}
&\frac{1}{2\xi}(\lambda_t - \lambda_{t-1})^2 - \frac{1}{2\xi}(\lambda_t - \lambda_{t+1})^2 - \frac{m_{t+1}^2}{2\xi} \\
&\geq (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}) - f(\theta_t) + \hat{f}(\theta_t))(-\lambda_t + \lambda_{t+1})
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{(C_\sigma^V)^2}{2h}\|\theta_t - \theta_{t-1}\|^2 - \frac{h}{2}(\lambda_{t+1} - \lambda_t)^2 \\
&+ \frac{(b_t - b_{t-1})}{2}(\lambda_{t+1}^2 - \lambda_t^2) - \frac{(b_t - b_{t-1})}{2}(\lambda_{t+1} - \lambda_t)^2 \\
&+ \frac{b_{t-1}b_0}{b_{t-1}+b_0}(\lambda_t - \lambda_{t-1})^2 + \frac{1}{b_{t-1}+b_0}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))^2 \\
&- \frac{\xi}{2}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))^2 - \frac{1}{2\xi}m_{t+1}^2. \qquad (102)
\end{aligned}
$$

From $\xi \leq \frac{1}{b_0} \leq \frac{2}{b_0+b_{t-1}}$, we have $\frac{1}{b_{t-1}+b_0}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))^2 - \frac{\xi}{2}(\nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_t) - \nabla_\lambda \tilde{V}_t(\theta_{t-1}, \lambda_{t-1}))^2 \geq 0$. Also, it can be shown that $\frac{b_{t-1}b_0}{b_{t-1}+b_0} \geq \frac{b_{t-1}b_0}{2b_0} = \frac{b_{t-1}}{2}$. Thus, it follows that

$$
\begin{aligned}
&\frac{1}{2\xi}(\lambda_t - \lambda_{t-1})^2 - \frac{1}{2\xi}(\lambda_t - \lambda_{t+1})^2 - \frac{m_{t+1}^2}{2\xi} \\
&\geq (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}) - f(\theta_t) + \hat{f}(\theta_t))(-\lambda_t + \lambda_{t+1}) \\
&\quad - \frac{(C_\sigma^V)^2}{2h}\|\theta_t - \theta_{t-1}\|^2 - \frac{h}{2}(\lambda_{t+1} - \lambda_t)^2 \\
&\quad + \frac{(b_t - b_{t-1})}{2}(\lambda_{t+1}^2 - \lambda_t^2) - \frac{(b_t - b_{t-1})}{2}(\lambda_{t+1} - \lambda_t)^2 \\
&\quad + \frac{b_{t-1}}{2}(\lambda_t - \lambda_{t-1})^2 - \frac{1}{2\xi}m_{t+1}^2. \qquad (103)
\end{aligned}
$$

Re-arrange the terms, it follows that

$$
\begin{aligned}
&- \frac{1}{2\xi}(\lambda_t - \lambda_{t+1})^2 - \frac{b_t - b_{t-1}}{2}\lambda_{t+1}^2 \\
&\geq (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}) - f(\theta_t) + \hat{f}(\theta_t))(-\lambda_t + \lambda_{t+1}) - \frac{(C_\sigma^V)^2}{2h}\|\theta_t - \theta_{t-1}\|^2 - \frac{h}{2}(\lambda_{t+1} - \lambda_t)^2 \\
&\quad - \frac{(b_t - b_{t-1})}{2}\lambda_t^2 - \frac{(b_t - b_{t-1})}{2}(\lambda_{t+1} - \lambda_t)^2 + \frac{b_{t-1}}{2}(\lambda_t - \lambda_{t-1})^2 - \frac{1}{2\xi}(\lambda_t - \lambda_{t-1})^2 \\
&\geq - \frac{1}{2\xi}(\lambda_t - \lambda_{t-1})^2 - \frac{(b_t - b_{t-1})}{2}\lambda_t^2 + (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}) - f(\theta_t) + \hat{f}(\theta_t))(-\lambda_t + \lambda_{t+1}) \\
&\quad - \frac{(C_\sigma^V)^2}{2h}\|\theta_t - \theta_{t-1}\|^2 - \frac{h}{2}(\lambda_{t+1} - \lambda_t)^2 + \frac{b_{t-1}}{2}(\lambda_t - \lambda_{t-1})^2, \qquad (104)
\end{aligned}
$$

where the last inequality is from the fact that b_t is decreasing.

