Appendix

A Decomposition of the State Space

A.1 The $\boldsymbol{E}_{\mathrm{u}} \oplus \boldsymbol{E}_{\mathrm{s}}$-decomposition

It is evident that the following two subspaces of $\mathbb{R}^n$ are invariant with respect to A, namely

$$E_{\mathrm{u}} := \bigoplus_{i \le k} E_i, \; E_{\mathrm{s}} := \bigoplus_{i > k} E_i$$

which we refer to as the *unstable subspace* and the *stable subspace* of A, respectively. Since the eigenspaces E_i sum to the whole $\mathbb{R}^n$ space, one natural decomposition is $\mathbb{R}^n = E_{\mathrm{u}} \oplus E_{\mathrm{s}}$; accordingly, each state can be uniquely decomposed as $x = x_{\mathrm{u}} + x_{\mathrm{s}}$, where $x_{\mathrm{u}} \in E_{\mathrm{u}}$ is called the *unstable component*, and $x_{\mathrm{s}} \in E_{\mathrm{s}}$ is called the *stable component*.

We also decompose A based on the $E_{\mathrm{u}} \oplus E_{\mathrm{s}}$-decomposition. Suppose E_{u} and E_{s} are represented by their *orthonormal* bases $Q_1 \in \mathbb{R}^{n \times k}$ and $Q_2 \in \mathbb{R}^{n \times (n-k)}$, respectively, namely

$$E_{\mathrm{u}} = \mathrm{col}(Q_1), \; E_{\mathrm{s}} = \mathrm{col}(Q_2).$$

Let $Q = [Q_1 \; Q_2]$ (which is invertible as long as A is diagonalizable), and let $R = [R_1^\top \; R_2^\top]^\top := Q^{-1}$. Further, let $\mathit{\Pi}_{\mathrm{u}} := Q_1 R_1$ and $\mathit{\Pi}_{\mathrm{s}} = Q_2 R_2$ be the *oblique* projectors onto E_{u} and E_{s} (along the other subspace), respectively. Since E_{u} and E_{s} are both invariant with regard to A, we know there exists $N_1 \in \mathbb{R}^{k \times k}$, $N_2 \in \mathbb{R}^{(n-k)\times(n-k)}$, such that

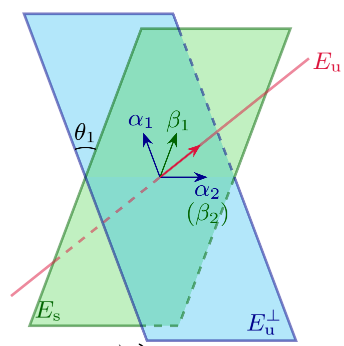
$$AQ = Q \begin{bmatrix} N_1 & \\ & N_2 \end{bmatrix} \;\Leftrightarrow\; N := \begin{bmatrix} N_1 & \\ & N_2 \end{bmatrix} = RAQ.$$

Let $z = [z_1^\top \; z_2^\top]^\top$ be the coordinate representation of x in the basis Q (i.e., $x = Qz$). The system dynamics in z-coordinates can be expressed as

$$\begin{bmatrix} z_{1,t+1} \\ z_{2,t+1} \end{bmatrix} = RAQ \begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} + RBu_t = \begin{bmatrix} N_1 & \\ & N_2 \end{bmatrix} \begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} + \begin{bmatrix} R_1 B \\ R_2 B \end{bmatrix} u_t.$$

The major advantage of this decomposition is that the dynamical matrix in z-coordinate is block diagonal, so it would be simpler to study the behavior of the open-loop system.

A.2 Geometric Interpretation: Principle Angles



Before going any further, we emphasize that Definition 3.1 is well-defined by itself, since singular values are preserved under orthonormal transformations.

It might seem unintuitive to interpret $\sigma_{\min}(P_2^\top Q_2)$ in Definition 3.1 as a measure of "closeness". However, this is closely related to the *principle angles* between subspaces that generalize the standard angle measures in lower dimensional cases. More specifically, we can recursively define the i^{th} principle angle θ_i $(i = 1, \cdots, n-k)$ as

$$\theta_i := \min \left\{ \arccos\left(\frac{\langle x, y \rangle}{\|x\|\|y\|} \right) \;\middle|\; \begin{array}{l} x \in E_{\mathrm{u}}^\perp, \; x \perp \mathrm{span}(x_1, \cdots, x_{i-1}); \\ y \in E_{\mathrm{s}}, \; y \perp \mathrm{span}(y_1, \cdots, y_{i-1}). \end{array} \right\} =: \angle(x_i, y_i), \tag{7}$$

where x_i and y_i $(i = 1, \cdots, n-k)$are referred to as the i^{th} principle vectors accordingly. Meanwhile, let $P_2^\top Q_2 = U \mathit{\Sigma} V^\top$ be the singular value decomposition (SVD), where $\mathit{\Sigma} = \mathrm{diag}(\sigma_1, \cdots, \sigma_{n-k})$ and $\sigma_1 \ge \cdots \ge \sigma_{n-k}$. Then by an equivalent recursive characterization of singular values, we have

$$\sigma_i = \max_{\substack{\|x\|=\|y\|=1 \\ \forall j<i:\; x \perp x_j,\; y \perp y_j}} x^\top P_2^\top Q_2 y =: \bar{x}_i^\top P_2^\top Q_2 \bar{y}_i.$$

Since P_2 and Q_2 are orthonormal, $\bar{x}_i$ and $\bar{y}_i$ can be regarded as coordinate representations of $x_i = P_2 \bar{x}_i$ and $y_i = Q_2 \bar{y}_i$, and it can be easily verified that x_i and y_i defined in this way are exactly the minimizers in (7). Hence we conclude that $\sigma_i = \cos\theta_i$. Therefore, $E_{\mathrm{u}}^\perp$ and E_{s} are ξ-close if and only if the all principle angles between $E_{\mathrm{u}}^\perp$ and E_{s} lie in the interval $[0, \arccos(1-\xi)]$; the above argument also shows that we can find orthonormal bases for $E_{\mathrm{u}}^\perp$ and E_{s} so that corresponding vectors form exactly the principle angles.

A.3 Characterization of ξ-close Subspaces

It is naturally expected that the geometric interpretation should inspire more relationships among $P_1 = Q_1, P_2, Q_2, R_1, R_2$ and N_2. We would like to emphasize that P_1, P_2 and Q_1 are not confined to bases consisting of eigenvectors (since they are even not necessarily orthonormal). Meanwhile, since they are only used in the stability guarantee proof, we are granted the freedom to select any orthonormal bases. For simplicity, we will stick to the convention that $P_1 = Q_1$ (and thus $M_1 = N_1$). Further, in Lemma A.1, such freedom is utilized to establish fundamental relationships between the bases in the above two decompositions. The results are concluded as follows.

Lemma A.1. *Suppose $E_{\mathrm{u}}^{\perp}$ and E_{s} are ξ-close. Then we shall select P_2 and Q_2 such that*

(1) $\sigma_{\min}(P_2^\top Q_2) \geq 1 - \xi$, $\|P_1^\top Q_2\| \leq \sqrt{2\xi}$, $\|P_2 - Q_2\| \leq \sqrt{2\xi}$.

(2) $\|R_2\| \leq \frac{1}{1-\xi}$, $\|N_2\| \leq \frac{1}{1-\xi}\|A\|$.

(3) $\|P_1^\top - R_1\| \leq \frac{\sqrt{2\xi}}{1-\xi}$, $\|R_1\| \leq \frac{\sqrt{2\xi}}{1-\xi} + 1$.

(4) $\|\Delta\| \leq \frac{2-\xi}{1-\xi}\sqrt{2\xi}\|A\|$.

Proof. (1) Following the above interpretation, take arbitrary orthonormal bases $\bar{P}_2$ and $\bar{Q}_2$ of $E_{\mathrm{u}}^{\perp}$ and E_{s}, respectively, and let $\bar{P}_2^\top \bar{Q}_2 = U\Sigma V^\top$ be the SVD, which translates to

$$(\bar{P}_2 U)^\top (\bar{Q}_2 V) = \Sigma =: \mathrm{diag}(\sigma_1, \cdots, \sigma_{n-k}).$$

Since U and V are orthonormal matrices, the columns of $\bar{P}_2 U$ and $\bar{Q}_2 V$ also form orthonormal bases of $E_{\mathrm{u}}^{\perp}$ and E_{s}, respectively. Then ξ-closeness basically says that there exist a basis $\{\alpha_1, \cdots, \alpha_{n-k}\}$ for $E_{\mathrm{u}}^{\perp}$, and a basis $\{\beta_1, \cdots, \beta_{n-k}\}$ for E_{s} (both are assumed to be orthonormal), such that

$$\langle \alpha_i, \beta_j \rangle = \delta_{ij}\sigma_i = \begin{cases} \sigma_i \geq 1 - \xi & \text{for any } i = j \\ 0 & \text{for any } i \neq j \end{cases},$$

and we also have $\Pi_2 \beta_i = \sigma_i \alpha_i$ and $\Pi_1 \alpha_i = \sigma_i \beta_i$ (recall that Π_1, Π_2 are orthogonal projectors onto subspaces $E_{\mathrm{u}}, E_{\mathrm{u}}^{\perp}$, respectively). Therefore, without loss of generality, we shall always select $P_2 = [\alpha_1 \ \cdots \ \alpha_{n-k}]$ and $Q_2 = [\beta_1 \ \cdots \ \beta_{n-k}]$, such that $P_2^\top Q_2 = \mathrm{diag}(\sigma_1, \cdots, \sigma_{n-k})$, and

$$\sigma_{\min}(P_2^\top Q_2) = \min_i |\sigma_i| \geq 1 - \xi.$$

Equivalently speaking, for any $\beta = Q_2 \eta \in E_{\mathrm{s}}$, we have (note that $\|\eta\| = \|\beta\|$)

$$\|P_2^\top \beta\| = \|P_2^\top Q_2 \eta\| \geq \sigma_{\min}(P_2^\top Q_2)\|\eta\| \geq (1 - \xi)\|\beta\|,$$

and consequently,

$$\|P_1^\top Q_2 \eta\| = \|P_1^\top \beta\| = \sqrt{\|\beta\|^2 - \|P_2^\top \beta\|^2} \leq \sqrt{2\xi}\|\beta\| = \sqrt{2\xi}\|\eta\|,$$

which further shows $\|P_1^\top Q_2\| \leq \sqrt{2\xi}$. To bound $\|P_2 - Q_2\|$, by definition we have

$$\begin{aligned}
\|P_2 - Q_2\| &= \max_{\|\eta\|=1} \|(P_2 - Q_2)\eta\| = \max_{\|\eta\|=1} \left\| \sum_i \eta_i (\alpha_i - \beta_i) \right\| \\
&= \max_{\|\eta\|=1} \sqrt{\sum_{i,j} \eta_i \eta_j (\alpha_i - \beta_i)^\top (\alpha_j - \beta_j)} \\
&= \max_{\|\eta\|=1} \sqrt{\sum_i 2(1 - \mu_i)\eta_i^2} \\
&\leq \max_{\|\eta\|=1} \sqrt{2\xi \sum_i \eta_i^2} = \sqrt{2\xi}.
\end{aligned}$$

Here $\eta = [\eta_1, \cdots, \eta_{n-k}]$ is an arbitrary vector in $\mathbb{R}^{n-k}$.

(2) By definition, $I = QR = Q_1 R_1 + Q_2 R_2$. Also recall that $P_1 = Q_1$, so we have $P_1^\top Q_1 = I$ and $P_2^\top Q_1 = \mathbf{0}$. Then by left-multiplying $P_2^\top$ to the equality, we have

$$P_2^\top = P_2^\top Q_1 R_1 + P_2^\top Q_2 R_2 = P_2^\top Q_2 R_2,$$

which further shows

$$\|R_2\| = \|(P_2^\top Q_2)^{-1} P_2^\top\| \leq \|(P_2^\top Q_2)^{-1}\| = \frac{1}{\sigma_{\min}(P_2^\top Q_2)} \leq \frac{1}{1-\xi}.$$

Therefore, since $N_2 = R_2 A Q_2$, we have

$$\|N_2\| = \|R_2 A Q_2\| \leq \|R_2\|\|A\|\|Q_2\| \leq \frac{1}{1-\xi}\|A\|.$$

(3) Similarly, by left-multiplying $P_1^\top$ to the equality, we have

$$P_1^\top = P_1^\top Q_1 R_1 + P_1^\top Q_2 R_2 = R_1 + P_1^\top Q_2 R_2,$$

which further shows

$$\|P_1^\top - R_1\| = \|P_1^\top Q_2 R_2\| \leq \|P_1^\top Q_2\|\|R_2\| \leq \frac{\sqrt{2\xi}}{1-\xi},$$

and therefore $\|R_1\| \leq \|P_1^\top - R_1\| + \|P_1^\top\| = 1 + \frac{\sqrt{2\xi}}{1-\xi}$.

(4) A combination of the above results gives

$$\begin{aligned}
\|\varDelta\| = \|P_1^\top A P_2\| &= \|P_1^\top A P_2 - R_1 A Q_2\| \\
&\leq \|P_1^\top A(P_2 - Q_2)\| + \|(P_1^\top - R_1) A Q_2\| \\
&\leq \|P_1^\top\|\|A\|\|P_2 - Q_2\| + \|P_1^\top - R_1\|\|A\|\|Q_2\| \\
&\leq \|A\|\sqrt{2\xi} + \frac{\sqrt{2\xi}}{1-\xi}\|A\| = \frac{2-\xi}{1-\xi}\sqrt{2\xi}\|A\|.
\end{aligned}$$

This completes the proof. □

B Solution to the Least Squares Problem in Stage 2

Lemma B.1 gives the explicit form for the solution to the least squares problem (see Algorithm 1).

Lemma B.1. *Given $D := [x_{t_0+1} \ \cdots \ x_{t_0+k}]$ and $\hat{P}_1 \hat{P}_1^\top = \hat{\varPi}_1 = D(D^\top D)^{-1} D^\top$, the solution*

$$\hat{M}_1 = \arg\min_{M_1} \sum_{t=t_0+1}^{t_0+k} \|\hat{P}_1^\top x_{t+1} - M_1 \hat{P}_1^\top x_t\|^2$$

is uniquely given by $\hat{M}_1 = \hat{P}_1^\top A \hat{P}_1$.

