Multi-Source Predictive Distributed Voltage Control (Supplementary Proofs)

1 Proof of Theorem III.1

We first give a complete statement for Theorem III.1 in the main paper.

Theorem 1.1 (ISS bound)**.** *Let L_H, μ_H be the maximum and minimum eigenvalue of the Karush-Kuhn-Tucker matrix defined in Lemma 1, and define*

$$\alpha_H := \left(\frac{L_H^2 - \mu_H^2}{L_H^2 + \mu_H^2}\right)^{-\frac{1}{2}}, \quad \rho_H := \left(\frac{L_H^2 - \mu_H^2}{L_H^2 + \mu_H^2}\right)^{\frac{1}{2}},$$

$$\delta_1 := \frac{1+\rho_H}{2}, \quad \Gamma := \sup_{d \in \mathcal{N}} \left(\frac{\rho_H}{\delta_1}\right)^d \mathsf{SE}(d),$$

$$\beta_1 := \frac{L_H\, \alpha_H\, \delta_1\, \Gamma}{\mu_H^2(1-\delta_1)}, \quad \Xi := \frac{2L\, L_H\, \alpha_H\, b_H\, \Gamma\, \mathsf{SE}(1)}{\mu_H^2(1-\delta_1)},$$

$$\beta_2 := \frac{2L}{\mu}, \quad \delta_2 := 1 - 2\left(\sqrt{1+\beta_2}+1\right)^{-1}.$$

Set $0 < \gamma \le 1 - \max\{\delta_1, \delta_2\} \in (0,1)$. For any $i \in \mathcal{N}$, if the locality radius κ and the prediction horizon k such that

$$k \ge \underline{k} = \frac{\ln \frac{2}{\gamma(1-\gamma)}}{\ln \frac{\alpha_H}{1-\gamma}}, \quad \kappa \ge 1 + \frac{\ln(2\underline{k} L N \beta_1(\beta_2+1))}{\ln \frac{1}{1-\gamma}}$$

where $0 < \gamma < 1$, $N := |\mathcal{N}|$. Then under Assumption 1, 3, 4 (in the main body), the global state x_t generated by Algorithm 1 (in the main body) satisfies, for all $t \in [T]$,

$$\|\mathrm{x}(t)\| \;\le\; C_x(1-\gamma)^{\max(t-k,0)}\|\mathrm{x}(0)\| + \frac{C_e}{\gamma} \sup_{0\le s\le t-1} e_k(s) + \frac{C}{\gamma},$$

where $C_x, C_e, C > 0$ are independent of t, and one of their definition is given in (27)*. Consequently, the closed-loop system is stable with respect to the weighted prediction error $e_k(s) := \sup_{i\in\mathcal{N}} \sum_{\tau=t}^{\bar{t}} (1-\gamma)^{\tau-t} \|(\xi(\widehat{s}_i(\tau|t)))_{(i,\kappa)} - (\xi(s(\tau)))_{(i,\kappa)}\|$, $s \in [t]$.*

Proof Outline. We begin the proof of Theorem III.1 with a three-step roadmap.

Step 1 Prove exponential-decay gap between the solution to the localized MPC problem (1a) versus the solution to the centralized one (2a). The main idea is to construct an equation 6 using the KKT conditions of both, and to use Lemma 1 to reveal the intrinsic properties of the KKT matrix. Then, the equations are combined to obtain the interim results.

Step 2 By referencing and extending existing error-based theoretical results (Corollary 1.1, Lemma 2, and Lemma 3), we reveal the performance gaps caused by different perturbations $\widehat{\mathbf{w}}, \mathbf{w}$.

Step 3 In final step, we proves the the upper bound of $\|\mathrm{x}(t)\|$ with respected to initial state $\|\mathrm{x}(0)\|$, weighted summation prediction error $\mathrm{e}_k(t)$. The main idea is to construct a induction hypothesis $\|\mathrm{x}(t)\| \le C_x(1-\gamma)^{t-k}\|\mathrm{x}(0)\| + \sum_{l=0}^{t+1} C_e(1-\gamma)^l \sup_{0\le s\le t-1} \mathrm{e}_k(s) + \frac{C}{\gamma}$

Now we begin the formal proof.

Step 1 We first recall the localized MPC problem used in Algorithm 1:

$$\Psi_{t,i}^{k}\left(\mathrm{x}_{(i,\kappa)}(t),\widehat{s}_i(t:\bar{t}|t)\right):=\underset{\widetilde{\mathrm{u}}(0:k-1),\widetilde{\mathrm{x}}(1:k)}{\arg\min}\sum_{\tau=0}^{k-1}(f_{\tau+t}(\widetilde{\mathrm{x}}(\tau))+g_{\tau+t}(\widetilde{\mathrm{u}}(\tau)))+F(\widetilde{\mathrm{x}}(k)) \tag{1a}$$

$$\text{s.t.}\quad \widetilde{\mathrm{x}}(\tau+1)=\widetilde{\mathrm{x}}(\tau)+X_{(i,\kappa)}\widetilde{\mathrm{u}}(\tau)+\xi_{(i,\kappa)}(\widehat{s}_i(t+\tau|t)) \tag{1b}$$

$$\widetilde{\mathrm{x}}(0)=\mathrm{x}_{(i,\kappa)}(t),\quad 0\le\tau\le k \tag{1c}$$

We also use the notation $\Psi\left(\mathrm{x}_{(i,\kappa)}(t),\widehat{s}_i(t:\bar{t}|t),\mathrm{y}_{(i,\kappa)}(t+k)\right)=\Psi_{t,i}^{k}\left(\mathrm{x}_{(i,\kappa)}(t),\widehat{s}_i(t:\bar{t}|t)\right)$ in the following content, where $\mathrm{y}_{(i,\kappa)}(t+k)$ denotes the terminal state. Then we construct the corresponding object without any structural sparsity:

$$\Psi_{t}^{k}\left(\mathrm{x}(t),\widehat{s}_i(t:\bar{t}|t)\right):=\underset{\widetilde{\mathrm{u}}(0:k-1),\widetilde{\mathrm{x}}(1:k)}{\arg\min}\sum_{\tau=0}^{k-1}(f_{\tau+t}(\widetilde{\mathrm{x}}(\tau))+g_{\tau+t}(\widetilde{\mathrm{u}}(\tau)))+F(\widetilde{\mathrm{x}}(k)) \tag{2a}$$

$$\text{s.t.}\quad \widetilde{\mathrm{x}}(\tau+1)=\widetilde{\mathrm{x}}(\tau)+X\widetilde{\mathrm{u}}(\tau)+\xi(\widehat{s}_i(t+\tau|t)) \tag{2b}$$

$$\widetilde{\mathrm{x}}(0)=\mathrm{x}(t),\quad 0\le\tau\le k \tag{2c}$$

Since the problems are FTOCP, we now investigate the Karush-Kuhn-Tucker (KKT) condition of Ψ_i and Ψ.