Now multiply $\frac{2}{\xi b_t}$ on both sides, we further have that

$$
\begin{aligned}
&- \frac{1}{\xi^2 b_t}(\lambda_t - \lambda_{t+1})^2 - \frac{1}{\xi}\left(1 - \frac{b_{t-1}}{b_t}\right)\lambda_{t+1}^2 \\
&\geq - \frac{1}{\xi^2 b_t}(\lambda_t - \lambda_{t-1})^2 - \frac{1}{\xi}\left(1 - \frac{b_{t-1}}{b_t}\right)\lambda_t^2 + \frac{2}{\xi b_t}(f(\theta_{t-1}) - \hat{f}(\theta_{t-1}) - f(\theta_t) + \hat{f}(\theta_t))(-\lambda_t + \lambda_{t+1}) \\
&\quad - \frac{(C_\sigma^V)^2}{h\xi b_t}\|\theta_t - \theta_{t-1}\|^2 - \frac{h}{\xi b_t}(\lambda_{t+1} - \lambda_t)^2 + \frac{1}{\xi}(\lambda_t - \lambda_{t-1})^2. \qquad (105)
\end{aligned}
$$

If we set $h = \frac{b_t}{2}$, eq. (105) can be rewritten as

$$
\begin{aligned}
&- \frac{1}{\xi^2 b_t}(\lambda_t - \lambda_{t+1})^2 - \frac{1}{\xi}\left(1 - \frac{b_{t-1}}{b_t}\right)\lambda_{t+1}^2 \\
&\geq - \frac{1}{\xi^2 b_t}(\lambda_t - \lambda_{t-1})^2 - \frac{1}{\xi}\left(1 - \frac{b_{t-1}}{b_t}\right)\lambda_t^2 + \frac{2}{\xi b_t}(f(\theta_{t-1}) - \hat{f}(\theta_{t-1}) - f(\theta_t) + \hat{f}(\theta_t))(-\lambda_t + \lambda_{t+1}) \\
&\quad - \frac{2(C_\sigma^V)^2}{\xi b_t^2}\|\theta_t - \theta_{t-1}\|^2 - \frac{1}{2\xi}(\lambda_{t+1} - \lambda_t)^2 + \frac{1}{\xi}(\lambda_t - \lambda_{t-1})^2. \qquad (106)
\end{aligned}
$$

Further we have that

$$-\frac{1}{\xi^2 b_{t+1}}(\lambda_t-\lambda_{t+1})^2+\left(\frac{1}{\xi^2 b_{t+1}}-\frac{1}{\xi^2 b_t}\right)(\lambda_t-\lambda_{t+1})^2-\frac{1}{\xi}\left(1-\frac{b_t}{b_{t+1}}\right)\lambda_{t+1}^2+\frac{1}{\xi}\left(\frac{b_{t-1}}{b_t}-\frac{b_t}{b_{t+1}}\right)\lambda_{t+1}^2$$
$$\geq -\frac{1}{\xi^2 b_t}(\lambda_t-\lambda_{t-1})^2-\frac{1}{\xi}\left(1-\frac{b_{t-1}}{b_t}\right)\lambda_t^2+\frac{2}{\xi b_t}(f(\theta_{t-1})-\hat{f}(\theta_{t-1})-f(\theta_t)+\hat{f}(\theta_t))(-\lambda_t+\lambda_{t+1})$$
$$-\frac{2(C_\sigma^V)^2}{\xi b_t^2}\|\theta_t-\theta_{t-1}\|^2-\frac{1}{2\xi}(\lambda_{t+1}-\lambda_t)^2+\frac{1}{\xi}(\lambda_t-\lambda_{t-1})^2. \tag{107}$$