Proof. Here we assume by default that the summation over t sums from t_0+1 to t_0+k. Since M_1 is a stationary point of $\mathcal{L}$, for any $\varDelta$ in the neighbourhood of O, we have

$$\begin{aligned}
0 \leq \mathcal{L}(M_1 + \varDelta) - \mathcal{L}(M_1) &= \sum_t \|\hat{y}_{1,t+1} - M_1 \hat{y}_{1,t} - \varDelta \hat{y}_{1,t}\|^2 - \sum_t \|\hat{y}_{1,t+1} - M_1 \hat{y}_{1,t}\|^2 \\
&= \sum_t \langle \varDelta \hat{y}_{1,t}, \hat{y}_{1,t+1} - M_1 \hat{y}_{1,t} \rangle + O(\|\varDelta\|^2) \\
&= \sum_t \operatorname{tr}\left(\hat{y}_{1,t}^\top \varDelta^\top (\hat{y}_{1,t+1} - A \hat{y}_{1,t})\right) + O(\|\varDelta\|^2) \\
&= \sum_t \operatorname{tr}\left(\varDelta^\top (\hat{y}_{1,t+1} - M_1 \hat{y}_{1,t}) \hat{y}_{1,t}^\top\right) + O(\|\varDelta\|^2) \\
&= \operatorname{tr}\left(\varDelta^\top \sum_t (\hat{y}_{1,t+1} - M_1 \hat{y}_{1,t}) \hat{y}_{1,t}^\top\right) + O(\|\varDelta\|^2).
\end{aligned}$$

Since it always holds for any $\varDelta$, we must have

$$\sum_t (\hat{y}_{1,t+1} - M_1 \hat{y}_{1,t}) \hat{y}_{1,t}^\top \ \Leftrightarrow \ M_1 \sum_t \hat{y}_{1,t} \hat{y}_{1,t}^\top = \sum_t \hat{y}_{1,t+1} \hat{y}_{1,t}^\top.$$

Plugging in $\hat{y}_{1,t} = \hat{P}_1^\top x_t$ and $\hat{y}_{1,t+1} = \hat{P}_1^\top A x_t$, we further have

$$M_1 \hat{P}_1^\top X \hat{P}_1 = M_1 \sum_t \hat{P}_1^\top x_t x_t^\top \hat{P}_1 = \sum_t \hat{P}_1^\top A x_t x_t^\top \hat{P}_1 = \hat{P}_1^\top A X \hat{P}_1,$$

where $X := \sum_t x_t x_t^\top = DD^\top$. Since the columns of $\hat{P}_1$ form an orthonormal basis of $\hat{E}_\mathrm{u}$, for any $x \in \hat{E}_\mathrm{u}$, $\hat{P}_1^\top x$ is the coordinate of x under that basis. The columns of D are linearly independent, so the columns of $\hat{P}_1^\top D$ are also linearly independent, which further yields

$$\operatorname{rank}(\hat{P}_1^\top X \hat{P}_1) = \operatorname{rank}\left((\hat{P}_1^\top D)(\hat{P}_1^\top D)^\top\right) = \operatorname{rank}(\hat{P}_1^\top D) = k.$$

Therefore, $\hat{P}_1^\top X \hat{P}_1$ is invertible, and M_1 is explicitly given by

$$M_1 = (\hat{P}_1^\top A X \hat{P}_1)(\hat{P}_1^\top X \hat{P}_1)^{-1}.$$

Note that $\hat{\Pi}_1 = \hat{P}_1 \hat{P}_1^\top$ is the projector onto subspace $\mathrm{col}(D)$, we must have

$$\hat{P}_1 \hat{P}_1^\top X = (\hat{\Pi}_1 D) D^\top = DD^\top = X,$$

which yields

$$M_1 = (\hat{P}_1^\top A (\hat{P}_1 \hat{P}_1^\top X) \hat{P}_1)(\hat{P}_1^\top X \hat{P}_1)^{-1} = (\hat{P}_1^\top A \hat{P}_1)(\hat{P}_1^\top X \hat{P}_1)(\hat{P}_1^\top X \hat{P}_1)^{-1} = \hat{P}_1^\top A \hat{P}_1.$$

This completes the proof of Lemma B.1. □

It might help understanding to note that, when $\hat{P}_1 = P_1$, for any $x_t, x_{t+1} \in E_\mathrm{u}$ we have

$$P_1^\top A x_t = y_{t+1} = M_1 y_t = M_1 P_1^\top x_t,$$

which requires $P_1^\top A = M_1 P_1^\top$, or equivalently $M_1 = P_1^\top A P_1$ (recall $P_1^\top P_1 = I$).

C Transformation of B with Arbitrary Columns

In the remaining sections of this paper, we have always regarded B as an n-by-k matrix (i.e., $m = k$). In this section, we will show that other cases can be handled in a similar way under proper transformations. This is trivial for the case where $m > k$, since we can simply select k linearly independent columns from B, and pad 0's in u_t for all unselected entries.

For the case where $m < k$, let $d = \lceil k/m \rceil$. Intuitively, we can "pack" every d consecutive steps to obtain a system with sufficient number of control inputs. More specifically, let

$$\tilde{x}_t = \begin{bmatrix} x_{td} \\ x_{td+1} \\ \vdots \\ x_{(t+1)d-1} \end{bmatrix}, \ \tilde{u}_t = \begin{bmatrix} u_{td-1} \\ u_{td} \\ \vdots \\ u_{(t+1)d-2} \end{bmatrix},$$

$$\tilde{A} = \begin{bmatrix} \mathbf{0} & & A \\ & \ddots & \vdots \\ & \mathbf{0} & A^{d-1} \\ & & A^d \end{bmatrix}, \ \tilde{B} = \begin{bmatrix} B & & & \\ AB & B & & \\ \vdots & \vdots & \ddots & \\ A^{d-1}B & A^{d-2}B & \cdots & B \end{bmatrix},$$

and consider the transformed system with dynamics

$$\tilde{x}_{t+1} = \tilde{A}\tilde{x}_t + \tilde{B}\tilde{u}_t.$$

The instability index of $\tilde{A}$ is still k, with $|\tilde{\lambda}_i| = |\lambda_i|^d$ ($i = 1, \cdots, n$). Norms of $\tilde{A}$ and $\tilde{B}$ satisfy

$$\|\tilde{A}\| \le \sqrt{\sum_{i=1}^{d} \|A^i\|^2} = \|A^d\| O(d), \quad \|\tilde{B}\| \le \|B\| \sqrt{\sum_{i=1}^{d} (d-i)\|A^i\|^2} = \|A^d\| \|B\| O(d).$$

Since $d \le k \ll n$, the above transformation only multiplies the bounds by a small constant.

D Proof of Lemma 5.3

Lemma 5.3 is actually a direct corollary of the following lemma, for which we first need to define $\mathrm{gap}_i(A)$, the *(bipartite) spectral gap around λ_i with respect to A*, namely

$$\mathrm{gap}_i(A) := \begin{cases} \min_{\lambda_j \in \lambda(A_2)} |\lambda_i - \lambda_j| & \lambda_i \in \lambda(A_1) \\ \min_{\lambda_j \in \lambda(A_1)} |\lambda_i - \lambda_j| & \lambda_i \in \lambda(A_2) \end{cases},$$

where $\lambda(A)$ denotes the spectrum of A.

Lemma D.1. *For 2-by-2 block matrices A and E in the form*

$$A = \begin{bmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & A_2 \end{bmatrix},\ E = \begin{bmatrix} \mathbf{0} & E_{12} \\ E_{21} & \mathbf{0} \end{bmatrix},$$

we have

$$|\lambda_i(A+E) - \lambda_i(A)| \le \frac{\kappa_{\mathrm{d}}(A)\kappa_{\mathrm{d}}(A+E)}{\mathrm{gap}_i(A)} \|E_{12}\| \|E_{21}\|.$$

Here $\kappa_{\mathrm{d}}(A)$ is the condition number of the matrix consisting of A's eigenvectors as columns.

Proof. The proof of the lemma can be found in existing literature like [53]. □

Proof of Lemma 5.3. Lemma D.1 basically guarantees that every eigenvalue of $A+E$ is within a distance of $O(\|E_{12}\|\|E_{21}\|)$ from some eigenvalue of A. Hence, by defining $\chi(A+E)$ as the maximum coefficient, namely

$$\chi(A+E) := \frac{\kappa_{\mathrm{d}}(A)\kappa_{\mathrm{d}}(A+E)}{\min_i\{\mathrm{gap}_i(A)\}},$$

we shall guarantee $|\rho(A+E) - \rho(A)| \le \chi(A+E)\|E_{12}\|\|E_{21}\|$. □

E Proof of Theorem 5.1 and its Corollary

The main idea of this proof is to diagonalize A and write the open-loop system dynamics using the basis formed by the eigenvectors of A. Then, we provide an explicit expression for $\hat{\Pi}_1$ and Π_1, based on which we can bound the error. To further derive a bound for $\|\hat{P}_1 - P_1\|$, one only needs to notice that norms are preserved under orthonormal coordinate transformations, so it only suffices to find a specific pair of bases of $E_{\mathrm{u}}^{\perp}$ and E_{s} that are close to each other — and the pair of bases formed by principle vectors (see Appendix A) is exactly what we want. This leads to Corollary 5.2 that is repeatedly used in subsequent proofs.

Without loss of generality, we shall write all matrices in the basis formed by unit eigenvectors $\{w_1, \cdots, w_n\}$ of A. Otherwise, let $W = [w_1\ \cdots\ w_n]$, and perform change-of-coordinate by setting $\tilde{D} := W^{-1}DW$, $\tilde{\Pi}_1 := W^{-1}\Pi_1 W$, which further gives

$$\tilde{\hat{\Pi}}_1 = \tilde{D}(\tilde{D}^\top \tilde{D})^{-1}\tilde{D}^\top = (W^{-1}DW)(W^{-1}D^\top DW)^{-1}(W^{-1}D^\top W) = W^{-1}\hat{\Pi}_1 W.$$

Note that $\|W^{-1}\hat{\Pi}_1 W - W^{-1}\Pi_1 W\| \le \|W\|\|W^{-1}\|\|\hat{\Pi}_1 - \Pi_1\|$, where the upper bound is only magnified by a constant factor of $\kappa_{\mathrm{d}}(A) = \|W\|\|W^{-1}\|$ that is completely determined by A. Therefore, it is largely equivalent to consider $(\tilde{D}, \tilde{\Pi}_1, \tilde{\hat{\Pi}}_1)$ instead of $(D, \Pi_1, \hat{\Pi}_1)$.

Note that the matrix $D = [x_{t_0+1}\ \cdots\ x_{t_0+k}]$ can be written as

$$D = \begin{bmatrix} d_1 & \lambda_1 d_1 & \cdots & \lambda_1^{k-1} d_1 \\ d_2 & \lambda_2 d_2 & \cdots & \lambda_2^{k-1} d_2 \\ \vdots & \vdots & \ddots & \vdots \\ d_n & \lambda_n d_n & \cdots & \lambda_n^{k-1} d_n \end{bmatrix},$$

where $x_{t_0+1} =: [d_1, \cdots, d_n]^\top$. We first present a lemma characterizing some well-known properties of Vandermonde matrices that we need in the proof.

Lemma E.1. *Given a Vandermonde matrix in variables $x_1, \cdots, x_n$ of order n*

$$V := V_n(x_1, \cdots, x_n) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix},$$

its determinant is given by

$$\det(V) = \sum_{\pi} (-1)^{\mathrm{sgn}(\pi)} x^0_{\pi(i_1)} x^1_{\pi(i_2)} \cdots x^{n-1}_{\pi(i_n)} = \prod_{j<\ell} (x_\ell - x_j), \tag{8}$$

and its (u, v)-cofactor is given by

$$\mathrm{cof}_{u,v}(V) = \begin{vmatrix} 1 & \cdots & 1 & 1 & \cdots & 1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_1^{u-2} & \cdots & x_{v-1}^{u-2} & x_{v+1}^{u-2} & \cdots & x_n^{u-2} \\ x_1^{u} & \cdots & x_{v-1}^{u} & x_{v+1}^{u} & \cdots & x_n^{u} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & \cdots & x_{v-1}^{n-1} & x_{v+1}^{n-1} & \cdots & x_n^{n-1} \end{vmatrix} = \sigma_{u,v} \prod_{j<\ell\neq v} (x_\ell - x_j). \tag{9}$$

Here coefficients $\sigma_{u,v}$ are given by $\sigma_{u,v} := s_{n-u}(x_1, \cdots, x_{v-1}, x_{v+1}, \cdots, x_n)$, where function s_m is defined by $s_m(y_1, \cdots, y_n) := \sum_{i_1 < \cdots < i_m} y_{i_1} \cdots y_{i_m}$.

Proof of Lemma E.1. The proof of (8) can be found in any standard linear algebra textbook, and that of (9) can be found in [54]. □

It is evident that the entries in D display a similar pattern as those of a Vandermonde matrix. Based on this observation, we shall further derive the explicit form of $\hat{\Pi}_1$ as in the next lemma.

Lemma E.2. *The projector $\hat{\Pi}_1 = D(D^\top D)^{-1} D^\top$ has explicit form*

$$(\hat{\Pi}_1)_{uv} = \frac{\displaystyle\sum_{\substack{i_2 < \cdots < i_k \\ \forall j : i_j \neq u, v}} \alpha_{u, i_2, \cdots, i_k} \alpha_{v, i_2, \cdots, i_k}}{\displaystyle\sum_{i_1 < \cdots < i_k} \alpha^2_{i_1, \cdots, i_k}},$$

where the summand $\alpha_{i_1, \cdots, i_k}$ (with ordered subscript) is defined as

$$\alpha_{i_1, \cdots, i_k} := \prod_j d_{i_j} \prod_{j<\ell} (\lambda_{i_\ell} - \lambda_{i_j}).$$

Proof of Lemma E.2. We start by deriving the explicit form of $(D^\top D)^{-1}$. Note that the determinant (which is also the denominator in the lemma) is given by

$$\begin{aligned} \det(D^\top D) &= \sum_{i_1, \cdots, i_k} \begin{vmatrix} \lambda^0_{i_1} d^2_{i_1} & \lambda^1_{i_2} d^2_{i_2} & \cdots & \lambda^{k-1}_{i_k} d^2_{i_k} \\ \lambda^1_{i_1} d^2_{i_1} & \lambda^2_{i_2} d^2_{i_2} & \cdots & \lambda^{k}_{i_k} d^2_{i_k} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda^{k-1}_{i_1} d^2_{i_1} & \lambda^k_{i_2} d^2_{i_2} & \cdots & \lambda^{2k-2}_{i_k} d^2_{i_k} \end{vmatrix} \\ &= \sum_{i_1, \cdots, i_k} d^2_{i_1} \cdots d^2_{i_k} \lambda^0_{i_1} \lambda^1_{i_2} \cdots \lambda^{k-1}_{i_k} \prod_{j<\ell} (\lambda_{i_\ell} - \lambda_{i_j}) \\ &= \sum_{i_1 < \cdots < i_k} d^2_{i_1} \cdots d^2_{i_k} \prod_{j<\ell} (\lambda_{i_\ell} - \lambda_{i_j}) \sum_{\pi} (-1)^{\mathrm{sgn}(\pi)} \lambda^0_{\pi(j_1)} \lambda^1_{\pi(j_2)} \cdots \lambda^{k-1}_{\pi(j_k)} \\ &= \sum_{i_1 < \cdots < i_k} d^2_{i_1} \cdots d^2_{i_k} \prod_{j<\ell} (\lambda_{i_\ell} - \lambda_{i_j})^2 \\ &= \sum_{i_1 < \cdots < i_k} \alpha^2_{i_1, \cdots, i_k}, \end{aligned}$$