Let $\psi:=(\widetilde{\mathrm{x}}(0),\widetilde{\mathrm{u}}(0),\widetilde{\mathrm{x}}(1),\ldots,\widetilde{\mathrm{u}}(T-1),\widetilde{\mathrm{x}}(T))$ be the regrouped trajectory of primal variables from Ψ and ψ^{c} be the same regroup order from ψ; $J_k(\psi):=\sum_{\tau=0}^{k-1}(f_\tau(\widetilde{\mathrm{x}}(\tau))+g_\tau(\widetilde{\mathrm{u}}(\tau)))+F(\widetilde{\mathrm{x}}(k))$ denoting the accumulative cost functions, and the constraint Jacobian function:

$$Y(A,B)=\begin{bmatrix} I & & & & & & \\ -A & -B & I & & & & \\ & & -A & -B & I & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & -A & -B & I \end{bmatrix}. \tag{3}$$

Let η,η^{c} denote the trajectory of dual variables in ψ and ψ^{c}, respectively. It is suffice to write down the KKT condition:

$$\begin{bmatrix}\nabla J_k(\psi)+(Y(I_{(i,\kappa)},X_{(i,\kappa)}))^\top\eta\\ Y(I_{(i,\kappa)},X_{(i,\kappa)})\psi\end{bmatrix}=\underbrace{\begin{bmatrix}0\\ \begin{bmatrix}\mathrm{x}_{(i,\kappa)}(t)\\ \xi_{(i,\kappa)}(\widehat{s}_i(t:t+k-1|t))\end{bmatrix}\end{bmatrix}}_{:=\omega_{(i,\kappa)}}, \tag{4}$$

$$\begin{bmatrix}\nabla J_k(\psi^{\mathrm{c}})+(Y(I,X))^\top\eta^{\mathrm{c}}\\ Y(I,X)\psi^{\mathrm{c}}\end{bmatrix}=\underbrace{\begin{bmatrix}0\\ \begin{bmatrix}\mathrm{x}(t)\\ \xi(\widehat{s}_i(t:t+k-1|t))\end{bmatrix}\end{bmatrix}}_{:=\omega}. \tag{5}$$

By assumption of cost functions that $J_k(\cdot)$ is μ-strongly convex, L-smooth, and twice continuously-differentiable, Lemma 1 in [2] tells us that there exists a symmetric matrix $M(\psi,\psi^{\mathrm{c}})$ satisfying the condition that $\mu I\preceq M(\psi,\psi^{\mathrm{c}})\preceq LI$ and $\nabla J_k(\psi)-\nabla J_k(\psi^{\mathrm{c}})=M(\psi,\psi^{\mathrm{c}})(\psi-\psi^{\mathrm{c}})$. Subtracting the left-hand sides of (4) and (5) we get

$$\begin{bmatrix}M(\psi,\psi^{\mathrm{c}}) & (Y(I_{(i,\kappa)},X_{(i,\kappa)}))^\top\\ Y(I_{(i,\kappa)},X_{(i,\kappa)}) & 0\end{bmatrix}\begin{bmatrix}\psi\\ \eta\end{bmatrix}-\begin{bmatrix}M(\psi,\psi^{\mathrm{c}}) & (Y(I,X))^\top\\ Y(I,X) & 0\end{bmatrix}\begin{bmatrix}\psi^{\mathrm{c}}\\ \eta^{\mathrm{c}}\end{bmatrix}=\omega-\omega_{(i,\kappa)}. \tag{6}$$

We now employ the following lemma to proceed the proof.

Lemma 1 (Theorem A.3 in [4]). *Consider the matrices M and Y defined above, and topology group $\mathcal{G} = (\mathcal{N}, \mathcal{E})$. Then there exist constants $0 \leq \mu_H \leq L_H$ and $b_H \in \mathbb{N}_+$ that the constructed KKT matrix $H := \begin{bmatrix} M & Y \\ Y & 0 \end{bmatrix}$ satisfies a lower-upper bound $\mu_H I \preceq H \preceq L_H I$ and its inverse matrix is bounded by a exponentially decaying form with respected to $\mathcal{G}$ by $\|H^{-1}_{ij}\| \leq \frac{L_H}{\mu_H^2}\left(\frac{L_H^2-\mu_H^2}{L_H^2+\mu_H^2}\right)^{\frac{d_{\mathcal{G}}(i,j)-b_H}{2b_H}}$.*

The proof of Lemma 1 is in [4]. Combining this result with (6) yields the following upper bound:

$$
\begin{aligned}
&\|[\psi-\psi^{\mathrm{c}}]_i\| \leq \left\|\left[\begin{bmatrix}\psi\\ \eta\end{bmatrix} - \begin{bmatrix}\psi^{\mathrm{c}}\\ \eta^{\mathrm{c}}\end{bmatrix}\right]_i\right\| \\
&\leq \left\|\sum_{j\in\mathcal{N}} (H^{\mathrm{c},-1})_{ij}\left((\omega-\omega_{(i,\kappa)})_j - \sum_{\ell\in\mathcal{N}}(H_{j\ell}-H^{\mathrm{c}}_{j\ell})\begin{bmatrix}\psi\\ \eta\end{bmatrix}_\ell\right)\right\| \\
&\leq \frac{L_H}{\mu_H^2}\sum_{j\in(\mathcal{N}-\mathcal{N}^\kappa_{\mathcal{G}(i)})} \alpha_H \rho_H^{d_{\mathcal{G}}(i,j)}\left(\|\omega-\omega_{(i,\kappa)}\| + 2L\sum_{\ell\in\mathcal{N}^{b_H}_{\mathcal{G}}(j)}\sum_{v\in\mathcal{N}^\kappa_{\mathcal{G}}(i)}\|(H^{-1})_{\ell v}\|\|\omega_{(i,\kappa)}\|\right) \\
&\leq \frac{L_H}{\mu_H^2}\sum_{d\geq\kappa+1}\mathsf{SE}(d)\alpha_H\rho_H^d\left(\|\omega-\omega_{(i,\kappa)}\| + 2Lb_H\mathsf{SE}(1)\sum_{s\leq 2\kappa}p(s)\alpha_H\rho_H^s\|\omega_{(i,\kappa)}\|\right) \\
&\leq \underbrace{\frac{L_H\alpha_H\delta_1\Gamma}{\mu_H^2(1-\delta_1)}\left(\|\omega-\omega_{(i,\kappa)}\| + \frac{2LL_H\alpha_H b_H\Gamma\mathsf{SE}(1)}{\mu_H^2(1-\delta_1)}\|\omega_{(i,\kappa)}\|\right)}_{:=\beta_1(\|\mathrm{x}(t)\|+\Xi D_k)}\delta_1^\kappa
\end{aligned}
$$

where $D_k := \sum_{\tau=0}^{k-1}\|\mathrm{w}(t+\tau)\|$ and

$$
\alpha_H := \left(\frac{L_H^2-\mu_H^2}{L_H^2+\mu_H^2}\right)^{-\frac{1}{2}}, \quad \rho_H := \left(\frac{L_H^2-\mu_H^2}{L_H^2+\mu_H^2}\right)^{\frac{1}{2b_H}}, \quad \Gamma = \sup_{d\in\mathcal{N}}\left(\frac{\rho_H}{\delta_1}\right)^d\mathsf{SE}(d), \quad \delta_1 = \frac{1+\rho_H}{2}
$$

