ECE 598: Interplay between Control and Machine Learning **Spring 2019**

Lecture 12

Zames-Falb IQCs for Convergence Rate Analysis

Lecturer: Bin Hu, Date:02/26/2019

In the last lecture, we have introduced the Zames-Falb IQCs that can be used to show the boundedness of the states, i.e. $V(\xi_k) \leq V(\xi_0)$. In this lecture, we will modify the Zames-Falb IQCs to show convergence rate bounds in the form of $V(\xi_k) \leq \rho^{2k} V(\xi_0)$. For this purpose, we need a stronger notion of IQCs. Specifically, we will introduce ρ-hard IQCs.

12.1 Convergence Rate Analysis Using ρ-Hard IQCs

First, let's formally define ρ-hard IQCs.

Definition 1. *Let Ψ be an LTI system governed by the state-space model*

$$\begin{aligned} \psi_{k+1} &= A_\psi \psi_k + B_{\psi 1} v_k + B_{\psi 2} w_k \\ r_k &= C_\psi \psi_k + D_{\psi 1} v_k + D_{\psi 2} w_k \end{aligned} \tag{12.1}$$

where $\det(A_\psi - I) \neq 0$. *Suppose* $M = M^T \in \mathbb{R}^{n_r \times n_r}$. *Given the reference points* (v^*, w^*), *we specify* (ψ^*, r^*) *by solving the following fixed point condition:*

$$\begin{aligned} \psi^* &= A_\psi \psi^* + B_{\psi 1} v^* + B_{\psi 2} w^* \\ r^* &= C_\psi \psi^* + D_{\psi 1} v^* + D_{\psi 2} w^* \end{aligned} \tag{12.2}$$

The operator Δ satisfies the time domain ρ-hard IQC defined by $(\Psi, M, \rho, v^*, w^*)$ if the following inequality holds for all $w = \Delta(v)$ and $N \geq 0$

$$\sum_{k=0}^{N} \rho^{-2k} (r_k - r^*)^\mathsf{T} M (r_k - r^*) \leq 0 \tag{12.3}$$

where r is the output of the state-space model (12.1) *with inputs* (v, w) *and an initial condition* $\psi_0 = \psi^*$.

Again, typically control papers will use "$\geq$" in (12.3). Here we want to interpret (12.3) as a supply rate condition and hence use "$\leq$" instead.

The dependence on ρ. The condition (12.3) depends on ρ. As $k \to \infty$, the term ρ^{-2k} blows up to infinity. If $\rho = 1$, then we recover the notion of standard hard IQCs. Usually Ψ itself also depends on ρ. We will demonstrate this by an example.

Graphical interpretation. Notice that ρ-hard IQCs yield a similar graphical interpretation. In Figure 12.1, let the input and output signals of Δ be filtered through Ψ with the initial condition $\psi_0 = \psi^*$. The ρ-hard IQC condition (12.3) just enforces a quadratic inequality on the signal r.

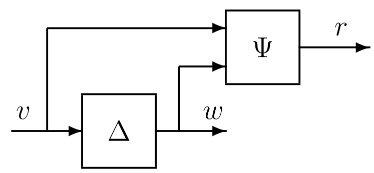

Figure 12.1. Graphical Interpretation for ρ-Hard IQCs

Now we are ready to modify the dissipation inequality framework for convergence rate analysis. As shown in Figure 12.2, we remove Δ and enforce the constraint (12.3) on the filtered signal r. We have $r = \Psi(v, w) = \Psi(G(w), w)$ and r must satisfy the constraint (12.3). Again, we only need to analyze the composite system $\Psi(G(w), w)$ with input w and the output r.

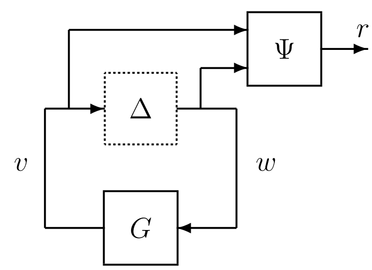

Figure 12.2. System G Extended to Include Filter Ψ

Suppose G is LTI and governed by

$$\begin{aligned} \xi_{k+1} &= A\xi_k + Bw_k \\ v_k &= C\xi_k \end{aligned}$$

In the last lecture, we have already shown that the augmented system $\Psi(G(w), w)$ is described by the following state space model

$$\begin{aligned} \eta_{k+1} &= \hat{A}\eta_k + \hat{B}w_k \\ r_k &= \hat{C}\eta_k + \hat{D}w_k \end{aligned} \tag{12.4}$$

where $\eta_k = \begin{bmatrix} \xi_k \\ \psi_k \end{bmatrix}$, $\hat{A} = \begin{bmatrix} A & 0 \\ B_{\psi 1}C & A_\psi \end{bmatrix}$, $\hat{B} = \begin{bmatrix} B \\ B_{\psi 2} \end{bmatrix}$, $\hat{C} = \begin{bmatrix} D_{\psi 1}C & C_\psi \end{bmatrix}$, and $\hat{D} = D_{\psi 2}$.