and the (u, v)-cofactor $\mathrm{cof}_{u,v}(D^\top D)$ is given by

$$
\begin{aligned}
\mathrm{cof}_{u,v}(D^\top D) &= (-1)^{u+v} \sum_{i_1,\cdots,i_{k-1}} \begin{vmatrix}
\lambda_{i_1}^{0} d_{i_1}^2 & \cdots & \lambda_{i_{v-1}}^{v-2} d_{i_{v-1}}^2 & \lambda_{i_v}^{v} d_{i_v}^2 & \cdots & \lambda_{i_{k-1}}^{k-1} d_{i_{k-1}}^2 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\lambda_{i_1}^{u-2} d_{i_1}^2 & \cdots & \lambda_{i_{v-1}}^{u+v-4} d_{i_{v-1}}^2 & \lambda_{i_v}^{u+v-2} d_{i_v}^2 & \cdots & \lambda_{i_{k-1}}^{u+k-3} d_{i_{k-1}}^2 \\
\lambda_{i_1}^{u} d_{i_1}^2 & \cdots & \lambda_{i_{v-1}}^{u+v-2} d_{i_{v-1}}^2 & \lambda_{i_v}^{u+v} d_{i_v}^2 & \cdots & \lambda_{i_{k-1}}^{u+k-1} d_{i_{k-1}}^2 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\lambda_{i_1}^{k-1} d_{i_1}^2 & \cdots & \lambda_{i_{u+v-2}}^{k+v-3} d_{i_{v-1}}^2 & \lambda_{i_v}^{k+v-1} d_{i_v}^2 & \cdots & \lambda_{i_{k-1}}^{2k-2} d_{i_{k-1}}^2
\end{vmatrix} \\
&= (-1)^{u+v} \sum_{i_1,\cdots,i_{k-1}} d_{i_1}^2 \cdots d_{i_{k-1}}^2 \lambda_{i_1}^0 \cdots \lambda_{i_{v-1}}^{v-2} \lambda_{i_v}^{v} \cdots \lambda_{i_{k-1}}^{k-1} s_{k-u} \prod_{j<\ell} (\lambda_{i_\ell} - \lambda_{i_j}) \\
&= (-1)^{u+v} \sum_{i_1<\cdots<i_{k-1}} s_{k-u} \cdot d_{i_1}^2 \cdots d_{i_{k-1}}^2 \prod_{j<\ell} (\lambda_{i_\ell} - \lambda_{i_j}) \cdot \\
&\qquad\qquad \sum_{\pi} (-1)^{\mathrm{sgn}(\pi)} \lambda_{\pi(i_1)}^0 \cdots \lambda_{\pi(i_{v-1})}^{v-2} \lambda_{\pi(i_v)}^{v} \cdots \lambda_{\pi(i_{k-1})}^{k-1} \\
&= (-1)^{u+v} \sum_{i_1<\cdots<i_{k-1}} s_{k-u} s_{k-v} \cdot d_{i_1}^2 \cdots d_{i_{k-1}}^2 \prod_{j<\ell} (\lambda_{i_\ell} - \lambda_{i_j})^2,
\end{aligned}
$$

where $s_{k-u}(\lambda_{i_1}, \cdots, \lambda_{i_{k-1}})$ is abbreviated to s_{k-u}.

Note that symmetry of $D^\top D$ guarantees $\mathrm{cof}_{v,u}(D^\top D) = \mathrm{cof}_{u,v}(D^\top D)$, so we have

$$
(D^\top D)^{-1}_{u,v} = \frac{\mathrm{cof}_{v,u}(D^\top D)}{\det(D^\top D)} = \frac{\mathrm{cof}_{u,v}(D^\top D)}{\det(D^\top D)}.
$$

And eventually we shall derive that

$$
\begin{aligned}
\hat{P}_{u,v} &= \sum_{p,q} D_{u,p} (D^\top D)^{-1}_{p,q} D^\top_{q,v} \\
&= \frac{1}{\det(D^\top D)} \sum_{p,q} D_{u,p} D_{v,q} \,\mathrm{cof}_{u,v}(D^\top D) \\
&= \frac{1}{\det(D^\top D)} \sum_{p,q} \lambda_u^{p-1} d_u \lambda_v^{q-1} d_v \cdot (-1)^{p+q} \sum_{i_1<\cdots<i_{k-1}} s_{k-p} s_{k-q} \cdot d_{i_1}^2 \cdots d_{i_{k-1}}^2 \prod_{j<\ell} (\lambda_{i_\ell} - \lambda_{i_j})^2 \\
&= \frac{1}{\det(D^\top D)} \sum_{i_1<\cdots<i_{k-1}} d_u d_v d_{i_1}^2 \cdots d_{i_{k-1}}^2 \prod_{j<\ell} (\lambda_{i_\ell} - \lambda_{i_j})^2 \sum_{p=1}^{k} (-1)^p \lambda_u^{p-1} s_{k-p} \sum_{q=1}^{k} (-1)^q \lambda_v^{q-1} s_{k-q} \\
&= \frac{1}{\det(D^\top D)} \sum_{i_1<\cdots<i_{k-1}} d_u d_{i_1} \cdots d_{i_{k-1}} \prod_{j<\ell} (\lambda_{i_\ell} - \lambda_{i_j}) \prod_{\ell} (\lambda_{i_\ell} - \lambda_u) \cdot \\
&\qquad\qquad d_v d_{i_1} \cdots d_{i_{k-1}} \prod_{j<\ell} (\lambda_{i_\ell} - \lambda_{i_j}) \prod_{\ell} (\lambda_{i_\ell} - \lambda_v) \\
&= \frac{1}{\det(D^\top D)} \sum_{\substack{i_2<\cdots<i_k \\ \forall j: i_j \neq u,v}} \alpha_{u,i_2,\cdots,i_k} \alpha_{v,i_2,\cdots,i_k},
\end{aligned}
$$

which is in exact the same form as stated in the lemma. □

Now we shall go back to the proof of the main result of this section.

Proof of Theorem 5.1. Recall that $d_i = \lambda_i^{t_0+1} x_{0,i}$. For the clarity of notations, let

$$
\theta_{i_1,i_2,\cdots,i_k} := \frac{\alpha_{i_1,i_2,\cdots,i_k}}{\alpha_{1,2,\cdots,k}},
$$

and it is evident that $|\theta_{i_1,i_2,\cdots,i_k}| = 1$ only if $(i_1, i_2, \cdots, i_k)$ is a permutation of $(1, 2, \cdots, k)$. For any other $(i_1, i_2, \cdots, i_k)$, by the definition in Lemma E.2 we have

$$|\theta_{i_1,i_2,\cdots,i_k}| \le c_{i_1,i_2,\cdots,i_k} \cdot r^{\sum_j \mathbf{1}_{i_j>k} t_0} \le c \cdot r^{t_0},$$

where $r = \frac{|\lambda_{k+1}|}{|\lambda_k|}$ and $c := \max_{i_1,\cdots,i_k} \{c_{i_1,i_2,\cdots,i_k}\}$. Therefore, since there are $\binom{n}{k}$ different k-tuples $(i_1, \cdots, i_k)$ such that $i_1 < \cdots < i_k$, we have

$$\sum_{i_1<\cdots<i_k} \theta^2_{i_1,\cdots,i_k} - \theta^2_{1,\cdots,k} < c\binom{n}{k} r^{2t_0}.$$

Now we can bound the entries in $\hat{\Pi}_1$. For any $\varepsilon > 0$, we shall select t_0 such that $c\binom{n}{k}r^{2t_0} < \frac{\varepsilon}{n^2}$, where the denominator is always bounded by

$$1 \le \sum_{i_1<\cdots<i_k} \theta^2_{i_1,\cdots,i_k} \le 1 + \frac{\varepsilon}{n^2}.$$

For the nominator, note that for each δ there are fewer entries with exponent δ in the nominator than in the denominator, so we can bound the denominator as

$$\left| \sum_{\substack{i_2<\cdots<i_k \\ \forall j: i_j \ne u,v}} \theta_{u,i_2,\cdots,i_k} \theta_{v,i_2,\cdots,i_k} \right| \le \begin{cases} c\binom{n}{k}r^{2t_0} + 1 & u = v \le k \\ c\binom{n}{k}r^{2t_0} & \text{otherwise} \end{cases}.$$

Therefore, when $u = v \le k$, we have $\sum_{\substack{i_2<\cdots<i_k \\ \forall j: i_j \ne u}} \theta^2_{u,i_2,\cdots,i_k} \ge 1$, which shows

$$\left.\begin{aligned} (\hat{\Pi}_1)_{uv} \ge \left(1 + \frac{\varepsilon}{n^2}\right)^{-1} \ge 1 - \frac{\varepsilon}{n^2} \\ (\hat{\Pi}_1)_{uv} \le 1 + \frac{\varepsilon}{n^2} \end{aligned}\right\} \;\Rightarrow\; \left|(\hat{\Pi}_1)_{uv} - (\Pi_1)_{uv}\right| \le \frac{\varepsilon}{n^2};$$

for all other cases, the nominator cannot sum over a permutation of $(1, \cdots, k)$, which gives

$$\left|(\hat{\Pi}_1)_{uv} - (\Pi_1)_{uv}\right| = \left|(\hat{\Pi}_1)_{uv}\right| \le \frac{\varepsilon}{n^2}.$$

Therefore, the overall estimation error is bounded by

$$\|\hat{\Pi}_1 - \Pi_1\| \le \sum_{u,v} \left|(\hat{\Pi}_1)_{uv} - (\Pi_1)_{uv}\right| \le \varepsilon.$$

Recall that the bound is subject to a change-of-basis transformation, and in the general scenario where the eigenvectors of A are not mutually orthogonal, the original prediction error bound should be multiplied by $\kappa_{\mathrm{d}}(A)$. Therefore, to achieve error threshold ε for predictions on Π_i, it is required that $c\binom{n}{k}r^{2t_0} < \frac{\varepsilon}{\kappa_{\mathrm{d}}(A)n^2}$, or equivalently, by *Stirling's Formula*,

$$t_0 > \frac{\log \kappa_{\mathrm{d}}(A) + \log \frac{cn^2}{\varepsilon} + \log \binom{n}{k}}{2 \log \frac{1}{r}} = O\left(\frac{k \log n - \log \varepsilon + \log \kappa_{\mathrm{d}}(A)}{2 \log \frac{|\lambda_k|}{|\lambda_{k+1}|}} \right). \tag{10}$$

This completes the proof. □

Proof of Corollary 5.2. We first construct a specific pair of orthonormal bases $(P_1^*, \hat{P}_1^*)$ that satisfy the corollary. To start with, take an arbitrary initial pair of orthonormal basis $(P_1^\circ, \hat{P}_1^\circ)$, and consider the SVD $(P_1^\circ)^\top \hat{P}_1^\circ = U\Sigma V^\top$, which is equivalent to $(P_1^\circ U)^\top (\hat{P}_1^\circ V) = \Sigma$. Note that the columns of $P_1^\circ U = [w_1 \; \cdots w_k]$ and $\hat{P}_1^\circ V = [\hat{w}_1 \; \cdots \hat{w}_k]$ form orthonormal bases of $\mathrm{col}(\Pi_1)$ and $\mathrm{col}(\hat{\Pi}_1)$, respectively; furthermore, these bases project onto each other accordingly by subscripts, namely

$$\Pi_1 \hat{w}_i = \sigma_i w_i, \; \hat{\Pi}_1 w_i = \sigma_i \hat{w}_i.$$

Now we set $P_1^* := P_1^\circ U$ and $\hat{P}_1^* := \hat{P}_1^\circ V$. Note that

$$|1 - \sigma_i| = \|(\hat{\Pi}_1 - \Pi_1)\hat{w}_i\| < \varepsilon,$$

which shows, by properties of projection matrix Π_1,

$$\|w_i - \hat{w}_i\| = \sqrt{\|w_i - \Pi_1 \hat{w}_i\|^2 + \|\Pi_1 \hat{w}_i - \hat{w}_i\|^2} = \sqrt{|1-\sigma_i|^2 + \|(\hat{\Pi}_1 - \Pi_1)\hat{w}_i\|^2} < \sqrt{2}\varepsilon,$$

and thus

$$\|P_1^* - \hat{P}_1^*\| = \max_{\|z\|=1} \|(P_1^* - \hat{P}_1^*)z\| = \max_{\|z\|=1} \left\| \sum_i z_i (w_i - \hat{w}_i) \right\| \le \sqrt{k} \cdot \sqrt{2}\varepsilon.$$

To further generalize the proposition to any arbitrary $\hat{P}_1$, we only have to note that there exists an orthonormal matrix T that maps the basis $\hat{P}_1^*$ to $\hat{P}_1 = \hat{P}_1^* T$. Now take $P_1 = P_1^* T$, and we have

$$\|\hat{P}_1 - P_1\| = \|(\hat{P}_1^* - P_1^*)T\| = \|\hat{P}_1^* - P_1^*\| < \sqrt{2k}\varepsilon.$$

As for the estimation error bound for M_1, we can directly write

$$\begin{aligned} \|P_1^\top A P_1 - \hat{P}_1^\top A \hat{P}_1\| &\le \|P_1^\top A P_1 - P_1^\top A \hat{P}_1\| + \|P_1^\top A \hat{P}_1 - \hat{P}_1^\top A \hat{P}_1\| \\ &\le \|A\|\|P_1 - \hat{P}_1\| + \|A\|\|P_1 - \hat{P}_1\| \\ &< 2\|A\|\delta, \end{aligned}$$

This completes the proof of the corollary. □

Recall that we are allowed to take any orthonormal basis P_1 for E_{u}. Hence we shall always assume by default that P_1 in the proofs are selected as shown in the proof above.

We finish this section with simple but frequently-used bounds on $\|\hat{P}_1^\top P_1\|$ and $\|\hat{P}_1^\top P_2\|$. These factors represent an additional error introduced by using the inaccurate projector $\hat{P}_1$.

Proposition E.1. *Under the premises of Corollary 5.2, $\|I_k - \hat{P}_1^\top P_1\| < \delta$, $\|\hat{P}_1^\top P_2\| < \delta$.*

Proof. Note that $P_1^\top P_1 = I_k$ and $P_1^\top P_2 = O$, it is evident that

$$\begin{aligned} \|I_k - \hat{P}_1^\top P_1\| &= \|(P_1 - \hat{P}_1)^\top P_1\| < \delta, \\ \|\hat{P}_1^\top P_2\| &= \|(\hat{P}_1 - P_1)^\top P_2\| = \|\hat{P}_1 - P_1\| < \delta. \end{aligned}$$

This finishes the proof. □

F Proof of Theorem 4.2

We first consider a warm-up case where A is symmetric, which provides some intuition for the general case. In this case, the eigenvectors of A are mutually orthogonal, which guarantees $E_{\mathrm{u}}^\perp = E_{\mathrm{s}}$ (i.e., they are 0-close to each other) and thus $\Delta = \mathbf{0}$. This allows us to select $\tau = 1$, $\omega = 0$ and $\alpha = 1$, and the closed-loop dynamical matrix simplifies to

$$\hat{L}_1 = \begin{bmatrix} M_1 + P_1^\top B \hat{K}_1 \hat{P}_1^\top P_1 & P_1^\top B \hat{K}_1 \hat{P}_1^\top P_2 \\ P_2^\top B \hat{K}_1 \hat{P}_1^\top P_1 & M_2 + P_2^\top B \hat{K}_1 \hat{P}_1^\top P_2 \end{bmatrix}. \tag{11}$$

The norm of the top-left block is in the order of $O(\delta)$ based on the estimation error bound (see Theorem F.1) $\|\hat{B}_1 - B_1\| = O(\sqrt{k}\delta)$, which characterizes how well the controller can eliminate the unstable component. The spectrum of the bottom-right block can be viewed as a perturbation (note that $\|\hat{P}_1^\top P_2\| = O(\delta)$ is small by Proposition E.1) to a stable matrix M_2 (recall $\rho(M_2) = |\lambda_{k+1}|$), which should also be stable as long as δ is small enough. Meanwhile, the top-right block is also approximately zero, since only projection error contributes to the top-right block (again $\|\hat{P}_1^\top P_2\| = O(\delta)$). The above observations together show that $\hat{L}_1$ is in the order of

$$\hat{L}_1 = \begin{bmatrix} O(\delta) & O(\delta) \\ O(1) & |\lambda_{k+1}| + O(\delta) \end{bmatrix}, \tag{12}$$

which is almost lower-triangular. Therefore, we can apply the block perturbation bound to bound the spectrum of $\hat{L}_1$.