Step 2 In the following content, we will present several auxiliary corollary and lemmas for further proof.

Corollary 1.1. *Let ψ and $\psi^\star$ denote the optimal primal solution to $\Psi^k_{t,i}(\mathrm{x}(t),\widehat{\omega})$ and $\Psi^k_t(\mathrm{x}^\star(t),\omega)$, and ψ and $\psi^\star$ is partitioned by the temporal topology $\mathcal{G}_k := \{\{0,\dots,k-1\},\{(0,1),\dots,(k-1,k)\}\}$. Let $\tau = 0,\dots,k-1$, and we have:*

$$
\|[\psi-\psi^\star](\tau)\| \leq \alpha_H\left(\rho_H^\tau\|\mathrm{x}(t)-\mathrm{x}^\star(t)\| + \sum_{l=0}^{k-1}\rho_H^{|l-\tau|}\|\mathrm{w}(t+l)-\widehat{\mathrm{w}}(t+l)\|\right)
$$

where α_H and ρ_H follows from the previous content. Specially, if $\psi^\star$ is the solution to $\Psi^k_t(0,0)$, then

$$
\|\psi(\tau)\| \leq \alpha_H\left(\rho_H^\tau\|\mathrm{x}(t)\| + \sum_{l=0}^{k-1}\rho_H^{|l-\tau|}\|\widehat{\mathrm{w}}(t+l)\|\right)
$$

Proof.

$$
\begin{aligned}
\|\psi(\tau)-\psi^\star(\tau)\| &\leq \left\|\begin{bmatrix}\psi(\tau)\\ \eta(\tau)\end{bmatrix} - \begin{bmatrix}\psi^{\mathrm{c}}(\tau)\\ \eta^{\mathrm{c}}(\tau)\end{bmatrix}\right\| = \left\|[H^{\mathrm{c},-1}(\widehat{\omega}-\omega)](\tau)\right\| \\
&\leq \alpha_H\left(\rho_H^\tau\|\mathrm{x}(t)-\mathrm{x}^\star(t)\| + \sum_{l=0}^{k-1}\rho_H^{|l-\tau|}\|\mathrm{w}(t+l)-\widehat{\mathrm{w}}(t+l)\|\right).
\end{aligned}
$$

□

Lemma 2. *Under Assumption 4 in the main content, for any state-perturbation groups* $(\mathrm{x}, \mathbf{w}, \mathrm{y})$ *and* $(\mathrm{x}', \mathbf{w}', \mathrm{y}')$ *with* $0 \le t \le T-k$*, we have*

$$\begin{aligned}&\|\Psi(\mathrm{x}, \mathbf{w}, \mathrm{y})_{\widehat{\mathrm{x}}(\tau)} - \Psi(\mathrm{x}', \mathbf{w}', \mathrm{y}')_{\widehat{\mathrm{x}}(\tau)}\| \\ &\le \beta_2 \left(\delta_2^{\tau} \|\mathrm{x} - \mathrm{x}'\| + \delta_2^{k-\tau} \|\mathrm{y} - \mathrm{y}'\| + \sum_{l=0}^{k-1} \delta_2^{|l-\tau|} \|\mathrm{w}(t+l) - \mathrm{w}'(t+l)\| \right)\end{aligned}$$

where $\beta_2 = \frac{2L}{\mu}$ *and* $\delta_2 = 1 - 2\left(\sqrt{1+\beta_2}+1\right)$.

This lemma follows directly from Theorem 3.3 in [1], therefore, we omit the proof.

Lemma 3. *Suppose the assumption of Theorem III.1 hold. We denote* $\mathbf{w}$ *and* $\widehat{\mathbf{w}}$ *as the ground-truth disturbance and prediction trajectories over any MPC horizon, respectively. Let* $\mathrm{x}(t), \mathrm{x}^\star(t)$ *be the one-step states updated by localized policy with prediction, and centralized optimal policy, respectively. Then for any* $k, \tau \in [T]$ *with condition* $\tau \le k$*, and* $t < T-k$*, we have*

$$\begin{aligned}&\left\|\Psi_t^{k+1}(\mathrm{x}(t), \widehat{\mathbf{w}})_{\widehat{\mathrm{x}}(\tau)} - \Psi_t^{k}(\mathrm{x}^\star(t), \widehat{\mathbf{w}})_{\widehat{\mathrm{x}}(\tau)}\right\| \\ &\le \beta_2^2 \sum_{l=0}^{k-1} \|\varepsilon(t+l-1)\|(1-\gamma)^{\tau+l} \\ &+ LN\beta_1\beta_2(\|\mathrm{x}(t-1)\| + \Xi D_k)(1-\gamma)^{\tau+\kappa} \\ &+ \alpha_H\beta_2\left((1-\gamma)^k(\|\mathrm{x}(t)\| + \alpha_H\rho_H\|\mathrm{x}(t-1)\|) + \frac{6}{1-\rho_H}\sup_{0\le l\le T-1}\|\widehat{\mathrm{w}}(t)\|\right)(1-\gamma)^{k-\tau}\end{aligned}$$

where $\mathrm{x}^\star(t) := \Psi_{t-1}^{k+1}(\mathrm{x}(t-1))_{\widehat{\mathrm{x}}(1)}$, $1-\gamma := \max(\delta_1, \delta_2) \in (0,1)$*, and* $\varepsilon(t) := \mathrm{w}(t) - \widehat{\mathrm{w}}(t)$.