Re-arranging the terms in eq. (107) implies that

$$-\frac{1}{\xi^2 b_{t+1}}(\lambda_t-\lambda_{t+1})^2-\frac{1}{\xi}\left(1-\frac{b_t}{b_{t+1}}\right)\lambda_{t+1}^2-\left(-\frac{1}{\xi^2 b_t}(\lambda_t-\lambda_{t-1})^2-\frac{1}{\xi}\left(1-\frac{b_{t-1}}{b_t}\right)\lambda_t^2\right)$$
$$\geq -\left(\frac{1}{\xi^2 b_{t+1}}-\frac{1}{\xi^2 b_t}\right)(\lambda_t-\lambda_{t+1})^2-\frac{1}{\xi}\left(\frac{b_{t-1}}{b_t}-\frac{b_t}{b_{t+1}}\right)\lambda_{t+1}^2$$
$$-\frac{2(C_\sigma^V)^2}{\xi b_t^2}\|\theta_t-\theta_{t-1}\|^2-\frac{1}{2\xi}(\lambda_{t+1}-\lambda_t)^2+\frac{1}{\xi}(\lambda_t-\lambda_{t-1})^2$$
$$+\frac{2}{\xi b_t}(f(\theta_{t-1})-\hat{f}(\theta_{t-1})-f(\theta_t)+\hat{f}(\theta_t))(-\lambda_t+\lambda_{t+1})$$
$$\geq -\frac{7}{10\xi}(-\lambda_t+\lambda_{t+1})^2-\frac{2(C_\sigma^V)^2}{\xi b_t^2}\|\theta_t-\theta_{t-1}\|^2+\frac{1}{\xi}(\lambda_t-\lambda_{t-1})^2+\frac{1}{\xi}\left(\frac{b_t}{b_{t+1}}-\frac{b_{t-1}}{b_t}\right)\lambda_{t+1}^2$$
$$+\frac{2}{\xi b_t}(f(\theta_{t-1})-\hat{f}(\theta_{t-1})-f(\theta_t)+\hat{f}(\theta_t))(-\lambda_t+\lambda_{t+1}), \tag{108}$$

where the last inequality is from $\frac{1}{b_{t+1}}-\frac{1}{b_t}\leq\frac{\xi}{5}$. Recall in Lemma 8, we showed that

$$V_\sigma^L(\theta_{t+1},\lambda_{t+1})-V_\sigma^L(\theta_t,\lambda_t)$$
$$\geq (f(\theta_{t-1})-\hat{f}(\theta_{t-1}))(\lambda_{t+1}-\lambda_t)+\langle\theta_{t+1}-\theta_t,-\hat{g}(\theta_t,\lambda_{t+1})+g(\theta_t,\lambda_{t+1})\rangle-\frac{\xi(C_\sigma^V)^2}{2}\|\theta_t-\theta_{t-1}\|^2$$
$$+\left(\frac{\mu_t}{2}+\nu\right)\|\theta_{t+1}-\theta_t\|^2+\frac{b_{t-1}}{2}(\lambda_t^2-\lambda_{t+1}^2)-\frac{1}{\xi}(\lambda_{t+1}-\lambda_t)^2-\frac{1}{2\xi}(\lambda_t-\lambda_{t-1})^2. \tag{109}$$

Combine both inequality together, and we further have that

$$-\frac{8}{\xi^2 b_{t+1}}(\lambda_t-\lambda_{t+1})^2-\frac{8}{\xi}\left(1-\frac{b_t}{b_{t+1}}\right)\lambda_{t+1}^2-\left(-\frac{8}{\xi^2 b_t}(\lambda_t-\lambda_{t-1})^2-\frac{8}{\xi}\left(1-\frac{b_{t-1}}{b_t}\right)\lambda_t^2\right)$$
$$+V_\sigma^L(\theta_{t+1},\lambda_{t+1})-V_\sigma^L(\theta_t,\lambda_t)$$
$$\geq -\frac{28}{5\xi}(-\lambda_t+\lambda_{t+1})^2-\frac{16(C_\sigma^V)^2}{\xi b_t^2}\|\theta_t-\theta_{t-1}\|^2+\frac{8}{\xi}(\lambda_t-\lambda_{t-1})^2+\frac{8}{\xi}\left(\frac{b_t}{b_{t+1}}-\frac{b_{t-1}}{b_t}\right)\lambda_{t+1}^2$$
$$+\frac{16}{b_t\xi}(f(\theta_{t-1})-\hat{f}(\theta_{t-1})-f(\theta_t)+\hat{f}(\theta_t))(-\lambda_t+\lambda_{t+1})$$
$$+(f(\theta_{t-1})-\hat{f}(\theta_{t-1}))(\lambda_{t+1}-\lambda_t)+\langle\theta_{t+1}-\theta_t,-\hat{g}(\theta_t,\lambda_{t+1})+g(\theta_t,\lambda_{t+1})\rangle-\frac{\xi(C_\sigma^V)^2}{2}\|\theta_t-\theta_{t-1}\|^2$$
$$+\left(\frac{\mu_t}{2}+\nu\right)\|\theta_{t+1}-\theta_t\|^2+\frac{b_{t-1}}{2}(\lambda_t^2-\lambda_{t+1}^2)-\frac{1}{\xi}(\lambda_{t+1}-\lambda_t)^2-\frac{1}{2\xi}(\lambda_t-\lambda_{t-1})^2$$
$$=S_t+\left(-\frac{16(C_\sigma^V)^2}{\xi b_t^2}-\frac{\xi(C_\sigma^V)^2}{2}\right)\|\theta_t-\theta_{t-1}\|^2+\left(-\frac{28}{5\xi}-\frac{1}{\xi}\right)(-\lambda_t+\lambda_{t+1})^2+\frac{b_{t-1}}{2}(\lambda_t^2-\lambda_{t+1}^2)$$
$$+\left(\frac{8}{\xi}-\frac{1}{2\xi}\right)(\lambda_t-\lambda_{t-1})^2+\frac{8}{\xi}\left(\frac{b_t}{b_{t+1}}-\frac{b_{t-1}}{b_t}\right)\lambda_{t+1}^2+\left(\frac{\mu_t}{2}+\nu\right)\|\theta_{t+1}-\theta_t\|^2, \tag{110}$$