We start by showing the estimation error bound for B_1, which is straight-forward since $\Delta = \mathbf{0}$. Note that the upper bound of the norm of our controller $\hat{K}_1$ appears as a natural corollary of it.

Proposition F.1. *Under the premises of Theorem 4.2,* $\|\hat{B}_1 - B_1\| < 4\|A\|\sqrt{k}\delta$.

Proof. Note that the column vector b_i has estimation error bound

$$\begin{aligned}
\|b_i - \hat{b}_i\| &= \frac{1}{\|x_{t_i}\|}\left\|\left(P_1^\top x_{t_i+1} - M_1 P_1^\top x_{t_i}\right) - \left(\hat{P}_1^\top x_{t_i+1} - \hat{M}_1\hat{P}_1^\top x_{t_i}\right)\right\| \\
&\le \frac{1}{\|x_{t_i}\|}\left(\|(P_1^\top - \hat{P}_1^\top)Ax_{t_i}\| + \|(M_1P_1^\top - \hat{M}_1\hat{P}_1^\top)x_{t_i}\|\right) \\
&\le \|P_1^\top - \hat{P}_1^\top\|\|A\| + \|M_1P_1^\top - M_1\hat{P}_1^\top\| + \|M_1\hat{P}_1^\top - \hat{M}_1\hat{P}_1^\top\| \\
&< \|A\|\delta + \|M_1\|\|P_1^\top - \hat{P}_1^\top\| + \|M_1 - \hat{M}_1\| \\
&< \|A\|\delta + \|A\|\delta + 2\|A\|\delta = 4\|A\|\delta,
\end{aligned}$$

where we repeatedly apply Corollary 5.2 and the fact that $\|M_1\| \le \|A\|$. Then, to bound the error of the whole matrix, we simply apply the definition

$$\|\hat{B}_1 - B_1\| = \max_{\|u\|=1}\|(\hat{B}_1 - B_1)u\| \le \max_{\|u\|=1}\sum_{i=1}^{k}|u_i|\|\hat{b}_i - b_i\| < 4\|A\|\sqrt{k}\delta.$$

This completes the proof. □

Corollary F.1. *Under the premises of Theorem 4.2, when (13) holds,* $\|\hat{K}_1\| < \frac{2\|A\|}{c\|B\|}$.

Proof. By Proposition F.1, it is evident that

$$\sigma_{\min}(\hat{B}_1) \ge \sigma_{\min}(B_1) - \|\hat{B}_1 - B_1\| > (c - 4\|A\|\sqrt{k}\delta)\|B\| > \frac{c}{2}\|B\|,$$

where the last inequality requires

$$\delta < \frac{c}{8\|A\|\sqrt{k}}. \tag{13}$$

Recall that $\hat{K}_1 = \hat{B}_1^{-1}\hat{M}_1$, and note that $\|\hat{B}_1^{-1}\| \le \frac{1}{\sigma_{\min}(\hat{B}_1)}$, so we have

$$\|\hat{K}_1\| = \|\hat{B}_1^{-1}\hat{M}_1\| \le \frac{\|\hat{P}_1^\top A\hat{P}_1\|}{\sigma_{\min}(\hat{B}_1)} < \frac{2\|A\|}{c\|B\|}.$$

This completes the proof. □

Recall that to apply Lemma 5.3, we need a bound on the spectral radii of diagonal blocks. The top-left block has already been eliminated to approximately $\mathbf{0}$ by the design of $\hat{K}_1$, but the bottom-right block needs some extra work — although M_2 is known to be stable, the inaccurate projection introduces an extra error that perturbs the spectrum. To bound the perturbed spectral radius, we will apply the following perturbation bound known as Bauer-Fike Theorem.

Lemma F.2 (Bauer-Fike)**.** *Suppose* $A \in \mathbb{R}^{n\times n}$ *is diagonalizable, then for any* $E \in \mathbb{R}^{n\times n}$*, we have*

$$|\rho(A) - \rho(A+E)| \le \max_{\hat{\lambda}\in\lambda(A+E)}\min_{\lambda\in\lambda(A)}|\lambda - \hat{\lambda}| \le \kappa_{\mathrm{d}}(A)\|E\|,$$

where $\kappa_{\mathrm{d}}(A)$ *is the condition number of the matrix consisting of A's eigenvectors as columns (i.e., if* $A = S\Lambda S^{-1}$ *with diagonal* Λ*, then* $\kappa_{\mathrm{d}}(A) = \mathrm{cond}(S)$*), and* $\lambda(A)$ *denotes the spectrum of* A.

Proof. The proof is well-known and can be found in, e.g., [55]. □

Now we are ready to prove the main theorem for any symmetric dynamical matrix A.

Proof of Theorem 4.2. With $\tau = 1$, the controlled dynamics under estimated controller $\hat{K}_1$ becomes

$$\hat{L}_1 = \begin{bmatrix} M_1 + P_1^\top B\hat{K}_1\hat{P}_1^\top P_1 & P_1^\top B\hat{K}_1\hat{P}_1^\top P_2 \\ P_2^\top B\hat{K}_1\hat{P}_1^\top P_1 & M_2 + P_2^\top B\hat{K}_1\hat{P}_1^\top P_2 \end{bmatrix}.$$

We first guarantee that the diagonal blocks are stable. For the top-left block,

$$\begin{aligned}
\|M_1 + P_1^\top B\hat{K}_1\| &= \|M_1 - B_1\hat{B}_1^{-1}\hat{M}_1\hat{P}_1^\top P_1\| \\
&\le \|M_1 - \hat{M}_1\| + \|\hat{M}_1 - B_1\hat{B}_1^{-1}\hat{M}_1\| + \|B_1\hat{B}_1^{-1}\hat{M}_1(I_k - \hat{P}_1^\top P_1)\| \\
&\le \|M_1 - \hat{M}_1\| + \|\hat{B}_1 - B_1\|\|\hat{K}_1\| + \|B\|\|\hat{K}_1\|\|I_k - \hat{P}_1^\top P_1\| \\
&< 2\|A\|\delta + \frac{8\|A\|^2\sqrt{k}}{c\|B\|}\delta + \frac{2\|A\|}{c}\delta \qquad (14) \\
&= \frac{2\big(4\sqrt{k}\|A\| + (c+1)\|B\|\big)\|A\|}{c\|B\|}\delta,
\end{aligned}$$

where in (14) we apply Corollary 5.2, Corollary F.1, and Proposition E.1. Meanwhile, for the bottom-right block, note that the norm of the error term is bounded by

$$\|P_2^\top B\hat{K}_1\hat{P}_1^\top P_2\| \le \|B\|\|\hat{B}_1^{-1}\|\|\hat{M}_1\|\|\hat{P}_1^\top P_2\| \le \frac{2\|A\|}{c}\delta.$$

Hence, by Lemma F.2, the spectral radius of the bottom-right block is bounded by

$$\rho(M_2 + P_2^\top B\hat{K}_1\hat{P}_1^\top P_2) \le \rho(M_2) + \tfrac{2}{c}\kappa_{\mathrm{d}}(M_2)\|A\|\delta < 1,$$

where we require (recall that $\rho(M_2) = |\lambda_{k+1}|$)

$$\delta < \frac{c(1 - |\lambda_{k+1}|)}{2\kappa_{\mathrm{d}}(M_2)\|A\|}. \qquad (15)$$

To apply the lemma, it only suffices to bound the spectral norms of off-diagonal blocks. Note that the top-right block is bounded by

$$\|P_1^\top B\hat{K}_1\hat{P}_1^\top P_2\| \le \|B\|\|\hat{K}_1\|\|\hat{P}_1^\top P_2\| < \frac{2\|A\|}{c}\delta,$$

and the bottom-left block is bounded by

$$\|P_2^\top B\hat{K}_1\hat{P}_1^\top P_1\| \le \|B\|\|\hat{K}_1\| \le \frac{2\|A\|}{c}.$$

Now, by Lemma 5.3, we can guarantee that

$$\rho(\hat{L}_1) \le \max\left\{\frac{2\big(4\sqrt{k}\|A\| + 2(c+1)\|B\|\big)\|A\|}{c\|B\|}\delta, |\lambda_{k+1}| + \|B\|\|\hat{K}_1\|\delta\right\} + \frac{4\|A\|^2\chi(\hat{L}_1)}{c^2}\delta < 1,$$

where we require

$$\delta < \min\left\{\frac{1}{\frac{2\left(4\sqrt{k}\|A\|+2(c+1)\|B\|\right)\|A\|}{c\|B\|} + \frac{4\|A\|^2\chi(\hat{L}_1)}{c^2}}, \frac{1 - |\lambda_{k+1}|}{\frac{2\|A\|}{c} + \frac{4\|A\|^2\chi(\hat{L}_1)}{c^2}}\right\}. \qquad (16)$$

So far, it is still left to recollect all the constraints we need on δ (see (13), (15) and (16)), i.e.,

$$\delta < \min\left\{\frac{c}{8\|A\|\sqrt{k}}, \frac{c(1 - |\lambda_{k+1}|)}{2\kappa_{\mathrm{d}}(M_2)\|A\|}, \frac{1 - |\lambda_{k+1}|}{\frac{2\|A\|}{c} + \frac{4\|A\|^2\chi(\hat{L}_1)}{c^2}}, \frac{1}{\frac{2\left(4\sqrt{k}\|A\|+2(c+1)\|B\|\right)\|A\|}{c\|B\|} + \frac{4\|A\|^2\chi(\hat{L}_1)}{c^2}}\right\},$$

which can be simplified (but weakened) to

$$\delta < \frac{c^2(1 - |\lambda_{k+1}|)}{16\sqrt{k}\kappa_{\mathrm{d}}(M_2)\|A\|(\|A\| + \|B\|)\chi(\hat{L}_1)} = O(k^{-1/2}). \qquad (17)$$

We shall rewrite the bound equivalently in terms of t_0 (recall (10) in Appendix E) as

$$t_0 > \frac{\log(cn^2\binom{n}{k}) - \log\frac{c^2(1-|\lambda_{k+1}|)}{16\sqrt{2}k\kappa_{\mathrm{d}}(M_2)\|A\|(\|A\|+\|B\|)\chi(\hat{L}_1)}}{2\log\frac{|\lambda_k|}{|\lambda_{k+1}|}} = O\left(\frac{k\log n}{\log\frac{|\lambda_k|}{|\lambda_{k+1}|}}\right), \qquad (18)$$

since $\kappa_{\mathrm{d}}(A) = 1$. This completes the proof of Theorem 4.2. $\square$

G Proof of the Main Theorem

For the general case, the analysis becomes more challenging for two reasons: on the one hand, we have to apply τ-hop control with τ possibly larger than 1, which potentially increases the norm of B_τ and $\hat{K}_1$; on the other hand, the top-right corner will no longer be $O(\delta)$ with a non-zero Δ (in fact, Δ_τ is in the order of $|\lambda_1|^\tau$ that grows exponentially with respect to τ). To settle these issues, we first introduce two key observations on bounds of major factors:

(1) For an arbitrary matrix X, although $\|X\|$ might be significantly larger than $\rho(X)$, we always have $\|X^t\| = O(\rho(X)^t)$ when t is large enough. This is formally proven as Gelfand's Formula (see Lemma G.1), and helps to establish bounds like $\|M_1\| = O(|\lambda_1|^\tau)$, $\|M_2\| = O(|\lambda_{k+1}|^\tau)$, $\|\Delta_\tau\| = O(|\lambda_1|^\tau)$, $\|P_2^\top A^{\tau-1}\| = O(|\lambda_{k+1}|^\tau)$, and $\|\hat{M}_1^\tau - M_1^\tau\| = O(|\lambda_1|^\tau \delta)$.

(2) When the system runs with 0 control inputs for a long period (specifically, for ω time steps), eventually we will see the unstable component expanding and the stable component shrinking, and consequently $\frac{\|P_2^\top A^\omega x\|}{\|A^\omega x\|} = O(|\lambda_k|^{-\omega})$. This cancels out the exponentially exploding $\|\Delta_\tau\|$, and helps to establish the estimation bound $\|\hat{B}_\tau - B_\tau\| = O(|\lambda_1|^\tau \delta)$.

With these in hand, we are ready to upper bound the norms of the blocks in $\hat{L}_\tau$:

(1) *The top-left and bottom-right blocks*: similar to the warm-up case, only to note that dynamical matrices are lifted to their τ^{th} power, and thus $\|\hat{B}_\tau - B_\tau\|$ carries an additional factor of $|\lambda_1|^\tau$.

(2) *The bottom-left block*: $P_2^\top A^{\tau-1}$ contributes an $O(|\lambda_{k+1}|^\tau)$ factor that decays exponentially, while $\hat{K}_1$ contributes an $O(|\lambda_1|^\tau)$ factor that explodes exponentially. The overall bound is in the order of $O(|\lambda_1 \lambda_{k+1}/\lambda_k|^\tau)$, and decays with respect to τ if $|\lambda_1 \lambda_{k+1}| < 1$.

(3) *The top-right block*: the first term is in the order of $O(|\lambda_1|^\tau)$, and the second term is in the order of $O(|\lambda_1 \lambda_{k+1}/\lambda_k|^\tau \delta)$. This block is in the order of $O(|\lambda_1|^\tau)$ when δ is small enough.

Therefore, the closed-loop dynamical matrix is actually in the order of

$$\hat{L}_\tau = \begin{bmatrix} O(|\lambda_1|^{2\tau}\delta) & O\big(|\lambda_1|^\tau + |\lambda_1 \lambda_{k+1}/\lambda_k|^\tau \delta\big) \\ O(|\lambda_1 \lambda_{k+1}/\lambda_k|^\tau) & O\big(|\lambda_{k+1}|^\tau + |\lambda_1 \lambda_{k+1}|^\tau \delta\big) \end{bmatrix}. \tag{19}$$

Finally, by Lemma 5.3, asymptotic stability is guaranteed when $|\lambda_1|^2 |\lambda_{k+1}| < |\lambda_k|$ (i.e., the norm of the bottom-left block decays faster than the norm of the top-right block grows), in which case we can set τ to be some constant determined by A and B, and δ in the order of $O(|\lambda_1|^{-2\tau})$.

Technically, we would like to bound the spectral radius of the matrix

$$\hat{L}_\tau = \begin{bmatrix} M_1^\tau + P_1^\top A^{\tau-1} B \hat{K}_1 \hat{P}_1^\top P_1 & \Delta_\tau + P_1^\top A^{\tau-1} B \hat{K}_1 \hat{P}_1^\top P_2 \\ P_2^\top A^{\tau-1} B \hat{K}_1 \hat{P}_1^\top P_1 & M_2^\tau + P_2^\top A^{\tau-1} B \hat{K}_1 \hat{P}_1^\top P_2. \end{bmatrix}$$

using Lemma 5.3. The proof is split into two major building blocks: on the one hand, we introduce the well-known Gelfand's Formula to bound matrices appearing with exponents; on the other hand, we establish the estimation error bound for B_τ (parallel to Lemma F.1) and proceed to bound $\|\hat{K}_1\|$, for which we rely on the instability results shown in Section G.2. Finally, a combination of these building blocks naturally establishes the main theorem.