Proof. We first define that:

$$\mathrm{y} := \Psi_t^{k+1}(\mathrm{x}(t), \widehat{\mathbf{w}})_{\widehat{\mathrm{x}}(k)}, \quad \mathrm{y}' := \Psi_t^{k}(\mathrm{x}(t), \widehat{\mathbf{w}})_{\widehat{\mathrm{x}}(k)}, \tag{7}$$

Then it can be observed that

$$\Psi_t^{k+1}(\mathrm{x}(t), \widehat{\mathbf{w}})_{\widehat{\mathrm{x}}(\tau)} = \Psi_t^{k}(\mathrm{x}(t), \widehat{\mathbf{w}}, \mathrm{y})_{\widehat{\mathrm{x}}(\tau)}, \quad \Psi_t^{k}(\mathrm{x}^\star(t))_{\widehat{\mathrm{x}}(\tau)} = \Psi_t^{k}(\mathrm{x}^\star(t), \widehat{\mathbf{w}}, \mathrm{y}')_{\widehat{\mathrm{x}}(\tau)},$$

By Lemma 2, it is straightforward to derive that

$$\begin{aligned}&\left\|\Psi_t^{k+1}(\mathrm{x}(t), \widehat{\mathbf{w}})_{\widehat{\mathrm{x}}(\tau)} - \Psi_t^{k}(\mathrm{x}^\star(t), \widehat{\mathbf{w}})_{\widehat{\mathrm{x}}(\tau)}\right\| \\ &= \left\|\Psi_t^{k}(\mathrm{x}(t), \widehat{\mathbf{w}}, \mathrm{y})_{\widehat{\mathrm{x}}(\tau)} - \Psi_t^{k}(\mathrm{x}^\star(t), \widehat{\mathbf{w}}, \mathrm{y}')_{\widehat{\mathrm{x}}(\tau)}\right\| \\ &\le \beta_2\left(\delta_2^{\tau}\|\mathrm{x}(t) - \mathrm{x}^\star(t)\| + \delta_2^{k-\tau}\|\mathrm{y} - \mathrm{y}'\|\right) \\ &\le \beta_2\delta_2^{\tau}(\|\mathrm{x}^c(t) - \mathrm{x}^\star(t)\| + \|\mathrm{x}^c(t) - \mathrm{x}(t)\|) + \alpha_H\beta_2\delta_2^{k-\tau}\left(\rho_H^k(\|\mathrm{x}(t)\| + \|\mathrm{x}^c(t)\|) + \frac{2}{1-\rho_H}\sup_{0\le\tau\le T-1}\|\widehat{\mathrm{w}}(\tau)\|\right) \\ &\le \beta_2\delta_2^{\tau}\left(\beta_2\sum_{l=0}^{k-1}\delta_2^{l}\|\varepsilon(t+l-1)\| + LN\beta_1(\|\mathrm{x}(t-1)\| + \Xi D_k)(1-\gamma)^{\kappa}\right) \\ &\quad + \alpha_H\beta_2\delta_2^{k-\tau}\left(\rho_H^k(\|\mathrm{x}(t)\| + \|\mathrm{x}^c(t)\|) + \frac{4}{1-\rho_H}\sup_{0\le\tau\le T-1}\|\widehat{\mathrm{w}}(\tau)\|\right) \\ &\le \beta_2^2\sum_{l=0}^{k-1}(1-\gamma)^{\tau+l}\|\varepsilon(t+l-1)\| + LN\beta_1\beta_2(\|\mathrm{x}(t-1)\| + \Xi D_k)(1-\gamma)^{\tau+\kappa} \\ &\quad + \alpha_H\beta_2\left((1-\gamma)^k(\|\mathrm{x}(t)\| + \alpha_H\rho_H\|\mathrm{x}(t-1)\|) + \frac{6}{1-\rho_H}\sup_{0\le\tau\le T-1}\|\widehat{\mathrm{w}}(\tau)\|\right)(1-\gamma)^{k-\tau}\end{aligned}$$

where the second, third inequality both employ the result of Corollary 1.1. ☐

Step 3 It now suffices to proof the main results of Theorem III.1.

Proof of Theorem III.1. For $k \le t+1 \le T-k$,

$$
\begin{aligned}
\|\mathrm{x}(t+1)\| &= \sum_{i=1}^{N} \|x_i(t+1)\| \\
&\le \|\Psi_t^k(\mathrm{x}(t))_{\widehat{\mathrm{x}}(1)}\| + \sum_{i=1}^{N} (\|x_i^\star(t+1) - x_i^c(t+1)\| + \|x_i^c(t+1) - x_i(t+1)\|) \\
&\le \sum_{\tau=0}^{k-2} \left\| \Psi_{t-\tau}^k(\mathrm{x}(t-\tau))_{\widehat{\mathrm{x}}(\tau+1)} - \Psi_{t-\tau-1}^k(\mathrm{x}(t-\tau-1))_{\widehat{\mathrm{x}}(\tau+2)} \right\| + \left\| \Psi_{t-k+1}^k(\mathrm{x}(t-k+1))_{\widehat{\mathrm{x}}(k)} \right\| \\
&\quad + \sum_{l=0}^{k-1} (1-\gamma)^l \|\varepsilon(t+l)\| + LN\beta_1(\|\mathrm{x}(t)\| + \Xi D_k)(1-\gamma)^\kappa \\
&= \sum_{\tau=0}^{k-2} \left\| \Psi_{t-\tau}^k(\mathrm{x}(t-\tau))_{\widehat{\mathrm{x}}(\tau+1)} - \Psi_{t-\tau}^{k-1}(\mathrm{x}^\star(t-\tau))_{\widehat{\mathrm{x}}(\tau+1)} \right\| + \left\| \Psi_{t-k+1}^k(\mathrm{x}(t-k+1))_{\widehat{\mathrm{x}}(k)} \right\| \\
&\quad + \sum_{l=0}^{k-1} (1-\gamma)^l \|\varepsilon(t+l)\| + LN\beta_1(\|\mathrm{x}(t)\| + \Xi D_k)(1-\gamma)^\kappa \\
&\le \sum_{\tau=0}^{k-2} \Bigg(\beta_2^2 \sum_{l=0}^{k-1} (1-\gamma)^{\tau+l+1} \|\varepsilon(t-\tau+l-1)\| + LN\beta_1\beta_2 \big(\|\mathrm{x}(t-\tau-1)\| + \Xi D_k\big)(1-\gamma)^{\tau+\kappa+1} \\
&\quad + \alpha_H\beta_2 \Big((1-\gamma)^k \big(\|\mathrm{x}(t-\tau)\| + \alpha_H\rho_H \|\mathrm{x}(t-\tau-1)\|\big) + \frac{6}{1-\rho_H} \sup_{0\le s\le T-1} \|\widehat{\mathrm{w}}(s)\| \Big)(1-\gamma)^{k-\tau-2} \Bigg) \\
&\quad + \alpha_H^k \|\mathrm{x}(t-k+1)\| + \frac{2}{1-\rho_H} \sup_{0\le s\le T-1} \|\widehat{\mathrm{w}}(s)\| \\
&\quad + \sum_{l=0}^{k-1} (1-\gamma)^l \|\varepsilon(t+l)\| + LN\beta_1 \big(\|\mathrm{x}(t)\| + \Xi D_k\big)(1-\gamma)^\kappa.
\end{aligned} \tag{8}
$$

where the second inequality employs the first step's upper bound, and the third inequality employs the result of Lemma 3.