where $S_t\triangleq\frac{16}{b_t\xi}(f(\theta_{t-1})-\hat{f}(\theta_{t-1})-f(\theta_t)+\hat{f}(\theta_t))(-\lambda_t+\lambda_{t+1})+(f(\theta_{t-1})-\hat{f}(\theta_{t-1}))(\lambda_{t+1}-\lambda_t)+\langle\theta_{t+1}-\theta_t,-\hat{g}(\theta_t,\lambda_{t+1})+g(\theta_t,\lambda_{t+1})\rangle$. Now

$$-\frac{8}{\xi^2 b_{t+1}}(\lambda_t-\lambda_{t+1})^2-\frac{8}{\xi}\left(1-\frac{b_t}{b_{t+1}}\right)\lambda_{t+1}^2-\left(-\frac{8}{\xi^2 b_t}(\lambda_t-\lambda_{t-1})^2-\frac{8}{\xi}\left(1-\frac{b_{t-1}}{b_t}\right)\lambda_t^2\right)$$

$$
\begin{aligned}
&+ V_\sigma^L(\theta_{t+1},\lambda_{t+1}) - V_\sigma^L(\theta_t,\lambda_t) + \frac{b_t}{2}\lambda_{t+1}^2 - \frac{b_{t-1}}{2}\lambda_t^2 \\
&+ \left(-\frac{16(C_\sigma^V)^2}{\xi b_{t+1}^2} - \frac{\xi(C_\sigma^V)^2}{2}\right)\|\theta_{t+1}-\theta_t\|^2 - \left(-\frac{16(C_\sigma^V)^2}{\xi b_t^2} - \frac{\xi(C_\sigma^V)^2}{2}\right)\|\theta_t-\theta_{t-1}\|^2 \\
&+ \left(\frac{8}{\xi} - \frac{1}{2\xi}\right)(\lambda_{t+1}-\lambda_t)^2 - \left(\frac{8}{\xi} - \frac{1}{2\xi}\right)(\lambda_t-\lambda_{t-1})^2 \\
\geq\ & S_t + \left(\frac{\mu_t}{2} + \nu - \frac{16(C_\sigma^V)^2}{\xi b_{t+1}^2} - \frac{\xi(C_\sigma^V)^2}{2}\right)\|\theta_{t+1}-\theta_t\|^2 + \frac{b_t - b_{t-1}}{2}\lambda_{t+1}^2 \\
&+ \left(\frac{8}{\xi} - \frac{1}{2\xi} - \frac{28}{5\xi} - \frac{1}{\xi}\right)(\lambda_{t+1}-\lambda_t)^2 + \frac{8}{\xi}\left(\frac{b_t}{b_{t+1}} - \frac{b_{t-1}}{b_t}\right)\lambda_{t+1}^2 \\
=\ & S_t + \left(\frac{\mu_t}{2} + \nu - \frac{16(C_\sigma^V)^2}{\xi b_{t+1}^2} - \frac{\xi(C_\sigma^V)^2}{2}\right)\|\theta_{t+1}-\theta_t\|^2 + \frac{b_t - b_{t-1}}{2}\lambda_{t+1}^2 \\
&+ \frac{9}{10\xi}(\lambda_{t+1}-\lambda_t)^2 + \frac{8}{\xi}\left(\frac{b_t}{b_{t+1}} - \frac{b_{t-1}}{b_t}\right)\lambda_{t+1}^2, \qquad (111)
\end{aligned}
$$