G.1 Gelfand's Formula

In this section, we will show norm bounds for factors that contain matrix exponents. It is natural to apply the well-known Gelfand's formula as stated below.

Lemma G.1 (Gelfand's formula)**.** *For any square matrix X, we have*

$$\rho(X) = \lim_{t \to \infty} \|X^t\|^{1/t}. \tag{20}$$

In other words, for any $\varepsilon > 0$, there exists a constant $\zeta_\varepsilon(X)$ such that

$$\sigma_{\max}(X^t) = \|X^t\| \le \zeta_\varepsilon(X)(\rho(X) + \varepsilon)^t. \tag{21}$$

Further, if X is invertible, let $\lambda_{\min}(X)$ denote the eigenvalue of X with minimum modulus, then

$$\sigma_{\min}(X^t) \ge \frac{1}{\zeta_\varepsilon(X^{-1})} \left(\frac{|\lambda_{\min}(X)|}{1 + \varepsilon |\lambda_{\min}(X)|} \right)^t. \tag{22}$$

Proof. The proof of (20) can be easily found in existing literature (e.g., [56], Corollary 5.6.14), and (21) follows by the definition of limits. For (22), note that

$$\sigma_{\min}(X^t) = \frac{1}{\sigma_{\max}((X^{-1})^t)} \geq \frac{1}{\zeta_\varepsilon(X^{-1})(\rho(X^{-1})+\varepsilon)^t} = \frac{1}{\zeta_\varepsilon(X^{-1})}\left(\frac{|\lambda_{\min}(X)|}{1+\varepsilon|\lambda_{\min}(X)|}\right)^t,$$

where we apply $\sigma_{\min}(X^t) = \sigma_{\max}((X^{-1})^t)^{-1}$ and $\rho(X^{-1}) = |\lambda_{\min}(X)|^{-1}$. □

It is evident that $\rho(A) = \rho(M_1) = \rho(N_1) = |\lambda_1|$, $\lambda_{\min}(M_1) = \lambda_{\min}(N_1) = |\lambda_k|$ and $\rho(M_2) = \rho(N_2) = |\lambda_{k+1}|$ (recall that M_1 and M_2 inherits the unstable and stable eigenvalues, respectively). Therefore, we can use Gelfand's formula to bound the relevant factors appearing in $\hat{L}_\tau$.

Proposition G.1. *Under the premises of Theorem 4.1, the following results hold for any $t \in \mathbb{N}$:*

(1) $\|B_t\| \leq \zeta_{\varepsilon_1}(A)(|\lambda_1|+\varepsilon_1)^{t-1}\|B\|$;

(2) $\|P_2^\top A^t\| \leq \zeta_{\varepsilon_2}(M_2)(|\lambda_{k+1}|+\varepsilon_2)^t$;

(3) $\|\varDelta_t\| \leq C_\varDelta(|\lambda_1|+\varepsilon_1)^t$, *where* $C_\varDelta = \zeta_{\varepsilon_1}(M_1)\zeta_{\varepsilon_2}(M_2)\frac{(2-\xi)\sqrt{2\xi}\|A\|}{1-\xi}\frac{2|\lambda_{k+1}|}{|\lambda_1|+\varepsilon_1-|\lambda_{k+1}|-\varepsilon_2}$.

Here (and below) ε_1 and ε_2 are selected to be sufficiently small constants (see (47)).

Proof. (1) This is a direct corollary of Gelfand's Formula, since

$$\|B_t\| = \|P_1^\top A^{t-1}B\| \leq \|A^{t-1}\|\|B\| \leq \zeta_{\varepsilon_1}(A)(|\lambda_1|+\varepsilon_1)^{t-1}\|B\|.$$

(2) It only suffices to recall $\rho(M_2) = |\lambda_{k+1}|$, and note that

$$P_2^\top A^t = P_2^\top P M^t P^{-1} = [\mathbf{0}\ I_{n-k}]M^t P^\top = M_2^t P_2^\top.$$

Hence by Gelfand's formula we have $\|P_2^\top A^t\| = \|M_2^t\| \leq \zeta_{\varepsilon_2}(M_2)(|\lambda_{k+1}|+\varepsilon_2)^t$.

(3) This is a direct corollary of Lemma A.1(4) and Gelfand's formula, since

$$\begin{aligned}
\|\varDelta_t\| &= \left\|\sum_i M_1^i \varDelta M_2^{t-1-i}\right\| \leq \|\varDelta\|\sum_i \|M_1^i\|\|M_2^{t-1-i}\| \\
&\leq \zeta_{\varepsilon_1}(M_1)\zeta_{\varepsilon_2}(M_2)\frac{(2-\xi)\sqrt{2\xi}\|A\|}{1-\xi}\sum_i(\varepsilon_1+|\lambda_1|)^i(|\lambda_{k+1}|+\varepsilon_2)^{t-1-i} \\
&= C_\varDelta(|\lambda_1|+\varepsilon_1)^t.
\end{aligned}$$

This finishes the proof of the proposition. □

Proposition G.2. *Under the premises of Theorem 4.1,*

$$\|\hat{M}_1^\tau - M_1^\tau\| < 2\tau\|A\|\zeta_{\varepsilon_1}(A)^2(|\lambda_1|+\varepsilon_1)^{\tau-1}\delta.$$

Proof. Recall that Corollary 5.2 gives $\|M_1 - \hat{M}_1\| < 2\|A\|\delta$. Meanwhile, by Gelfand's Formula,

$$\begin{aligned}
\|M_1^t\| &= \|P^\top A^t P\| \leq \|A^t\| \leq \zeta_{\varepsilon_1}(A)(|\lambda_1|+\varepsilon_1)^t, \\
\|\hat{M}_1^t\| &= \|\hat{P}^\top A^t \hat{P}\| \leq \|A^t\| \leq \zeta_{\varepsilon_1}(A)(|\lambda_1|+\varepsilon_1)^t.
\end{aligned}$$

Then we have the following bound by telescoping

$$\begin{aligned}
\|M_1^\tau - \hat{M}_1^\tau\| &= \left\|\sum_{i=1}^\tau \left(M_1^i\hat{M}_1^{\tau-i} - M_1^{i-1}\hat{M}_1^{\tau-i+1}\right)\right\| \\
&\leq \sum_{i=1}^\tau \|M_1^{i-1}\|\|\hat{M}_1^{\tau-i}\|\|M_1 - \hat{M}_1\| \\
&< \tau\cdot\zeta_{\varepsilon_1}(A)^2(|\lambda_1|+\varepsilon_1)^{\tau-1}\cdot 2\|A\|\delta \\
&= 2\tau\|A\|\zeta_{\varepsilon_1}(A)^2(|\lambda_1|+\varepsilon_1)^{\tau-1}\delta.
\end{aligned}$$

This finishes the proof. □

Corollary G.2. *Under the premises of Theorem 4.1, when $\delta < \frac{1}{\tau}$,*

$$\|\hat{M}_1^\tau\| < \big(\zeta_{\varepsilon_1}(M_1)(|\lambda_1| + \varepsilon_1) + 2\|A\|\zeta_{\varepsilon_1}(A)\big)(|\lambda_1| + \varepsilon_1)^{\tau-1}.$$

Proof. A combination of Gelfand's Formula and Proposition G.2 yields

$$\begin{aligned}\|\hat{M}_1^\tau\| &\le \|M_1^\tau\| + \|\hat{M}_1^\tau - M_1^\tau\| \\ &\le \zeta_{\varepsilon_1}(M_1)(|\lambda_1| + \varepsilon_1)^\tau + 2\tau\|A\|\zeta_{\varepsilon_1}(A)^2(|\lambda_1| + \varepsilon_1)^{\tau-1}\delta \\ &< \big(\zeta_{\varepsilon_1}(M_1)(|\lambda_1| + \varepsilon_1) + 2\tau\|A\|\zeta_{\varepsilon_1}(A)\delta\big)(|\lambda_1| + \varepsilon_1)^{\tau-1},\end{aligned}$$

where the last inequality requires $\delta < \frac{1}{\tau}$. This completes the proof. ☐

G.2 Instability of the Unstable Component

We have been referring to E_s (and approximately, $E_\mathrm{u}^\perp$) as "stable", and E_u as "unstable". This leads us to think that the unstable component will constitute an increasing proportion of the state as the system evolves with zero control input. However, in some cases it might happen that the proportion of unstable component does not increase within the first few time steps, although eventually it will explode. This motivates us to formally characterize such instability of the unstable component.

In this section, we aim to establish a fundamental property of A^ω (for large enough ω, of course) that it "almost surely" increases the norm of the state. By "almost surely" we mean that the initial state should have non-negligible unstable component, which happens with probability $1 - \varepsilon$ when we uniformly sample the initial state from the surface of unit hyper-sphere in $\mathbb{R}^n$.

Throughout this section, we use γ to denote the ratio of the unstable component over the stable component within some state x (i.e., $\frac{\|R_1 x\|}{\|R_2 x\|}$). Note that

$$x = \Pi_\mathrm{u} x + \Pi_\mathrm{s} x = Q_1 R_1 x + Q_2 R_2 x,$$

where Q_1, Q_2 are orthonormal. Hence

$$\|R_1 x\| - \|R_2 x\| \le \|x\| \le \|R_1 x\| + \|R_2 x\|.$$

As a consequence, when $\frac{\|R_1 x\|}{\|R_2 x\|} > \gamma > 1$, we also know that

$$\frac{\|R_1 x\|}{\|x\|} \ge \frac{\|R_1 x\|}{\|R_1 x\| + \|R_2 x\|} > \frac{\gamma}{\gamma+1}, \quad \frac{\|R_2 x\|}{\|x\|} \le \frac{\|R_2 x\|}{\|R_1 x\| - \|R_2 x\|} < \frac{1}{\gamma - 1}.$$

The following results are presented to fit in the framework of an inductive proof. We first establish the inductive step, where Proposition G.3 shows that the unstable component eventually becomes dominant with a non-negligible initial γ, and Proposition G.4 shows that the unstable component will still constitute a non-negligible part after a control input of mild magnitude is injected. Meanwhile, Proposition G.5 shows that the initial unstable component is non-negligible with large probability.

Proposition G.3. *Given a dynamical matrix A and some constant $\gamma > 0$, for any state x such that $\frac{\|R_1 x\|}{\|R_2 x\|} > \gamma$, for any $\omega \in \mathbb{N}$, we have*

$$\frac{\|R_1 A^\omega x\|}{\|R_2 A^\omega x\|} > \gamma_\omega := C_\gamma \left(\frac{|\lambda_k|}{(1 + \varepsilon_3|\lambda_k|)(|\lambda_{k+1}| + \varepsilon_2)}\right)^\omega,$$

where $C_\gamma := \frac{1}{(1+\frac{1}{\gamma})\zeta_{\varepsilon_3}(N_1^{-1})\zeta_{\varepsilon_2}(N_2)\|R_2\|}$ is a constant related to γ. Specifically, for any $\gamma_+ > 0$, there exists a constant $\omega_0(\gamma, \gamma_+) = O(\log \frac{\gamma_+}{\gamma})$, such that for any $\omega > \omega_0(\gamma, \gamma_+)$, $\frac{\|R_1 x\|}{\|R_2 x\|} > \gamma_+$.

Proof. Recall that $R_1 A^\omega = N_1^\omega R_1$ and $R_2 A^\omega = N_2^\omega R_2$. By Gelfand's Formula we have

$$\begin{aligned}\frac{\|R_1 A^\omega x\|}{\|R_2 A^\omega x\|} &= \frac{\|N_1^\omega R_1 x\|}{\|N_2^\omega R_2 x\|} \ge \frac{\sigma_{\min}(N_1^\omega)\|R_1 x\|}{\|N_2^\omega\|\|R_2\|\|x\|} > \frac{\sigma_{\min}(N_1^\omega)}{(1+\frac{1}{\gamma})\|N_2^\omega\|\|R_2\|} \\ &\ge \frac{\big(|\lambda_k|/(1+\varepsilon_3|\lambda_k|)\big)^\omega}{(1+\frac{1}{\gamma})\zeta_{\varepsilon_3}(N_1^{-1})\zeta_{\varepsilon_2}(N_2)(|\lambda_{k+1}| + \varepsilon_2)^\omega\|R_2\|}\end{aligned}$$

$$= \frac{1}{(1+\frac{1}{\gamma})\zeta_{\varepsilon_3}(N_1^{-1})\zeta_{\varepsilon_2}(N_2)\|R_2\|} \left(\frac{|\lambda_k|}{(1+\varepsilon_3|\lambda_k|)(|\lambda_{k+1}|+\varepsilon_2)} \right)^{\omega}.$$

Therefore, we shall take

$$\omega_0(\gamma,\gamma_+) = \frac{\log \gamma_+/C_\gamma}{\log(|\lambda_k|)/\big((1+\varepsilon_3|\lambda_k|)(|\lambda_{k+1}|+\varepsilon_2)\big)} = O\left(\log\frac{\gamma_+}{\gamma}\right),$$

and the proof is completed. □

Corollary G.3. *Under the premises of Proposition G.3, for any $\omega > \omega_0(\gamma,\gamma_+)$,*

$$\frac{\|P_1^\top A^\omega x\|}{\|A^\omega x\|} > 1 - \frac{2}{\gamma_\omega - 1}, \quad \frac{\|P_2^\top A^\omega x\|}{\|A^\omega x\|} < \frac{1}{\gamma_\omega - 1}.$$

Proof. Note that we have decomposition $x = \Pi_{\mathrm{u}}x + \Pi_1\Pi_{\mathrm{s}}x + \Pi_2\Pi_{\mathrm{s}}x$, where $\|\Pi_{\mathrm{u}}x\| = \|R_1x\|$ and $\|\Pi_{\mathrm{s}}x\| = \|R_2x\|$. Hence, for any $\omega > \omega_0(\gamma,\gamma_+)$, we can show that

$$\begin{aligned}
\frac{\|P_1^\top A^\omega x\|}{\|A^\omega x\|} &= \frac{\|\Pi_{\mathrm{u}}A^\omega x + \Pi_1\Pi_{\mathrm{s}}A^\omega x\|}{\|A^\omega x\|} \\
&\geq \frac{\|\Pi_{\mathrm{u}}A^\omega x\| - \|\Pi_1\Pi_{\mathrm{s}}A^\omega x\|}{\|A^\omega x\|} \\
&\geq \frac{\|R_1A^\omega x\| - \|R_2A^\omega x\|}{\|A^\omega x\|} \\
&> \frac{\gamma_\omega}{\gamma_\omega+1} - \frac{1}{\gamma_\omega-1} > 1 - \frac{2}{\gamma_\omega-1},
\end{aligned}$$

and similarly,

$$\frac{\|P_2^\top A^\omega x\|}{\|A^\omega x\|} = \frac{\|\Pi_2\Pi_{\mathrm{s}}A^\omega x\|}{\|A^\omega x\|} \leq \frac{\|\Pi_{\mathrm{s}}A^\omega x\|}{\|A^\omega x\|} < \frac{1}{\gamma_\omega-1}.$$

The proof is completed. □

Proposition G.4. *Given dynamical matrices A, B and constants $\gamma > 0, \gamma_+ > 1$, for any state x such that $\frac{\|R_1x\|}{\|R_2x\|} > \gamma_+$, suppose we feed a control input $\|u\| \leq \alpha\|x\|$ and observe the next state $x' = Ax + Bu$, where α satisfies*

$$\alpha < \frac{\frac{\gamma_+}{\gamma_++1}\sigma_{\min}(M_1) - \frac{\gamma}{\gamma_+-1}\frac{1}{1-\xi}\|A\|}{(1+\frac{\sqrt{2\xi}}{1-\xi}+\frac{\gamma}{1-\xi})\|B\|}. \tag{23}$$

Then we can guarantee that $\frac{\|R_1x'\|}{\|R_2x'\|} > \gamma$.