By absorbing $\alpha_H^k \|\mathrm{x}(t-k+1)\|$ into the same structure via

$$
\alpha_H^k \|\mathrm{x}(t-k+1)\| \le \alpha_H^k (1-\gamma)^{-(k+1)} (1-\gamma)^{2k-(k-1)} \|\mathrm{x}(t-k+1)\|,
$$

and we define

$$
\begin{aligned}
C_1 &:= \alpha_H\beta_2 (1-\gamma)^{-2} + \alpha_H^2\rho_H\beta_2 (1-\gamma)^{-1} + \alpha_H^k (1-\gamma)^{-(k+1)}, \\
C_2 &:= LN\beta_1\beta_2 (1-\gamma) + LN\beta_1.
\end{aligned} \tag{9}
$$

Define the length-k sliding-window error

$$
\mathrm{e}_k(s) := \sum_{m=0}^{k-1} (1-\gamma)^m \|\varepsilon(s+m)\|. \tag{10}
$$

Using time-shift reindexing, one obtains

$$
\sum_{\tau=0}^{k-2} \sum_{l=0}^{k-1} (1-\gamma)^{\tau+l+1} \|\varepsilon(t-\tau+l-1)\| \le \beta_2^2 \sum_{\tau=0}^{k-1} (1-\gamma)^\tau \mathrm{e}_k(t-\tau),
$$

and

$$\sum_{l=0}^{k-1}(1-\gamma)^l\|\varepsilon(t+l)\| \le \sum_{\tau=0}^{k-1}(1-\gamma)^\tau \mathrm{e}_k(t-\tau)$$

and we define

$$C_3 := \beta_2^2 + 1. \tag{11}$$

From the terms involving $\widehat{\mathrm{w}}$ and D_k, we use geometric-series bounds

$$\sum_{\tau=0}^{k-2}(1-\gamma)^{k-\tau-2} \le \frac{1}{\gamma}, \qquad \sum_{\tau=0}^{k-2}(1-\gamma)^{\tau+\kappa+1} \le \frac{(1-\gamma)^{\kappa+1}}{\gamma},$$

and define

$$\begin{aligned} C_4 := &\left(\frac{6\alpha_H\beta_2}{1-\rho_H}\cdot\frac{1}{\gamma} + \frac{2}{1-\rho_H}\right)\sup_{0\le s\le T-1}\|\widehat{\mathrm{w}}(s)\| \\ &+ \frac{LN\beta_1\beta_2\Xi D_k}{\gamma} + LN\beta_1\Xi D_k. \end{aligned} \tag{12}$$

Then with the definitions (9)–(12), inequality (8) can be compactly written as

$$\|\mathrm{x}(t+1)\| \le \sum_{\tau=0}^{k-1}\big(C_1(1-\gamma)^{2k-\tau} + C_2(1-\gamma)^{\kappa+\tau}\big)\|\mathrm{x}(t-\tau)\| + \sum_{\tau=0}^{k-1}C_3(1-\gamma)^\tau \mathrm{e}_k(t-\tau) + C_4. \tag{13}$$

By shifting the index to start at $\tau = 1$, (13) is equivalently

$$\|\mathrm{x}(t+1)\| \le \sum_{\tau=1}^{k}\big(C_1(1-\gamma)^{2k-\tau+1} + C_2(1-\gamma)^{\kappa+\tau-1}\big)\|\mathrm{x}(t-\tau+1)\| + \sum_{\tau=0}^{k-1}C_3(1-\gamma)^\tau \mathrm{e}_k(t-\tau) + C_4. \tag{14}$$

It now suffices to proof the results of Theorem III.1 by induction. Suppose that for for each $\tau \in \{1,\dots,k\}$,

$$\|\mathrm{x}(t-\tau+1)\| \le C_x(1-\gamma)^{t-\tau+1-k}\|\mathrm{x}(0)\| + \sum_{l=0}^{t-\tau}C_e(1-\gamma)^l \sup_{0\le s\le t-\tau}\mathrm{e}_k(s) + \frac{C}{\gamma}. \tag{15}$$

Substituting (15) into (14) yields

$$\begin{aligned} \|\mathrm{x}(t+1)\| \le &\sum_{\tau=1}^{k}\Big(C_1(1-\gamma)^{2k-\tau+1} + C_2(1-\gamma)^{\kappa+\tau-1}\Big)\Bigg[C_x(1-\gamma)^{t-\tau-k+1}\|\mathrm{x}(0)\| \\ &\qquad + \sum_{l=0}^{t-\tau}C_e(1-\gamma)^l \sup_{0\le s\le t-\tau}\mathrm{e}_k(s) + \frac{C}{\gamma}\Bigg] \\ &+ \sum_{\tau=0}^{k-1}C_3(1-\gamma)^\tau \mathrm{e}_k(t-\tau) + C_4. \end{aligned} \tag{16}$$

Consider first the term multiplying $\|\mathrm{x}(0)\|$ (by (I)):

$$\begin{aligned} \text{(I)} &:= C_x\|\mathrm{x}(0)\|\sum_{\tau=1}^{k}\big(C_1(1-\gamma)^{2k-\tau+1} + C_2(1-\gamma)^{\kappa+\tau-1}\big)(1-\gamma)^{t-\tau-k+1} \\ &= C_x(1-\gamma)^{t-k+1}\|\mathrm{x}(0)\|\sum_{\tau=1}^{k}\big(C_1(1-\gamma)^{2k-\tau+1} + C_2(1-\gamma)^{\kappa+\tau-1}\big)(1-\gamma)^{-\tau}. \end{aligned} \tag{17}$$

In Theorem III.1, κ and k are chosen to satisfy that

$$\sum_{\tau=1}^{k}\left(C_1(1-\gamma)^{2k-\tau+1}+C_2(1-\gamma)^{\kappa+\tau-1}\right)(1-\gamma)^{-\tau}\leq 1. \tag{18}$$

Then from (17),

$$\text{(I)}\leq C_x(1-\gamma)^{t-k+1}\|\mathrm{x}(0)\|. \tag{19}$$

Next, collect all terms involving $\mathrm{e}_k(\cdot)$. From (16) we get two contributions. One is from the double sum by (II):

$$\text{(II)}:=\sum_{\tau=1}^{k}\big(C_1(1-\gamma)^{2k-\tau+1}+C_2(1-\gamma)^{\kappa+\tau-1}\big)\sum_{l=0}^{t-\tau}C_e(1-\gamma)^{l}\sup_{0\leq s\leq t-\tau}\mathrm{e}_k(s). \tag{20}$$