which then completes the proof. □

We now restate Theorem 2 with all the specific step sizes. The definitions of these constants can also be found in Section J.

Theorem 3. *(Restatement of Theorem 2) Set* $b_t = \frac{19}{20\xi t^{0.25}}$, $\mu_t = \xi(C_\sigma^V)^2 + \frac{16\tau(C_\sigma^V)^2}{\xi(b_{t+1})^2} - 2\nu$, $\beta_t = \frac{1}{\xi}, \alpha_t = \nu + \mu_t$, *where* $\xi > \frac{2\nu+(1+\Lambda^*)L_\sigma}{(C_\sigma^V)^2}$, ν *is any positive number and* τ *is any number greater than* 2. *Moreover, set* $\epsilon_{est} = \frac{1}{t^{0.5}L_\Omega}\frac{1}{32t^{0.25}\Lambda^*+2\Lambda^*+\frac{1}{\alpha_1}(1+\Lambda^*)C_\sigma^V}\frac{19^2\epsilon^2}{3200\xi(\tau-2)(C_\sigma^V)^2uL_\Omega} = \mathcal{O}(\frac{\epsilon^2}{t^{0.75}})$, *then*

$$
\min_{1\leq t\leq T}\|G_t^\sigma\|^2 \leq (1+\sqrt{2})\epsilon, \qquad (112)
$$

when $T = \mathcal{O}(\epsilon^{-4})$.

Proof. Denote by $p_t \triangleq \frac{8(\tau-2)(C_\sigma^V)^2}{\xi b_{t+1}^2}$ and $M_1 \triangleq \frac{16\tau^2}{(\tau-2)^2} + \frac{(\xi(C_\sigma^V)^2-\nu)^2}{64(\tau-2)^2(C_\sigma^V)^2\xi^2}$. Then it can be verified that $\nu + \frac{\mu_t}{2} - \frac{\xi(C_\sigma^V)^2}{2} - \frac{16(C_\sigma^V)^2}{\xi b_{t+1}^2} = p_t$. Then eq. (111) can be rewritten as

$$
\begin{aligned}
F_{t+1} - F_t \geq\ & S_t + p_t\|\theta_{t+1}-\theta_t\|^2 + \frac{b_t-b_{t-1}}{2}\lambda_{t+1}^2 \\
&+ \frac{9}{10\xi}(\lambda_{t+1}-\lambda_t)^2 + \frac{8}{\xi}\left(\frac{b_t}{b_{t+1}} - \frac{b_{t-1}}{b_t}\right)\lambda_{t+1}^2. \qquad (113)
\end{aligned}
$$

From the definition, we have that

$$
G_t^\sigma = \begin{bmatrix} \beta_t\left(\lambda_t - \prod_{[0,\Lambda^*]}\left(\lambda_t - \frac{1}{\beta_t}\left(\nabla_\lambda V_\sigma^L(\theta_t,\lambda_t)\right)\right)\right) \\ \alpha_t\left(\theta_t - \prod_\Theta\left(\theta_t + \frac{1}{\alpha_t}\left(\nabla_\theta V_\sigma^L(\theta_t,\lambda_t)\right)\right)\right) \end{bmatrix}, \qquad (114)
$$

and denote by

$$
\tilde{G}_t \triangleq \begin{bmatrix} \beta_t\left(\lambda_t - \prod_{[0,\Lambda^*]}\left(\lambda_t - \frac{1}{\beta_t}\left(\nabla_\lambda \tilde{V}_t(\theta_t,\lambda_t)\right)\right)\right) \\ \alpha_t\left(\theta_t - \prod_\Theta\left(\theta_t + \frac{1}{\alpha_t}\left(\nabla_\theta \tilde{V}_t(\theta_t,\lambda_t)\right)\right)\right) \end{bmatrix}. \qquad (115)
$$