Proof. The proposition can be shown by direct calculation. Let $z = Rx = [z_1^\top, z_2^\top]^\top$. Recall that

$$Rx' = z' = \begin{bmatrix} N_1z_1 + R_1Bu \\ N_2z_2 + R_2Bu \end{bmatrix},$$

and note that $\frac{\|z_1\|}{\|x\|} > \frac{\gamma_+}{\gamma_++1}$, $\frac{\|z_2\|}{\|x\|} < \frac{1}{\gamma_+-1}$ under the assumptions, so we have

$$\begin{aligned}
\frac{\|R_1x'\|}{\|R_2x'\|} &= \frac{\|N_1z_1 + R_1Bu\|}{\|N_2z_2 + R_2Bu\|} \geq \frac{\|N_1z_1\| - \|R_1Bu\|}{\|N_2z_2\| + \|R_2Bu\|} \\
&\geq \frac{\sigma_{\min}(N_1)\|z_1\| - \|R_1B\|\|u\|}{\|N_2\|\|z_2\| + \|R_2B\|\|u\|} \\
&\geq \frac{\sigma_{\min}(N_1)\frac{\gamma_+}{\gamma_++1}\|x\| - \alpha\|R_1\|\|B\|\|x\|}{\|N_2\|\frac{1}{\gamma_+-1}\|x\| + \alpha\|R_2\|\|B\|\|x\|} \\
&\geq \frac{\sigma_{\min}(M_1)\frac{\gamma_+}{\gamma_++1}\|x\| - \alpha(1+\frac{\sqrt{2\xi}}{1-\xi})\|B\|\|x\|}{\frac{1}{1-\xi}\|A\|\frac{1}{\gamma_+-1}\|x\| + \alpha\frac{1}{1-\xi}\|B\|\|x\|} \\
&> \gamma,
\end{aligned}$$

where we apply Lemma A.1 and the convention of taking $N_1 = M_1$. □

Proposition G.5. *Suppose a state x is sampled uniformly randomly from the unit hyper-sphere surface $\mathbb{B}_n \subset \mathbb{R}^n$, then for any constant $\gamma < \min\left\{\frac{1}{2}, \frac{1}{\sqrt{2/(\sigma_{\min}(R_1)k)+1}}\right\}$, we have*

$$\mathrm{Pr}_{x\sim\mathcal{U}(\mathbb{B}_n)}\left[\frac{\|R_1 x\|}{\|R_2 x\|} > \gamma\right] > 1-\theta(\gamma),$$

where $\theta(\gamma) = \frac{8\sqrt{2}}{\mathrm{B}(\frac{1}{2},\frac{n-1}{2})\sqrt{\sigma_{\min}(R_1)}}\gamma = O(\gamma)$ is a constant bounded linearly by γ.

Proof. Note that

$$\|R_1 x\| > \frac{\gamma}{1-\gamma}\|x\| \;\Rightarrow\; \|R_2 x\| < \|x\| + \|R_1 x\| < \frac{1}{1-\gamma}\|x\| \;\Rightarrow\; \frac{\|R_1 x\|}{\|R_2 x\|} > \gamma.$$

so we only have to show that $\mathrm{Pr}_{x\sim\mathcal{U}(\mathbb{B}_n)}\left[\|R_1 x\| \le \frac{\gamma}{1-\gamma}\right] < \theta(\gamma)$. Now let $R_1^\top R_1 = S^\top D S$ be the eigen-decomposition of $R_1^\top R_1$, where S is selected to be orthonormal such that

$$D = \mathrm{diag}(d_1, \cdots, d_k, 0, \cdots, 0).$$

Note that the vector $y = Sx =: [y_1, \cdots, y_n]$ also obeys a uniform distribution over $\mathbb{B}_n$, so we have

$$\begin{aligned}
\mathrm{Pr}\left[\|R_1 x\| \le \tfrac{\gamma}{1-\gamma}\right] &= \mathrm{Pr}\left[x^\top R_1^\top R_1 x \le (\tfrac{\gamma}{1-\gamma})^2\right] = \mathrm{Pr}\left[y^\top D y \le (\tfrac{\gamma}{1-\gamma})^2\right] \\
&\le \mathrm{Pr}\left[d_i y_i^2 \le \tfrac{1}{k}(\tfrac{\gamma}{1-\gamma})^2,\ \forall i = 1,\ldots,k\right] \\
&\le \sum_{i=1}^{k} \mathrm{Pr}\left[y_i^2 \le \tfrac{1}{d_i k}(\tfrac{\gamma}{1-\gamma})^2\right].
\end{aligned}$$

It suffices to bound the probability $\mathrm{Pr}_{y\sim\mathcal{U}(B)}\left[y_i^2 \le \eta\right]$. Note that y can be obtained by first sampling a Gaussian random vector $z \sim \mathcal{N}(0, I_n)$, and then normalize it to get $y = \frac{z}{\|z\|}$. Hence

$$\mathrm{Pr}_{y\sim\mathcal{U}(\mathbb{B}_n)}\left[y_i^2 \le \eta\right] = \mathrm{Pr}_{z\sim\mathcal{N}(0,I_n)}\left[z_i^2 \le \eta\|z\|^2\right] = \mathrm{Pr}_{z\sim\mathcal{N}(0,I_n)}\left[\frac{z_i^2}{\sum_{j\neq i} z_j^2} \le \frac{\eta}{1-\eta}\right],$$

where $w := \frac{z_i^2}{\sum_{j\neq i} z_j^2}$ is known to obey an F-distribution $w \sim \mathcal{F}(1, n-1)$. The c.d.f. of w is known to be $I_{w/(w+n-1)}(\frac{1}{2}, \frac{n-1}{2})$, where I denotes the *regularized incomplete Beta function*. Note that

$$I_{w/(w+n-1)}\left(\frac{1}{2}, \frac{n-1}{2}\right) = \frac{2w^{1/2}}{(n-1)^{1/2}\mathrm{B}(\frac{1}{2},\frac{n-1}{2})} - \frac{n w^{3/2}}{3(n-1)^{3/2}\mathrm{B}(\frac{1}{2},\frac{n-1}{2})} + O(n^{5/2}),$$

it can be shown that $I_{w/(w+n-1)}\left(\frac{1}{2}, \frac{n-1}{2}\right) < \frac{4\sqrt{w}}{\sqrt{n-1}\mathrm{B}(\frac{1}{2},\frac{n-1}{2})}$. Hence

$$\mathrm{Pr}_{y\sim\mathcal{U}(\mathbb{B}_n)}\left[y_i^2 \le \eta\right] = \mathrm{Pr}_{z\sim\mathcal{N}(0,I_n)}\left[\frac{z_i^2}{\sum_{j\neq i} z_j^2} \le \frac{\eta}{1-\eta}\right] < \frac{4\sqrt{\frac{\eta}{1-\eta}}}{\sqrt{n-1}\mathrm{B}(\frac{1}{2},\frac{n-1}{2})},$$

which further gives

$$\mathrm{Pr}\left[\|R_1 x\| \le \tfrac{\gamma}{1-\gamma}\right] < \sum_{i=1}^{k} \frac{4\sqrt{\frac{2}{d_i k}(\frac{\gamma}{1-\gamma})^2}}{\sqrt{n-1}\mathrm{B}(\frac{1}{2},\frac{n-1}{2})} < \frac{8\sqrt{2}}{\mathrm{B}(\frac{1}{2},\frac{n-1}{2})\sqrt{\sigma_{\min}(R_1)}}\gamma = O(\gamma)$$

where we require $\gamma < \min\left\{\frac{1}{2}, \frac{1}{\sqrt{2/(\sigma_{\min}(R_1)k)+1}}\right\}$. □

Combining the previous three propositions, we have shown in an inductive way that the algorithm guarantees $\frac{\|P_2^\top x_{t_i}\|}{\|x_{t_i}\|}$ is constantly upper bounded at each time step t_i ($i = 1, \cdots, k$), which is critical to the estimation error bound of B_τ. This is concluded as the following lemma.

Lemma G.4. *Under the premises of Theorem 4.1, for any constants ω, γ such that $\omega < t_0$ and $\gamma < \min\left\{\frac{1}{2}, \frac{1}{\sqrt{2/(\sigma_{\min}(R_1)k)+1}}\right\}$, the algorithm guarantees*

$$\frac{\|P_2^\top x_{t_i}\|}{\|x_{t_i}\|} < \frac{1}{\gamma_\omega - 1},\ \forall i = 1, \cdots, k$$

with probability $1-\theta(\gamma)$ over the initialization of x_0 on the unit hyper-sphere surface $\mathbb{B}_n$, where

$$\gamma_\omega := C_\gamma \left(\frac{|\lambda_k|}{(1+\varepsilon_3|\lambda_k|)(|\lambda_{k+1}|+\varepsilon_2)}\right)^\omega.$$

Proof. We proceed by showing that $\frac{\|R_1 x_{t_i}\|}{\|R_2 x_{t_i}\|} > \gamma_\omega$ for $i = 1, \cdots, k$ in an inductive way.

For the base case, it is guaranteed by Proposition G.5 that x_0 satisfies $\frac{\|R_1 x_0\|}{\|R_2 x_0\|} > \gamma$ with probability $1-\theta(\gamma)$, and Proposition G.3 further guarantees $\frac{\|R_1 x_{t_1}\|}{\|R_2 x_{t_1}\|} > \gamma_\omega$. Here we require $t_0 > \omega$.

For the inductive step, suppose we have shown $\frac{\|R_1 x_{t_i}\|}{\|R_2 x_{t_i}\|} > \gamma_\omega$. Since $\|u_{t_i}\| = \alpha\|x_{t_i}\|$, we have $\frac{\|R_1 x_{t_i+1}\|}{\|R_2 x_{t_i+1}\|} > \gamma$ by Proposition G.4, and again Proposition G.3 guarantees $\frac{\|R_1 x_{t_{i+1}}\|}{\|R_2 x_{t_{i+1}}\|} > \gamma_\omega$.

Now it only suffices to apply Corollary G.3 to complete the proof. □

G.3 Estimation Error of B_τ

Proposition G.6. *Under the premises of Theorem 4.1 and Lemma G.4, when (29) holds,*

$$\|\hat{B}_\tau - B_\tau\| < C_B(|\lambda_1|+\varepsilon_1)^{\tau-1}\delta,$$

where $C_B := \frac{2\sqrt{k}\zeta_{\varepsilon_1}(A)^2\left((2\tau+2)\|A\|+\|B\|\right)}{\alpha}$.

Proof. This is parallel to Lemma F.1. Note that we have to subtract an additional term (induced by non-zero Δ_τ in M^τ) to calculate the actual b_i, so we have

$$\begin{aligned}
\|b_i - \hat{b}_i\| &= \frac{1}{\alpha\|x_{t_i}\|}\left\|\left(P_1^\top x_{t_i+\tau} - M_1^\tau P_1^\top x_{t_i} - \Delta_\tau P_2^\top x_{t_i}\right) - \left(\hat{P}_1^\top x_{t_i+\tau} - \hat{M}_1^\tau \hat{P}_1^\top x_{t_i}\right)\right\| \\
&\le \frac{1}{\alpha\|x_{t_i}\|}\left(\|(P_1-\hat{P}_1)^\top(A^\tau x_{t_i} + B_\tau u_{t_i})\| + \|M_1^\tau P_1^\top x_{t_i} - \hat{M}_1^\tau \hat{P}_1^\top x_{t_i}\| + \|\Delta_\tau P_2^\top x_{t_i}\|\right) \\
&< \frac{1}{\alpha}\left(\zeta_{\varepsilon_1}(A)^2(|\lambda_1|+\varepsilon_1)^{\tau-1}\left((2\tau+2)\|A\|+\|B\|\right)\delta + \delta\right).
\end{aligned}$$

Here the first term is bounded by

$$\begin{aligned}
\|(P_1-\hat{P}_1)^\top(A^\tau x_{t_i} + B_\tau u_{t_i})\| &\le \|P_1 - \hat{P}_1\|(\|A^\tau\| + \|A^{\tau-1}B\|)\|x_{t_i}\| \\
&< \|x_{t_i}\|\zeta_{\varepsilon_1}(A)(|\lambda_1|+\varepsilon_1)^{\tau-1}(\|A\|+\|B\|)\delta,
\end{aligned}$$

where in the last inequality we apply Corollary 5.2; the second term is bounded by

$$\begin{aligned}
\|M_1^\tau P_1^\top x_{t_i} - \hat{M}_1^\tau \hat{P}_1^\top x_{t_i}\| &\le (\|M_1^\tau(P_1^\top - \hat{P}_1^\top)\| + \|(M_1^\tau - \hat{M}_1^\tau)\hat{P}_1^\top\|)\|x_{t_i}\| \\
&< \left(\zeta_{\varepsilon_1}(A)(|\lambda_1|+\varepsilon_1)^{\tau-1}\|A\|\delta \right.\\
&\qquad \left. + 2\tau\|A\|\zeta_{\varepsilon_1}(A)^2(|\lambda_1|+\varepsilon_1)^{\tau-1}\delta\right)\|x_{t_i}\| && (24) \\
&\le \|x_{t_i}\|\zeta_{\varepsilon_1}(A)^2(|\lambda_1|+\varepsilon_1)^{\tau-1}(2\tau+1)\|A\|\delta, && (25)
\end{aligned}$$

where in (24) we apply Proposition G.2, and in (25) we apply a simple fact that $\zeta_{\varepsilon_1}(A) \ge 1$; the third term is bounded by

$$\begin{aligned}
\frac{\|\Delta_\tau\|\|P_2^\top x_{t_i}\|}{\|x_{t_i}\|} &\le \frac{C_\Delta(|\lambda_1|+\varepsilon_1)^\tau}{\left[C_\gamma\left(\frac{|\lambda_k|}{(1+\varepsilon_3|\lambda_k|)(|\lambda_{k+1}|+\varepsilon_2)}\right)^\omega - 1\right]} && (26) \\
&< \frac{2C_\Delta(|\lambda_1|+\varepsilon_1)^\tau}{C_\gamma\left(\frac{|\lambda_k|}{(1+\varepsilon_3|\lambda_k|)(|\lambda_{k+1}|+\varepsilon_2)}\right)^\omega} && (27)
\end{aligned}$$

$$< \delta, \tag{28}$$

where in (26) we apply Lemma G.4, while in (27) and (28) we require

$$\omega > \max\left\{\frac{\log 2/C_\gamma}{\log\big(|\lambda_k|/(1+\varepsilon_3|\lambda_k|)(|\lambda_{k+1}|+\varepsilon_2)\big)}, \frac{\log(2C_\Delta)/(C_\gamma\delta)+\tau\log(|\lambda_1|+\varepsilon_1)}{\log\big(|\lambda_k|/(1+\varepsilon_3|\lambda_k|)(|\lambda_{k+1}|+\varepsilon_2)\big)}\right\}. \tag{29}$$