Since $\sup_{0\leq s\leq t-\tau}\mathrm{e}_k(s)\leq\sup_{0\leq s\leq t}\mathrm{e}_k(s)$, $\sum_{l=0}^{t-\tau}(1-\gamma)^l\leq\frac{1}{\gamma}$, we have

$$\text{(II)}\leq\Big(\sum_{\tau=1}^{k}\big(C_1(1-\gamma)^{2k-\tau+1}+C_2(1-\gamma)^{\kappa+\tau-1}\big)\Big)\cdot\frac{C_e}{\gamma}\cdot\sup_{0\leq s\leq t}\mathrm{e}_k(s). \tag{21}$$

Let us denote the direct error term by (III):

$$\text{(III)}:=\sum_{\tau=0}^{k-1}C_3(1-\gamma)^{\tau}\mathrm{e}_k(t-\tau). \tag{22}$$

Using the bound $\mathrm{e}_k(t-\tau)\leq\sup_{0\leq s\leq t}\mathrm{e}_k(s)$ and the geometric series sum $\sum_{\tau=0}^{k-1}(1-\gamma)^{\tau}\leq\frac{1}{\gamma}$, we have

$$\text{(III)}\leq\frac{C_3}{\gamma}\sup_{0\leq s\leq t}\mathrm{e}_k(s). \tag{23}$$

Finally, collect the constant terms (denoted by (IV)):

$$\begin{aligned}\text{(IV)}&:=\left[\sum_{\tau=1}^{k}\left(C_1(1-\gamma)^{2k-\tau+1}+C_2(1-\gamma)^{\kappa+\tau-1}\right)\right]\frac{C}{\gamma}+C_4\\&=\left[\sum_{\tau=1}^{k}\left(C_1(1-\gamma)^{2k-\tau+1}+C_2(1-\gamma)^{\kappa+\tau-1}\right)(1-\gamma)^{-\tau}(1-\gamma)^{\tau}\right]\frac{C}{\gamma}+C_4.\end{aligned} \tag{24}$$

From the condition (18), let $\rho:=\sum_{\tau=1}^{k}\left(C_1(1-\gamma)^{2k-\tau+1}+C_2(1-\gamma)^{\kappa+\tau-1}\right)(1-\gamma)^{-\tau}\leq 1$. Since $(1-\gamma)^{\tau}\leq 1-\gamma$ for $\tau\geq 1$, the coefficient of $\frac{C}{\gamma}$ in (24) is bounded by $\rho(1-\gamma)\leq 1-\gamma$. Thus,

$$\text{(IV)}\leq(1-\gamma)\frac{C}{\gamma}+C_4. \tag{25}$$

Combining the bounds for (I) from (19), (II) from (21), (III) from (23), and (IV) from (25), and substituting them back into (16), we obtain:

$$\begin{aligned}\|\mathrm{x}(t+1)\|\leq{}&C_x(1-\gamma)^{t-k+1}\|\mathrm{x}(0)\|\\&+\left((1-\gamma)\frac{C_e}{\gamma}+\frac{C_3}{\gamma}\right)\sup_{0\leq s\leq t}\mathrm{e}_k(s)\\&+(1-\gamma)\frac{C}{\gamma}+C_4.\end{aligned} \tag{26}$$

To satisfy the induction hypothesis form for $t+1$, we define the constants C_e and C such that the coefficients remain bounded. Specifically, we require:

$$(1-\gamma)\frac{C_e}{\gamma} + \frac{C_3}{\gamma} \le \frac{C_e}{\gamma} \quad \text{and} \quad (1-\gamma)\frac{C}{\gamma} + C_4 \le \frac{C}{\gamma}.$$

Solving these inequalities yields the conditions:

$$C_e \ge \frac{C_3}{\gamma}, \qquad C \ge \frac{C_4}{\gamma}. \tag{27}$$

We define $C_e := C_3/\gamma$ and $C := C_4/\gamma$. Thus, (26) implies that

$$\|\mathrm{x}(t+1)\| \le C_x(1-\gamma)^{(t+1)-k}\|\mathrm{x}(0)\| + \sum_{l=0}^{t+1} C_e(1-\gamma)^l \sup_{0\le s\le t} \mathrm{e}_k(s) + \frac{C}{\gamma}. \tag{28}$$

This completes the induction step.

For $0 \le t \le k$, the recursive inequality (13) is unrolled a finite number of times. Since the time horizon is finite, all states are bounded by the initial state, the accumulated error, and the noise terms.

By iterating (13) for t steps (where $t \le k$), and applying the geometric series bounds (e.g., $\sum_{\tau=0}^{k-1}(1-\gamma)^\tau \le \gamma^{-1}$), we obtain a uniform finite-time bound:

$$\|\mathrm{x}(t)\| \le \bar{C}_x\|\mathrm{x}(0)\| + \bar{C}_e \sup_{0\le s\le k} \mathrm{e}_k(s) + \bar{C}_c, \quad \forall t \in \{0,\dots,k\}, \tag{29}$$

where $\bar{C}_x, \bar{C}_e, \bar{C}_c > 0$ depend only on the system parameters $(\alpha_H, \beta_1, \beta_2, \rho_H, \gamma, L, N, \Xi, k, \kappa)$ and on $\sup_{0\le s\le T-1} \|\widehat{\mathrm{w}}(s)\|$, but are independent of t.

Concretely, one may choose

$$\bar{C}_x := 1 + C_1\frac{(1-\gamma)^{k+1}}{\gamma} + C_2\frac{(1-\gamma)^\kappa}{\gamma}, \tag{30}$$

$$\bar{C}_e := \frac{C_3}{\gamma}, \tag{31}$$

$$\bar{C}_c := \left(1 + \left(C_1\frac{(1-\gamma)^{k+1}}{\gamma^2} + C_2\frac{(1-\gamma)^\kappa}{\gamma^2}\right)\right) C_4, \tag{32}$$

To ensure the global theorem holds, we simply require the final constants C_x, C_e, C to be large enough to cover this initial transient. We impose the lower bounds: $C_x \ge \bar{C}_x, C_e \ge \bar{C}_e, C/\gamma \ge \bar{C}_c$. Under these conditions, for any $t \le k$, the bound (29) implies:

$$\begin{aligned}\|\mathrm{x}(t)\| &\le C_x\|\mathrm{x}(0)\| + C_e \sup_{0\le s\le k} \mathrm{e}_k(s) + \frac{C}{\gamma} \\ &= C_x(1-\gamma)^{\max(0,t-k)}\|\mathrm{x}(0)\| + \sum_{l=0}^{t} C_e(1-\gamma)^l \sup_{0\le s\le k} \mathrm{e}_k(s) + \frac{C}{\gamma}.\end{aligned} \tag{33}$$

Note that for $t \le k$, $(1-\gamma)^{\max(0,t-k)} = 1$. Thus, the form of the inductive hypothesis is satisfied for the initial phase.