It can be verified that

$$
\|G_t^\sigma\| - \|\tilde{G}_t\| \leq b_{t-1}|\lambda_t|. \qquad (116)
$$

From Theorem 4.2 in (Xu et al., 2020), it can be shown that

$$
\|\tilde{G}_t\|^2 \leq 2(\mu_t+\nu)^2\|\theta_{t+1}-\theta_t\|^2 + \left(2(C_\sigma^V)^2 + \frac{1}{\xi^2}\right)(\lambda_{t+1}-\lambda_t)^2, \qquad (117)
$$

and

$$M_1 \geq \frac{2(\nu+\mu_t)^2}{p_t^2}. \tag{118}$$

Hence

$$\|\tilde{G}_t\|^2 \leq M_1 p_t^2 \|\theta_{t+1}-\theta_t\|^2 + \left(2(C_\sigma^V)^2 + \frac{1}{\xi^2}\right)(\lambda_{t+1}-\lambda_t)^2. \tag{119}$$

Set $u_t \triangleq \frac{1}{\max\left\{M_1 p_t, \frac{10+20\xi^2(C_\sigma^V)^2}{9\xi}\right\}}$, then from eq. (113), we have that

$$u_t\|\tilde{G}_t\|^2 \leq F_{t+1} - F_t - S_t - \frac{b_t - b_{t-1}}{2}\lambda_{t+1}^2 - \frac{8}{\xi}\left(\frac{b_t}{b_{t+1}} - \frac{b_{t-1}}{b_t}\right)\lambda_{t+1}^2. \tag{120}$$

Summing the inequality above from $t=1$ to T, then

$$\begin{aligned}\sum_{t=1}^T u_t\|\tilde{G}_t\|^2 &\leq F_{T+1} - F_1 - \sum_{t=1}^T S_t + \frac{8}{\xi}\left(\frac{b_0}{b_1}\lambda_2^2 - \frac{b_T}{b_{T+1}}\lambda_{T+1}^2\right) + \left(\frac{b_0-b_T}{2}(\Lambda^*)^2\right)\\ &\leq F_{T+1} - F_1 - \sum_{t=1}^T S_t + \frac{8}{\xi}\frac{b_0}{b_1}(\Lambda^*)^2 + \left(\frac{b_0-b_T}{2}(\Lambda^*)^2\right),\end{aligned} \tag{121}$$

which is from b_t is decreasing and $\lambda_t < \Lambda^*$. Note that

$$\begin{aligned}\max_{t\geq 1}\max_{\theta\in\Theta,\lambda\in[0,\Lambda^*]} F_t &= \max\Bigg\{-\frac{8}{\xi^2 b_{t+1}}(\lambda_t-\lambda_{t+1})^2 - \frac{8}{\xi}\left(1-\frac{b_t}{b_{t+1}}\right)\lambda_{t+1}^2 + V_\sigma^L(\theta_{t+1},\lambda_{t+1}) + \frac{b_t}{2}\lambda_{t+1}^2\\ &\quad + \left(-\frac{16(C_\sigma^V)^2}{\xi b_{t+1}^2} - \frac{\xi(C_\sigma^V)^2}{2}\right)\|\theta_{t+1}-\theta_t\|^2 + \left(\frac{8}{\xi}-\frac{1}{2\xi}\right)(\lambda_{t+1}-\lambda_t)^2\Bigg\}\\ &\leq \frac{1.6}{\xi}(\Lambda^*)^2 + (1+\Lambda^*)(2C_\sigma) + \frac{b_1}{2}(\Lambda^*)^2 + \frac{15}{2\xi}(\Lambda^*)^2\\ &\triangleq F^*,\end{aligned} \tag{122}$$

which is from the definition of b_t, and $8(\frac{b_t}{b_{t+1}}-1) \leq 8(\frac{(t+1)^{0.25}}{t^{0.25}}-1) \leq 8(\frac{2^{0.25}}{1}-1) < 1.6$. Then plugging in the definition of b_t implies that

$$\sum_{t=1}^T u_t\|\tilde{G}_t\|^2 \leq F^* - F_1 - \sum_{t=1}^T S_t + \frac{8}{\xi}(\Lambda^*)^2 + \left(\frac{b_0}{2}(\Lambda^*)^2\right). \tag{123}$$