Finally, to bound the error of the whole matrix, we simply apply the definition

$$\begin{aligned}
\|\hat{B}_\tau - B_\tau\| &= \max_{\|u\|=1}\|(\hat{B}_\tau - B_\tau)u\| \le \max_{\|u\|=1}\sum_{i=1}^{k}|u_i|\|\hat{b}_i - b_i\| \\
&< \frac{\sqrt{k}}{\alpha}\left(\zeta_{\varepsilon_1}(A)^2(|\lambda_1|+\varepsilon_1)^{\tau-1}\big((2\tau+2)\|A\|+\|B\|\big)+1\right)\delta \\
&< \frac{2\sqrt{k}\zeta_{\varepsilon_1}(A)^2\big((2\tau+2)\|A\|+\|B\|\big)}{\alpha}(|\lambda_1|+\varepsilon_1)^{\tau-1}\delta.
\end{aligned}$$

This completes the proof. □

Corollary G.5. *Under the premises of Theorem 4.1 and Lemma G.4, when (29), (30) and (31) hold,*

$$\sigma_{\min}(\hat{B}_\tau) > \frac{c\|B\|}{4\zeta_{\varepsilon_3}(N_1^{-1})}\left(\frac{|\lambda_k|}{1+\varepsilon_3|\lambda_k|}\right)^{\tau-1}.$$

Proof. We apply the $E_{\mathrm{u}} \oplus E_{\mathrm{s}}$-decomposition. Note that

$$B_\tau = P_1^\top A^{\tau-1}B = P_1^\top(Q_1N_1^{\tau-1}R_1 + Q_2N_2^{\tau-1}R_2)B = N_1^{\tau-1}R_1B + P_1^\top Q_2N_2^{\tau-1}R_2B,$$

so by Gelfand's Formula and Lemma A.1 we have

$$\begin{aligned}
\sigma_{\min}(B_\tau) &= \sigma_{\min}(N_1^{\tau-1}R_1B + P_1^\top Q_2N_2^{\tau-1}R_2B) \\
&\ge \sigma_{\min}(N_1^{\tau-1})\sigma_{\min}(R_1B) - \|P_1^\top Q_2\|\|N_2^{\tau-1}\|\|R_2\|\|B\| \\
&\ge \frac{c\|B\|}{\zeta_{\varepsilon_3}(N_1^{-1})}\left(\frac{|\lambda_k|}{1+\varepsilon_3|\lambda_k|}\right)^{\tau-1} - \frac{\sqrt{2\xi}\zeta_{\varepsilon_2}(N_2)\|B\|}{1-\xi}(|\lambda_{k+1}|+\varepsilon_2)^{\tau-1} \\
&> \frac{c\|B\|}{2\zeta_{\varepsilon_3}(N_1^{-1})}\left(\frac{|\lambda_k|}{1+\varepsilon_3|\lambda_k|}\right)^{\tau-1}
\end{aligned}$$

where the last inequality requires

$$\frac{\sqrt{2\xi}\zeta_{\varepsilon_2}(N_2)\zeta_{\varepsilon_3}(N_1^{-1})}{c(1-\xi)}\left(\frac{(|\lambda_{k+1}|+\varepsilon_2)(1+\varepsilon_3|\lambda_k|)}{|\lambda_k|}\right)^{\tau-1} < \frac{1}{2},$$

or equivalently,

$$\tau > \frac{\log\frac{c(1-\xi)}{2\sqrt{2\xi}\zeta_{\varepsilon_2}(N_2)\zeta_{\varepsilon_3}(N_1^{-1})}}{\log\frac{(|\lambda_{k+1}|+\varepsilon_2)(1+\varepsilon_3|\lambda_k|)}{|\lambda_k|}} + 1. \tag{30}$$

Therefore, using Proposition G.6, $\sigma_{\min}(\hat{B}_\tau)$ is lower bounded by

$$\begin{aligned}
\sigma_{\min}(\hat{B}_\tau) &\ge \sigma_{\min}(B_\tau) - \|\hat{B}_\tau - B_\tau\| \\
&> \frac{c\|B\|}{2\zeta_{\varepsilon_3}(N_1^{-1})}\left(\frac{|\lambda_k|}{1+\varepsilon_3|\lambda_k|}\right)^{\tau-1} - C_B(|\lambda_1|+\varepsilon_1)^{\tau-1}\delta \\
&> \frac{c\|B\|}{4\zeta_{\varepsilon_3}(N_1^{-1})}\left(\frac{|\lambda_k|}{1+\varepsilon_3|\lambda_k|}\right)^{\tau-1},
\end{aligned}$$

where the last inequality requires

$$\delta < \frac{c\|B\|}{4\zeta_{\varepsilon_3}(N_1^{-1})C_B}\left(\frac{|\lambda_k|}{(1+\varepsilon_3|\lambda_k|)(|\lambda_1|+\varepsilon_1)}\right)^{\tau-1}. \tag{31}$$

This completes the proof. □

Finally, using the above bounds, we can easily upper bound the norm of our controller $\hat{K}_1$.

Proposition G.7. *Under the premises of Theorem 4.1, when (29), (30), (31) and $\delta < \frac{1}{\tau}$ hold,*

$$\|\hat{K}_1\| < C_K \left(\frac{(|\lambda_1| + \varepsilon_1)(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|}\right)^{\tau-1},$$

where $C_K := \frac{4\zeta_{\varepsilon_3}(N_1^{-1})\left(\zeta_{\varepsilon_1}(M_1)(|\lambda_1|+\varepsilon_1)+2\|A\|\zeta_{\varepsilon_1}(A)\right)}{c\|B\|}$.

Proof. Recall that the controller is constructed as $\hat{K}_1 = \hat{B}_\tau^{-1}\hat{M}_1^\tau \hat{P}_1^\top$, so we have

$$\|\hat{K}_1\| \le \|\hat{B}_\tau^{-1}\|\|\hat{M}_1^\tau\| = \frac{\|\hat{M}_1^\tau\|}{\sigma_{\min}(\hat{B}_\tau)},$$

and the bound is merely a combination of Corollary G.2 and Corollary G.5 whenever $\delta < \frac{1}{\tau}$. $\square$

G.4 Proof of Theorem 4.1

Now we are ready to combine the above building blocks and present the complete proof of Theorem 4.1. Note that, with all the bounds established above, the proof structure parallels that of Theorem 4.2, the special case with a symmetric dynamical matrix A.

Proof of Theorem 4.1. The proof is again based on Lemma 5.3. We first guarantee that the diagonal blocks are stable. For the top-left block,

$$\begin{aligned}
\|M_1^\tau + P_1^\top A^{\tau-1}B\hat{K}_1\| &= \|M_1^\tau - B_\tau \hat{B}_\tau^{-1}\hat{M}_1^\tau \hat{P}_1^\top P_1\| \\
&\le \|M_1^\tau - \hat{M}_1^\tau\| + \|(B_\tau - \hat{B}_\tau)\hat{B}_\tau^{-1}\hat{M}_1^\tau\| + \|B_\tau \hat{B}_\tau^{-1}\hat{M}_1^\tau (I - \hat{P}_1^\top P_1)\| \\
&\le \|M_1^\tau - \hat{M}_1^\tau\| + \|B_\tau - \hat{B}_\tau\|\|\hat{K}_1\| + \|B_\tau\|\|\hat{K}_1\|\|I - \hat{P}_1^\top P_1\| \\
&\le 2\tau\|A\|\zeta_{\varepsilon_1}(A)^2(|\lambda_1| + \varepsilon_1)^{\tau-1}\delta \\
&\qquad + C_B C_K \left(\frac{(|\lambda_1| + \varepsilon_1)^2(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|}\right)^{\tau-1}\delta \qquad (32)\\
&\qquad + \zeta_{\varepsilon_1}(A)\|B\|C_K \left(\frac{(|\lambda_1| + \varepsilon_1)^2(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|}\right)^{\tau-1}\delta \\
&< (C_B C_K + \zeta_{\varepsilon_1}(A)\|B\|C_K + 1)\left(\frac{(|\lambda_1| + \varepsilon_1)^2(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|}\right)^{\tau-1}\delta \qquad (33)\\
&< \frac{1}{2}, \qquad (34)
\end{aligned}$$

where in (32) we apply Propositions G.2, G.6, G.7, and E.1; in (33) we require

$$\frac{1}{\tau}\left(\frac{(|\lambda_1| + \varepsilon_1)^2(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|}\right)^{\tau-1} > 2\|A\|\zeta_{\varepsilon_1}(A)^2; \qquad (35)$$

and in (34) we require

$$\delta < \frac{1}{2(C_B C_K + \zeta_{\varepsilon_1}(A)\|B\|C_K + 1)}\left(\frac{(|\lambda_1| + \varepsilon_1)^2(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|}\right)^{-(\tau-1)}. \qquad (36)$$

For the bottom-right block, it is straight-forward to see that

$$\begin{aligned}
\|M_2^\tau + P_2^\top A^{\tau-1}B\hat{K}_1\hat{P}_1^\top P_2\| &\le \|M_2^\tau\| + \|P_2^\top A^{\tau-1}\|\|B\|\|\hat{K}_1\|\|\hat{P}_1^\top P_2\| \\
&\le \zeta_{\varepsilon_2}(M_2)(|\lambda_{k+1}| + \varepsilon_2)^\tau \\
&\qquad + \zeta_{\varepsilon_2}(M_2)\|B\|C_K\left(\frac{(|\lambda_1| + \varepsilon_1)(|\lambda_{k+1}| + \varepsilon_2)(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|}\right)^{\tau-1}\delta \\
&< 1
\end{aligned}$$

where the last inequality requires

$$\tau > \frac{\log 1/(4\zeta_{\varepsilon_2}(M_2))}{\log(|\lambda_{k+1}| + \varepsilon_2)}, \tag{37}$$

$$\delta < \frac{1}{4\zeta_{\varepsilon_2}(M_2)\|B\|C_K} \left(\frac{(|\lambda_1| + \varepsilon_1)(|\lambda_{k+1}| + \varepsilon_2)(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|} \right)^{-(\tau-1)}. \tag{38}$$

Now it only suffices to bound the spectral norms of off-diagonal blocks. Note that, by applying Proposition G.7 and Proposition G.1, the top-right block is bounded as

$$\begin{aligned} \|\varDelta_\tau + P_1^\top A^{\tau-1} B\hat{K}_1 \hat{P}_1^\top P_2\| &\le \|\varDelta_\tau\| + \|B_\tau\|\|\hat{K}_1\|\|\hat{P}_1^\top P_2\| \\ &< C_\varDelta(|\lambda_1| + \varepsilon_1)^\tau \\ &\quad + \zeta_{\varepsilon_1}(A)\|B\|C_K \left(\frac{(|\lambda_1| + \varepsilon_1)^2(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|} \right)^{\tau-1} \delta \\ &< (C_\varDelta + 1)(|\lambda_1| + \varepsilon_1)^\tau \end{aligned}$$

where the last inequality requires

$$\delta < \frac{(|\lambda_1| + \varepsilon_1)^2}{\zeta_{\varepsilon_1}(A)\|B\|C_K} \left(\frac{(|\lambda_1| + \varepsilon_1)^2(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|} \right)^{-\tau}; \tag{39}$$

and the bottom-left block is bounded as

$$\begin{aligned} \|P_2^\top A^{\tau-1} B\hat{K}_1 \hat{P}_1^\top P_1\| &\le \|P_2^\top A^{\tau-1}\|\|B\|\|\hat{K}_1\| \\ &< \zeta_{\varepsilon_2}(M_2)\|B\|C_K \left(\frac{(|\lambda_1| + \varepsilon_1)(|\lambda_{k+1}| + \varepsilon_2)(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|} \right)^{\tau-1}. \end{aligned}$$

Now, by Lemma 5.3, we can guarantee that

$$\rho(\hat{L}_\tau) \le \frac{1}{2} + \chi(\hat{L}_\tau) \frac{(C_\varDelta + 1)\zeta_{\varepsilon_2}(M_2)\|B\|C_K}{|\lambda_1| + \varepsilon_1} \left(\frac{(|\lambda_1| + \varepsilon_1)^2(|\lambda_{k+1}| + \varepsilon_2)(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|} \right)^{\tau-1} < 1,$$

which requires

$$\tau > \frac{\log \frac{2(|\lambda_1| + \varepsilon_1)}{\chi(\hat{L}_\tau)(C_\varDelta + 1)\zeta_{\varepsilon_2}(M_2)\|B\|C_K}}{\log \frac{(|\lambda_1| + \varepsilon_1)^2(|\lambda_{k+1}| + \varepsilon_2)(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|}}. \tag{40}$$

Note that the above constraint makes sense only if $|\lambda_1|^2|\lambda_{k+1}| < |\lambda_k|$.

So far, it is still left to recollect all the constraints we need on the parameters $\tau, \alpha, \delta, \gamma$ and ω. To start with, all constraints on τ (see (30), (35), (37) and (40)) can be summarized as

$$\begin{aligned} \tau > \max \Bigg\{ & \frac{\log \frac{c(1-\xi)}{2\sqrt{2\xi}\zeta_{\varepsilon_2}(N_2)\zeta_{\varepsilon_3}(N_1^{-1})}}{\log \frac{(|\lambda_{k+1}| + \varepsilon_2)(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|}} + 1, \frac{\log 1/(4\zeta_{\varepsilon_2}(M_2))}{\log(|\lambda_{k+1}| + \varepsilon_2)}, \frac{\log \frac{2(|\lambda_1| + \varepsilon_1)}{\chi(\hat{L}_\tau)(C_\varDelta + 1)\zeta_{\varepsilon_2}(M_2)\|B\|C_K}}{\log \frac{(|\lambda_1| + \varepsilon_1)^2(|\lambda_{k+1}| + \varepsilon_2)(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|}}, \\ & -\frac{1}{\log \frac{(|\lambda_1| + \varepsilon_1)^2(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|}} W_{-1} \left(-\frac{\log \frac{(|\lambda_1| + \varepsilon_1)^2(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|}}{2\|A\|\zeta_{\varepsilon_1}(A)^2 \frac{(|\lambda_1| + \varepsilon_1)^2(1 + \varepsilon_3|\lambda_k|)}{|\lambda_k|}} \right) \Bigg\}, \end{aligned}$$

where W_{-1} denotes the non-principle branch of the Lambert-W function. Here we utilize the fact that, for $x > \frac{1}{\log a}$, $y = \frac{a^x}{x}$ is monotone increasing with inverse function $x = -\frac{1}{\log a} W_{-1}\big(-\frac{\log a}{y}\big)$, which can be upper bounded by Theorem 1 in [57] as

$$-\frac{1}{\log a} W_{-1} \left(-\frac{\log a}{y} \right) < \frac{\log y - \log\log a + \sqrt{2(\log y - \log\log a)}}{\log a} < \frac{3(\log y - \log\log a)}{\log a}.$$