Finally, we consider $T-k < t \le T-1$. In this case, the summation in the dynamics is truncated because the future horizon $t+k$ extends beyond the available horizon T. Let $Y := T-t < k$. The dynamics are governed by a truncated version of (13):

$$\|\mathrm{x}(t+1)\| \le \sum_{\tau=1}^{Y} \Theta_\tau\|\mathrm{x}(t-\tau+1)\| + \sum_{\tau=0}^{Y-1} C_3(1-\gamma)^\tau \mathrm{e}_k(t-\tau) + C_4^{(Y)}, \tag{34}$$

where $\Theta_\tau := C_1(1-\gamma)^{2k-\tau+1} + C_2(1-\gamma)^{\kappa+\tau-1}$ denotes the state transition coefficients, and $C_4^{(Y)}$ represents the constant contribution over the truncated horizon Y with

$$C_4^{(L)} := \left(\frac{6\alpha_H\beta_2}{1-\rho_H} S_1(L-1,\gamma) + \frac{2}{1-\rho_H}\right) \sup_{0\le s\le T-1} \|\widehat{\mathrm{w}}(s)\| \\ + LN\beta_1\beta_2\Xi D_k (1-\gamma)^{\kappa+1} S_1(L-1,\gamma) + LN\beta_1\Xi D_k (1-\gamma)^{\kappa}, \tag{35}$$

and

$$S_1(L-1,\gamma) \le \frac{1}{\gamma} \quad \Rightarrow \quad C_4^{(L)} \le C_4.$$

Substituting (15) into (34):

$$\begin{aligned} \|\mathrm{x}(t+1)\| \le & \sum_{\tau=1}^{L} \left(C_1(1-\gamma)^{2k-\tau+1} + C_2(1-\gamma)^{\kappa+\tau-1}\right) \Big[C_x(1-\gamma)^{t-\tau-k+1}\|\mathrm{x}(0)\| \\ & + \sum_{l=0}^{t-\tau} C_e(1-\gamma)^l \sup_{0\le s\le t-\tau} \mathrm{e}_k(s) + \tfrac{C}{\gamma}\Big] \\ & + \sum_{\tau=0}^{L-1} C_3(1-\gamma)^\tau \mathrm{e}_k(t-\tau) + C_4^{(L)}. \end{aligned} \tag{36}$$

Crucially, the partial sums over L terms are strictly bounded by the sums over k terms used in the main induction step. Specifically:

(a) **Gain Condition:** The sum of state coefficients satisfies

$$\sum_{\tau=1}^{L} \Theta_\tau (1-\gamma)^{-\tau} < \sum_{\tau=1}^{k} \Theta_\tau (1-\gamma)^{-\tau} \le 1.$$

(b) **Error Term:** The direct error contribution is

$$\sum_{\tau=0}^{L-1} C_3(1-\gamma)^\tau \le \sum_{\tau=0}^{\infty} C_3(1-\gamma)^\tau = \frac{C_3}{\gamma} \le C_e.$$

(c) **Constant Term:** The truncated constant term satisfies $C_4^{(L)} \le C_4 \le C/\gamma$.

Thus, substituting the inductive hypothesis (15) into (36) and following the exact same procedure as in the case of $t+1 \le T-k$, we confirm that:

$$\|\mathrm{x}(t+1)\| \le C_x(1-\gamma)^{t+1-k}\|\mathrm{x}(0)\| + \sum_{l=0}^{t+1} C_e(1-\gamma)^l \sup_{0\le s\le t} \mathrm{e}_k(s) + \frac{C}{\gamma}. \tag{37}$$

This completes the proof of Theorem III.1. □

2 Proof of Theorem IV.1

The proof contains three steps. First, we show h_t is L_f-smooth by establishing $\|\nabla h_t(\lambda) - \nabla h_t(\lambda')\| \le L_f\|\lambda - \lambda'\|$ via the properties of mixing model $\xi(\cdot)$ and disturbance $\mathrm{w}(t)$, and set $\eta = 1/L_f$. Next, the k-step delayed update yields k number of OGD subsequences with $R_T \le kD/(2\eta) + \eta T/2(2W_\gamma^2\omega L_\xi S_\infty)^2$, implying $1/T\sum_{t\in[T]} h_t(\lambda(t-k)) \le (\varepsilon^\star)^2 + \mathcal{O}(\sqrt{1/T})$. Finally, by ISS Bound in Theorem III.1 and Jensen inequality, we complete the proof.

Step 1 Smoothness of $h_t(\lambda(t))$ Recall Algorithm 3, we use a confidence criterion function $h_t(\lambda) = (\epsilon_t(\lambda))^2$, where $\epsilon_t(\lambda) := \sum_{\tau=t-k}^{t-1}(1-\gamma)^{\tau-t+k}\|\xi(\lambda\circ\widehat{s}(\tau|t-k))-\mathrm{w}(\tau)\|$. We establish the smoothness constant L_f by analyzing the Lipschitz continuity of the gradient $\nabla h_t(\lambda)$. The gradient is given by $\nabla h_t(\lambda) = 2\epsilon_t(\lambda)\nabla\epsilon_t(\lambda)$. Consider two parameters $\lambda, \lambda' \in [0,1]^D$. By the triangle inequality:

$$
\begin{aligned}
&\|\nabla h_t(\lambda) - \nabla h_t(\lambda')\| = 2\|\epsilon_t(\lambda)\nabla\epsilon_t(\lambda) - \epsilon_t(\lambda')\nabla\epsilon_t(\lambda')\| \\
=&2\|\epsilon_t(\lambda)(\nabla\epsilon_t(\lambda) - \nabla\epsilon_t(\lambda')) + (\epsilon_t(\lambda) - \epsilon_t(\lambda'))\nabla\epsilon_t(\lambda')\| \\
\leq&2\underbrace{|\epsilon_t(\lambda)|}_{\text{(a)}}\underbrace{\|\nabla\epsilon_t(\lambda) - \nabla\epsilon_t(\lambda')\|}_{\text{(b)}} + 2\underbrace{|\epsilon_t(\lambda) - \epsilon_t(\lambda')|}_{\text{(c)}}\underbrace{\|\nabla\epsilon_t(\lambda')\|}_{\text{(d)}}.
\end{aligned}
$$

We bound each term based on the properties of ξ (assumed L_ξ-smooth):

(a) $|\epsilon_t(\lambda)| \leq 2\sum_{\tau\geq t}(1-\gamma)^{\tau-t}\omega \leq 2\omega/\gamma$.