If moreover set $u \triangleq \max\left\{M_1, \frac{10+20\xi^2(C_\sigma^V)^2}{9\xi p_2}\right\}$, then $u_t \geq \frac{1}{up_t}$, and hence

$$\frac{\sum_{t=1}^T \frac{1}{p_t}\|\tilde{G}_t\|^2}{\sum_{t=1}^T \frac{1}{p_t}} \leq \frac{u}{\sum_{t=1}^T \frac{1}{p_t}}\left(F^* - F_1 - \sum_{t=1}^T S_t + \frac{8}{\xi}(\Lambda^*)^2 + \left(\frac{b_0}{2}(\Lambda^*)^2\right)\right). \tag{124}$$

Plug in the definition of p_t then we have that

$$\frac{\sum_{t=1}^T \frac{1}{p_t}\|\tilde{G}_t\|^2}{\sum_{t=1}^T \frac{1}{p_t}} \leq \frac{3200\xi(\tau-2)(C_\sigma^V)^2 d}{19^2(\sqrt{T}-2)}\left(F^* - F_1 - \sum_{t=1}^T S_t + \frac{8}{\xi}(\Lambda^*)^2 + \left(\frac{b_0}{2}(\Lambda^*)^2\right)\right). \tag{125}$$

We moreover have that

$$\begin{aligned}|S_t| &= \Bigg|\frac{16}{b_t\xi}(f(\theta_{t-1}) - \hat{f}(\theta_{t-1}) - f(\theta_t) + \hat{f}(\theta_t))(-\lambda_t+\lambda_{t+1}) + (f(\theta_{t-1}) - \hat{f}(\theta_{t-1}))(\lambda_{t+1}-\lambda_t)\\ &\quad + \langle\theta_{t+1}-\theta_t, -\hat{g}(\theta_t,\lambda_{t+1}) + g(\theta_t,\lambda_{t+1})\rangle\Bigg|\\ &\leq 32t^{0.25}\Lambda^*(\Omega_{t-1}+\Omega_t) + 2\Lambda^*\Omega_{t-1} + \frac{1}{\alpha_t}(1+\Lambda^*)C_\sigma^V\Omega_t,\end{aligned} \tag{126}$$

where $\Omega_t \triangleq \max\left\{\|g(\theta_t,\lambda_{t+1})-\hat{g}(\theta_t,\lambda_{t+1})\|, |f(\theta_t)-\hat{f}(\theta_t)|\right\}$. Note that it has been shown in (Wang & Zou, 2022) that $\Omega_t \leq L_\Omega \max\left\{\|Q_{\sigma,r}-\hat{Q}_{\sigma,r}\|, \|Q_{\sigma,c}-\hat{Q}_{\sigma,c}\|\right\} = L_\Omega \epsilon_{\text{est}}$, and hence Ω_t can be controlled by setting ϵ_{est}.

Note that $\alpha_t = \nu + \mu_t$ is increasing, hence $\frac{1}{\alpha_t} \leq \frac{1}{\alpha_1}$. Hence if we set $\epsilon_{\text{est}} = \frac{1}{t^{0.5}L_\Omega}\frac{1}{32t^{0.25}\Lambda^*+2\Lambda^*+\frac{1}{\alpha_1}(1+\Lambda^*)C_\sigma^V}\frac{19^2\epsilon^2}{3200\xi(\tau-2)(C_\sigma^V)^2uL_\Omega} = \mathcal{O}(\frac{\epsilon^2}{t^{0.75}})$, then

$$|S_t| \leq \frac{1}{t^{0.5}}\frac{19^2\epsilon^2}{3200\xi(\tau-2)(C_\sigma^V)^2uL_\Omega}, \tag{127}$$

and hence

$$\left|\sum_{t=1}^T S_t\right| \leq \sqrt{T}\frac{19^2\epsilon^2}{3200\xi(\tau-2)(C_\sigma^V)^2uL_\Omega}. \tag{128}$$