By gathering different constants, we have

$$\tau > \frac{\log \frac{\sqrt{\xi}}{1-\xi} + \log \frac{1}{c} + \log \chi(\hat{L}_\tau) + 5\log \bar{\zeta} + \log \frac{\|A\|}{|\lambda_1| - |\lambda_{k+1}|} + C_\tau}{\log \frac{|\lambda_k|}{|\lambda_1|^2|\lambda_{k+1}|}} = O(1), \tag{41}$$

where we define $\bar{\zeta} := \max\{\zeta_{\varepsilon_1}(A), \zeta_{\varepsilon_2}(M_2), \zeta_{\varepsilon_2}(N_2), \zeta_{\varepsilon_3}(N_1^{-1})\}$, and C_τ is a numerical constant. Note that we have to guarantee the denominator to be positive, which gives rise to the additional assumption $|\lambda_1|^2|\lambda_{k+1}| < |\lambda_k|$. Meanwhile, for any $\ell \in \mathbb{N}$, we shall select γ such that

$$\gamma = O(k^{-\ell}), \quad \gamma < \min\left\{\frac{1}{2}, \frac{1}{\sqrt{2/(\sigma_{\min}(R_1)k)}+1}\right\}, \tag{42}$$

and select α such that (see (23), and we have already guaranteed $\gamma_\omega > 2$ in (29))

$$\alpha < \frac{\frac{2}{3}\sigma_{\min}(M_1) - \frac{\gamma}{1-\xi}\|A\|}{(1+\frac{\sqrt{2\xi}}{1-\xi}+\frac{\gamma}{1-\xi})\|B\|} = O(1). \tag{43}$$

Now constraints on δ (see (31), (36), (38) and (39)) can be summarized as

$$\begin{aligned}\delta < \min\Bigg\{ &\frac{c\|B\|}{4\zeta_{\varepsilon_3}(N_1^{-1})C_B}\left(\frac{|\lambda_k|}{(1+\varepsilon_3|\lambda_k|)(|\lambda_1|+\varepsilon_1)}\right)^{\tau-1}, \\ &\frac{1}{2(C_BC_K+\zeta_{\varepsilon_1}(A)\|B\|C_K+1)}\left(\frac{(|\lambda_1|+\varepsilon_1)^2(1+\varepsilon_3|\lambda_k|)}{|\lambda_k|}\right)^{-(\tau-1)}, \\ &\frac{1}{4\zeta_{\varepsilon_2}(M_2)\|B\|C_K}\left(\frac{(|\lambda_1|+\varepsilon_1)(|\lambda_{k+1}|+\varepsilon_2)(1+\varepsilon_3|\lambda_k|)}{|\lambda_k|}\right)^{-(\tau-1)}, \\ &\frac{(|\lambda_1|+\varepsilon_1)^2}{\zeta_{\varepsilon_1}(A)\|B\|C_K}\left(\frac{(|\lambda_1|+\varepsilon_1)^2(1+\varepsilon_3|\lambda_k|)}{|\lambda_k|}\right)^{-\tau}\Bigg\},\end{aligned}$$

which can be simplified to (C_δ is a constant collecting minor factors)

$$\delta < \frac{C_\delta \alpha c}{\sqrt{k}\bar{\zeta}^3(\|A\|+\|B\|)}|\lambda_1|^{-2\tau} = O(k^{-1/2}|\lambda_1|^{-2\tau}), \tag{44}$$

or we can rewrite the bound equivalently in terms of t_0 (recall (10) in Appendix E) as

$$\begin{aligned} t_0 &> \frac{\log(n^2\binom{n}{k}) + \log k + \log\kappa_{\mathrm{d}}(A) + 2\tau\log|\lambda_1| + 3\log\bar{\zeta} + \log(\|A\|+\|B\|) + \log\frac{\sqrt{2}}{C_\delta\alpha}}{2\log\frac{|\lambda_k|}{|\lambda_{k+1}|}} \\ &= O\left(\frac{2\tau\log|\lambda_1| + k\log n + \log\kappa_{\mathrm{d}}(A)}{\log\frac{|\lambda_k|}{|\lambda_{k+1}|}}\right), \end{aligned} \tag{45}$$

Finally, we select ω such that (see (29), and note that $C_\gamma = O(\gamma) = O(k^{-\ell})$)

$$\omega > \max\left\{\frac{\log\frac{2}{C_\gamma}}{\log\frac{|\lambda_k|}{(1+\varepsilon_3|\lambda_k|)(|\lambda_{k+1}|+\varepsilon_2)}}, \frac{\log\frac{2C_\Delta}{C_\gamma\delta}+\tau\log(|\lambda_1|+\varepsilon_1)}{\log\frac{|\lambda_k|}{(1+\varepsilon_3|\lambda_k|)(|\lambda_{k+1}|+\varepsilon_2)}}\right\},$$

which can be reorganized as

$$\omega > \frac{\log\frac{1}{C_\gamma} + \log\frac{\sqrt{\xi}}{1-\xi} + 2\log\bar{\zeta} + \log\frac{\|A\|}{|\lambda_1|-|\lambda_{k+1}|} + \log\frac{1}{\delta} + C_\omega}{\log\frac{|\lambda_k|}{|\lambda_{k+1}|}} = O(\ell\log k). \tag{46}$$

Note that here $\varepsilon_1, \varepsilon_2, \varepsilon_3$ are taken to be small enough, so that

$$|\lambda_{k+1}| + \varepsilon_2 < 1, \quad |\lambda_1| + \varepsilon_1)^2(|\lambda_{k+1}|+\varepsilon_2) < \frac{|\lambda_k|}{1+\varepsilon_3|\lambda_k|}, \quad \varepsilon_3|\lambda_k| < 1. \tag{47}$$

Also, the probability of sampling an admissible x_0 is $1-\theta(\gamma) = 1 - O(k^{-\ell})$ by the union bound.

Finally, by (41), (45) and (46), we conclude that Algorithm 1 terminates within

$$t_0 + k(1+\omega+\tau) > \frac{1}{2\log\frac{|\lambda_k|}{|\lambda_{k+1}|}}\left(\underbrace{\log(n^2\tbinom{n}{k})}_{O(k\log n)} + \underbrace{2k\log\frac{1}{C_\gamma}}_{O(k\log k)} + \log k\right) + k$$

$$
\begin{aligned}
&+ \frac{\log \kappa_{\mathrm{d}}(A) + 2\tau \log|\lambda_1| + 3\log\bar{\zeta} + \log(\|A\| + \|B\|) + \log\frac{\sqrt{2}}{C_\delta \alpha}}{2\log\frac{|\lambda_k|}{|\lambda_{k+1}|}} \\
&+ \frac{k\left(\log\frac{\sqrt{\xi}}{1-\xi} + 2\log\bar{\zeta} + \log\frac{\|A\|}{|\lambda_1| - |\lambda_{k+1}|} + \log\frac{1}{\delta} + C_\omega\right)}{\log\frac{|\lambda_k|}{|\lambda_{k+1}|}} \\
&+ \frac{k\left(\log\frac{\sqrt{\xi}}{1-\xi} + \log\frac{1}{c} + \log\chi(\hat{L}_\tau) + 5\log\bar{\zeta} + \log\frac{\|A\|}{|\lambda_1| - |\lambda_{k+1}|} + C_\tau\right)}{\log\frac{|\lambda_k|}{|\lambda_1|^2|\lambda_{k+1}|}} \\
= {} & O(k\log n),
\end{aligned}
$$

time steps, which completes the proof. □

For the convenience of readers, we provide a table summarizing all constants appearing in the bound.

Table 1: Lists of parameters and constants appearing in the bound.

(a) Algorithmic parameters (introduced in Algorithm algorithm 1).

Constant	**Appearance**	**Explanation**
t_0	Stage 1	t_0 initialization steps to separate unstable components
ω	Stage 3	ω heat-up steps in each iteration of learning B_τ
α	Stage 3	$\|u_{t_i}\| = \alpha\|x_{t_i}\|$ to keep non-negligible unstable component
τ	Stage 4	τ steps between consecutive control inputs are injected

(b) System parameters (as functions of dynamical matrices).

Constant	**Definition**	**Explanation**
λ_i	Assumption 4.1	(complex) eigenvalue of A with i^{th} largest modulus
$\|A\|, \|B\|$	Notation	2-norm of dynamical matrices A and B
c	Assumption 4.3	c effective controllability over the unstable subspace E_{u}, i.e., $\sigma_{\min}(R_1 B) > c\|B\|$
ξ	Definition 3.1	$E_{\mathrm{u}}^{\perp}$ and E_{s} are ξ-close subspaces, i.e., $\sigma_{\min}(P_2^\top Q_2) > 1 - \xi$
$\chi(\cdot)$	Lemma D.1	perturbation constant for 2-by-2 block diagonal matrices
$\zeta_\varepsilon(\cdot)$	Lemma G.1	Gelfand constant for the norm of matrix exponents
$\kappa_{\mathrm{d}}(\cdot)$	Notation	the diagonalization condition number, i.e., condition number of the matrix formed by eigenvectors as columns

(c) Shorthand notations (introduced in proofs).

Constant	**Definition**	**Explanation**
C_Δ	Proposition G.1	$C_\Delta := \zeta_{\varepsilon_1}(M_1)\zeta_{\varepsilon_2}(M_2)\frac{(2-\xi)\sqrt{2\xi}\|A\|}{1-\xi}\frac{2\|\lambda_{k+1}\|}{\|\lambda_1\|+\varepsilon_1-\|\lambda_{k+1}\|-\varepsilon_2}$
C_γ	Proposition G.3	$C_\gamma := \frac{1}{(1+\frac{1}{\gamma})\zeta_{\varepsilon_3}(N_1^{-1})\zeta_{\varepsilon_2}(N_2)\|R_2\|}$ (γ is taken according to (42))
C_B	Proposition G.6	$C_B := \frac{2\sqrt{k}\zeta_{\varepsilon_1}(A)^2\left((2\tau+2)\|A\|+\|B\|\right)}{\alpha}$
C_K	Proposition G.7	$C_K := \frac{4\zeta_{\varepsilon_3}(N_1^{-1})\left(\zeta_{\varepsilon_1}(M_1)(\|\lambda_1\|+\varepsilon_1)+2\|A\|\zeta_{\varepsilon_1}(A)\right)}{c\|B\|}$
$\bar{\zeta}$	below (41)	$\bar{\zeta} := \max\{\zeta_{\varepsilon_1}(A), \zeta_{\varepsilon_2}(M_2), \zeta_{\varepsilon_2}(N_2), \zeta_{\varepsilon_3}(N_1^{-1})\}$
$C_\tau, C_\delta, C_\omega$	(41), (44), (46)	collection of numerical constants in (41), (44), (46)

H An Illustrative Example with Additive Noise

Finally, we include an illustrative experiment that shows the performance of our LTS_0 algorithm.

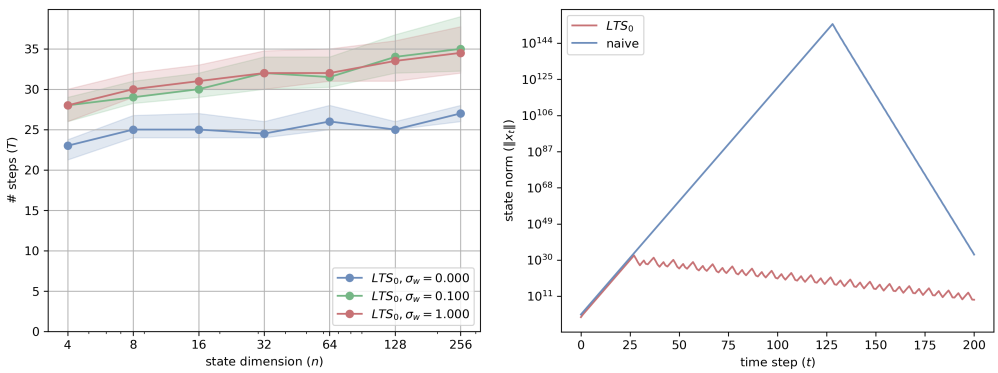
Settings. We evaluate the algorithm in LTI systems with additive noise

$$x_{t+1} = Ax_t + Bu_t + w_t, \quad \text{where } w_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_w^2 I).$$

Here σ_w characterizes the variance (and thus the magnitude) of the noise. The dynamical matrices are randomly generated: A is generated based on its eigen-decomposition $A = V\Lambda V^{-1}$, where the eigenvalues $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$ are randomly generated by selecting $\lambda_{1:k} \sim \mathcal{U}(1, \lambda_{\max})$ and $\lambda_{k+1:n} \sim \frac{|\lambda_k|}{|\lambda_1|^2} \cdot \mathcal{U}(-1, 1)$ (to ensure $|\lambda_1|^2 |\lambda_{k+1}| < |\lambda_k|$), and the eigenvectors $V = [v_1, \cdots, v_n]$ are generated by random perturbation to a random orthogonal matrix (to avoid tiny ξ); meanwhile, B is generated by random sampling i.i.d. entries from $\mathcal{U}(0, 1)$. For comparability and reproducibility, throughout the experiment we set $k = 3$ and use 0 as the initial random seed.

To compare the performance in different settings, 30 data points are collected for each pair of σ_w and n. It is observed that our algorithm might cause numerical instability issues (e.g., $\mathrm{cond}(D^\top D)$ could be large), so we simply ignore such cases and repeat until 30 data points are collected. The parameters of the algorithm are determined in an adaptive way that minimizes the number of running steps: we search for the minimum t_0 that yields estimation error smaller than δ, search for the minimum τ such that $K = B_\tau^{-1} M_1^\tau P_1^\top$ stabilizes the system, and the ω heat-up steps in Stage 3 could be ended earlier if we already observe $\|\hat{P}_1^\top x\| / \|x\|$ larger than a certain threshold.

Our experimental results are presented in Figure 1 below.



(a) Running steps of LTS_0

(b) State norms along one trajectory

Figure 1: Experimental results. *In* (*a*)*, the line shows the median of running steps, and the shadow marks the range between upper and lower quartiles* (*the horizontal axis is in log scale*)*. In* (*b*)*, the trajectories of our algorithm and the naive approach are compared in a randomly-generated system with* $n = 128$ *and* $\sigma_w = 0$ (*the vertical axis is in log scale*)*.*

Performance under different n and σ_w. Figure 1a shows the number of running steps of LTS_0 that is needed to learn a stabilizing controller. It is evident that the number of running steps grow almost linearly with regard to $\log n$, which is in accordance with Theorem 4.1.

As for the effect of noise, it is observed that the algorithm needs more steps in systems with noise than in those without noise; nevertheless, the magnitude of noise does not have much influence on the number of running steps. This is also reasonable since the increase is mainly attributed to t_0 — it takes more initial steps to push the state close enough to E_u, such that the estimation error of P_1 drops to acceptable level; however, as the E_u-component grows exponentially fast over time while w_t is i.i.d., the magnitude of noise only plays a minor role in the increase. Noise becomes negligible in later stages due to the disproportionate magnitudes of states and noise.

Analysis of comparison of trajectories. In Figure 1b we study an exemplary trajectory of our LTS_0 algorithm, and compare it against that of the naive approach, which first identifies the system and then designs a controller to nullify the unstable eigenvalues by standard pole-placement method. It is evident that our algorithm needs significantly fewer steps, and thus induces far smaller state norms, to learn a controller that effectively stabilizes the system. It is also observed that our controller decreases state norm in a zig-zag manner, which is due to the τ-hop design our algorithm adopts. Nevertheless, a potential drawback of our controller design is that the spectral radius of the controlled system is larger (since we cannot precisely nullify all unstable eigenvalues), resulting in a slower stabilizing rate than the naive approach (compare the decreasing parts of the curves).