(d) $\|\nabla\epsilon_t(\lambda)\| \leq L_\xi S_\infty/\gamma$.

(c) $|\epsilon_t(\lambda) - \epsilon_t(\lambda')| \leq L_\xi S_\infty\|\lambda - \lambda'\|/\gamma$.

(b) $\|\nabla\epsilon_t(\lambda) - \nabla\epsilon_t(\lambda')\| \leq L_\xi S_\infty^2\|\lambda - \lambda'\|/\gamma$.

Substituting these into the inequality:

$$
\begin{aligned}
\|\nabla h_t(\lambda) - \nabla h_t(\lambda')\| &\leq 2(2\omega/\gamma)(L_\xi S_\infty^2/\gamma)\|\lambda - \lambda'\| + 2(L_\xi S_\infty\|\lambda - \lambda'\|/\gamma)(L_\xi S_\infty/\gamma) \\
&= 2S_\infty^2\left(2\omega L_\xi + L_\xi^2\right)\|\lambda - \lambda'\|/\gamma^2.
\end{aligned}
$$

Thus, $h_t(\lambda)$ is L_f-smooth with $L_f = 2S_\infty^2(L_\xi^2 + 2\omega L_\xi)/\gamma^2$. Similarly, we get $\|\nabla h_t(\lambda)\| \leq 2\omega L_\xi S_\infty/\gamma^2$.

Step 2: k-Step Delayed Descent Essentially, the update rule $\lambda(t) = \mathtt{Proj}(\lambda(t-k) - \eta\nabla h_t(\lambda(t-k)))$ decouples the learning process into k subsequences. Let the time step set $\mathcal{T}_r = \{t \in [k, T+k-1] \mid t \bmod k = r\}$ for $r \in \{0, \ldots, k-1\}$. For each subsequence r, the updates form a standard OGD process on the convex functions $\{h_t\}_{t\in\mathcal{T}_r}$. Applying the standard OGD regret bound in [3]:

$$
\sum_{t\in\mathcal{T}_r}(h_t(\lambda(t-k)) - h_t(\lambda^\star)) \leq \frac{D}{2\eta} + \frac{\eta}{2}\sum_{t\in\mathcal{T}_r}\|\nabla h_t(\lambda(t-k))\|^2.
$$

Summing over all k subsequences yields the total regret R_T:

$$
\begin{aligned}
R_T(\lambda) &= \sum_{r=0}^{k-1}\sum_{t\in\mathcal{T}_r}(h_t(\lambda(t-k)) - h_t(\lambda^\star)) \\
&\leq \sum_{r=0}^{k-1}\left(\frac{D}{2\eta} + \frac{\eta|\mathcal{T}_r|\left(2\omega L_\xi S_\infty\right)^2}{2\gamma^2}\right) \\
&= \frac{kD}{2\eta} + \frac{\eta T\left(2\omega L_\xi S_\infty\right)^2}{2\gamma^2}.
\end{aligned}
$$

Dividing by T and substituting η, the average stage cost is bounded by:

$$
\begin{aligned}
\frac{1}{T}\sum_{t=0}^{T-1}h_t(\lambda(t-k)) &\leq h(\lambda^\star) + \frac{kD}{2\eta T} + \frac{\eta\left(2W_\gamma^2\omega L_\xi S_\infty\right)^2}{2} \\
&= (\varepsilon^\star)^2 + \mathcal{O}\left(\sqrt{\frac{1}{T}}\right). \qquad (38)
\end{aligned}
$$

where the equality employs the condition of η.

Step 3: Global ISS Bound Derivation From the proof intermediate (13) underlying Theorem 1, the closed-loop state satisfies, for $t > k$,

$$\|\mathrm{x}(t)\| \leq C_x(1-\gamma)^{\max\{t-k,0\}}\|\mathrm{x}(0)\| + \frac{C}{\gamma} + \frac{C_e}{\gamma}\sum_{\tau=k}^{t-1}\gamma(1-\gamma)^{t-\tau}\sqrt{h_\tau(\lambda(\tau-k))},$$

where $\sqrt{h_\tau(\lambda(\tau-k))}$ identifies the stage prediction error $\varepsilon(\tau)$ in Theorem III.1.

Let $S_t := \sum_{\tau=0}^{t-1}\alpha_{t,\tau}\sqrt{h_\tau(\lambda(\tau-k))}$ with $\alpha_{t,\tau} := \gamma(1-\gamma)^{t-1-\tau}$. By Jensen's inequality,

$$S_t \leq \sqrt{\sum_{\tau=0}^{t-1}\alpha_{t,\tau}h_\tau(\lambda(\tau-k))}. \tag{39}$$

For a horizon $t \in [T]$, Step 2 implies

$$\frac{1}{t}\sum_{\tau=0}^{t-1}h_\tau(\lambda(\tau-k)) = (\varepsilon^\star)^2 + \mathcal{O}\left(t^{-1/2}\right), \tag{40}$$

Then the weighted average satisfies $\sum_{\tau=0}^{t-1}\alpha_{t,\tau}h_\tau(\lambda(\tau-k)) = (\varepsilon^\star)^2 + \mathcal{O}(t^{-1/2})$ for large time step t. Thus, $S_t = \varepsilon^\star + \mathcal{O}(t^{-1/4})$ via $\sqrt{a^2+b^2} \leq a+b$ for $\forall a, b \geq 0$. Substituting yields

$$\|\mathrm{x}(t)\| \leq \mathcal{O}((1-\gamma)^t) + \frac{C + C_e\varepsilon^\star}{\gamma} + \mathcal{O}(t^{-1/4}),$$

as in Theorem IV.1. □

References

[1] Yiheng Lin, Yang Hu, Guanya Shi, Haoyuan Sun, Guannan Qu, and Adam Wierman. Perturbation-based regret analysis of predictive control in linear time varying systems. *Advances in Neural Information Processing Systems*, 34:5174–5185, 2021.

[2] Guannan Qu and Na Li. On the exponential stability of primal-dual gradient dynamics. *IEEE Control Systems Letters*, 3(1):43–48, 2018.

[3] Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2025.

[4] Sungho Shin, Yiheng Lin, Guannan Qu, Adam Wierman, and Mihai Anitescu. Near-optimal distributed linear-quadratic regulator for networked systems. *SIAM Journal on Control and Optimization*, 61(3):1113–1135, 2023.