Thus plug in eq. (125) and we have that

$$\frac{\sum_{t=1}^T \frac{1}{p_t}\|\tilde{G}_t\|^2}{\sum_{t=1}^T \frac{1}{p_t}} \leq \frac{3200\xi(\tau-2)(C_\sigma^V)^2u}{19^2(\sqrt{T}-2)}K + \epsilon^2, \tag{129}$$

where $K = F^* - F_1 + \frac{8}{\xi}(\Lambda^*)^2 + \left(\frac{b_1}{2}(\Lambda^*)^2\right)$. When $T = (2 + \frac{3200\xi(\tau-2)(C_\sigma^V)^2uK}{19^2\epsilon^2})^2$, we have that

$$\frac{\sum_{t=1}^T \frac{1}{p_t}\|\tilde{G}_t\|^2}{\sum_{t=1}^T \frac{1}{p_t}} \leq 2\epsilon^2. \tag{130}$$

Similarly to Theorem 4.2 in (Xu et al., 2020), if $t > \frac{19^4(\Lambda^*)^4}{2\cdot 10^4\xi^4\epsilon^4}$, then $b_{t-1} < \frac{\epsilon}{\Lambda^*}$ and $b_{t-1}\lambda_t < \epsilon$. Hence combine with eq. (116) we finally have that

$$\min_{1\leq t\leq T}\|G_t^\sigma\| \leq (1+\sqrt{2})\epsilon, \tag{131}$$

when $T = \max\left\{\frac{7(\Lambda^*)^4}{\xi^4\epsilon^4}, \left(2+\frac{9\xi(\tau-2)(C_\sigma^V)^2uK}{\epsilon^2}\right)^2\right\} = \mathcal{O}(\epsilon^{-4})$. □

Remark 1. *Note that the sample complexity of robust TD algorithm to achieve an ϵ_{est}-error bound is $\mathcal{O}(\epsilon_{est}^{-2})$, hence the sample complexity at the time step t is $\mathcal{O}(\epsilon_{est}^{-2}) = \mathcal{O}(\frac{t^{1.5}}{\epsilon^4})$. Thus the total sample complexity to find an ϵ-stationary solution is $\sum_{t=1}^T \frac{t^{1.5}}{\epsilon^4} = \mathcal{O}(\epsilon^{-14})$. This great increasing of complexity is due to the estimation of robust value functions.*

J Constants and Notations

In this section, we summarize the definitions of all the constants we used in this paper.

$$\begin{aligned}
L_V &= \frac{k|\mathcal{A}|}{(1-\gamma)^2},\\
C_\sigma &= \frac{1}{1-\gamma}(1+2\gamma\delta\frac{\log|\mathcal{S}|}{\sigma}),\\
C_\sigma^V &= \frac{1}{1-\gamma}|\mathcal{A}|kC_\sigma,\\
k_B &= \frac{1}{1-\gamma+\gamma\delta}\left(|\mathcal{A}|C_\sigma l + |\mathcal{A}|kC_\sigma^V\right) + \frac{2|\mathcal{A}|^2\gamma(1-\delta)}{(1-\gamma+\gamma\delta)^2}k^2C_\sigma,\\
L_\sigma &= k_B + \frac{\gamma\delta}{1-\gamma}\left(\sqrt{|\mathcal{S}|}k_B + 2\sigma|\mathcal{S}|C_\sigma^V\frac{1}{1-\gamma+\gamma\delta}k|\mathcal{A}|C_\sigma\right),\\
b_t &= \frac{19}{20\xi t^{0.25}},
\end{aligned}$$

$$
\begin{aligned}
M_1 &= \frac{16\tau^2}{(\tau-2)^2} + \frac{(\xi(C_\sigma^V)^2-\nu)^2}{64(\tau-2)^2(C_\sigma^V)^2\xi^2}, \\
u &= \max\left\{M_1, \frac{10+20\xi^2(C_\sigma^V)^2}{9\xi p_2}\right\}, \\
F^* &= \frac{1.6}{\xi}(\Lambda^*)^2 + (1+\Lambda^*)(2C_\sigma) + \frac{b_1}{2}(\Lambda^*)^2 + \frac{15}{2\xi}(\Lambda^*)^2, \\
K &= F^* - F_1 + \frac{8}{\xi}(\Lambda^*)^2 + \left(\frac{b_1}{2}(\Lambda^*)^2\right), \\
\mu_t &= \xi(C_\sigma^V)^2 + \frac{16\tau(C_\sigma^V)^2}{\xi(b_{t+1})^2} - 2\nu, \\
\beta_t &= \frac{1}{\xi}, \\
\alpha_t &= \nu + \mu_t.
\end{aligned} \tag{132}
$$