If there exists a positive definite matrix P such that

$$\begin{bmatrix} \hat{A}^\mathsf{T} P \hat{A} - \rho^2 P & \hat{A}^\mathsf{T} P \hat{B} \\ \hat{B}^\mathsf{T} P \hat{A} & \hat{B}^\mathsf{T} P \hat{B} \end{bmatrix} \leq \begin{bmatrix} \hat{C} & \hat{D} \end{bmatrix}^\mathsf{T} M \begin{bmatrix} \hat{C} & \hat{D} \end{bmatrix} \tag{12.5}$$

then the exponential dissipation inequality $V(\eta_{k+1}) - \rho^2 V(\eta_k) \leq S(\eta_k, w_k)$ holds with $V(\eta_k) = (\eta_k - \eta^*)^\mathsf{T} P(\eta_k - \eta^*)$ and $S(\eta_k, w_k) = r_k^\mathsf{T} M r_k$. This dissipation inequality can be rewritten as $\rho^{-2k} V(\eta_{k+1}) - \rho^{-2k+2} V(\eta_k) \leq \rho^{-2k} S(\eta_k, w_k)$. Based on the ρ-hard IQC condition (12.3), one have $\rho^{-2N} V(\eta_{N+1}) \leq \rho^2 V(\eta_0)\ \forall N$. Therefore, we have $V(\eta_k) \leq \rho^{2k} V(\eta_0)$.

The function V_k is not monotonically decreasing! Notice that the ρ-hard IQCs do not lead to the conclusion $V(\xi_{k+1}) \leq \rho^2 V(\xi_k)$ in general. Therefore, V is not a Lyapunov function. The function V may increase for certain k but $V(\xi_k)$ is bounded above by $\rho^{2k} V(\xi_0)$. Allowing V to be non-monotone makes the analysis less conservative.

12.2 Weighted Off-by-One IQC

Suppose Δ maps v to w as $w_k = \nabla f(v_k)$ where f is L-smooth and m-strongly convex. What is the most commonly-used ρ-hard IQC for such Δ?

First, it is obvious that pointwise quadratic constraints directly lead to ρ-hard IQCs for any ρ. Hence we can choose $r_k = \begin{bmatrix} v_k \\ w_k \end{bmatrix}$ and obtain the following condition:

$$\sum_{k=0}^{N} \rho^{-2k} \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}^\mathsf{T} \begin{bmatrix} 2mLI & -(m+L)I \\ -(m+L)I & 2I \end{bmatrix} \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix} \leq 0,\ \forall k \tag{12.6}$$

Again, (12.6) is conservative in the sense that it does not reflect the fact that the function f is not time-varying. Just imagine $w_k = \nabla f_k(v_k)$ where f_k is L-smooth and m-strongly convex for all k. Assume $\nabla f_k(v^*) = 0$ for all k (all the functions at the different time steps share the same global min). Then (12.6) still holds. This condition does not exploit the fact that the function f is not changing over time.

We can modify the Zames-Falb IQCs to fix the above issue. When f is L-smooth and m-strongly convex, we can actually prove the following inequality for $w = \Delta(v)$:

$$\sum_{k=0}^{N} \rho^{-2k} (-m(v_k - v^*) + (w_k - w^*))^\mathsf{T} (L(v_k - v^*) - (w_k - w^*) - \rho^2 L(v_{k-1} - v^*) + \rho^2 (w_{k-1} - w^*)) \geq 0 \tag{12.7}$$

where v_{-1} is defined to be v^* satisfying $\nabla f(v^*) = 0$, and $w_{-1} = \nabla f(v_{-1}) = 0$. In addition, we have $w^* = \nabla f(v^*) = 0$.

We skip the proof for the above ρ-hard IQC. Let's try to rewrite the above inequality in a filter form. If we choose $r_k = \begin{bmatrix} Lv_k - w_k - \rho^2 L v_{k-1} + \rho^2 w_{k-1} \\ -m v_k + w_k \end{bmatrix}$ and $M = \begin{bmatrix} 0 & -I \\ -I & 0 \end{bmatrix}$, then (12.3) and (12.7) are just the same. The question becomes how to generate $r_k = \begin{bmatrix} Lv_k - w_k - \rho^2 Lv_{k-1} + \rho^2 w_{k-1} \\ -mv_k + w_k \end{bmatrix}$. Notice r_k can only explicitly depend on ψ_k, v_k, and w_k. Since r_k cannot explicitly depend on v_{k-1} and w_{k-1}, we need to use ψ_k to memorize $(Lv_{k-1} - w_{k-1})$. We will have $\psi_k = Lv_{k-1} - w_{k-1}$ and hence $\psi_{k+1} = Lv_k - w_k$. Therefore, we have the following filter dynamics:

$$\begin{aligned} \psi_{k+1} &= Lv_k - w_k \\ r_k &= \begin{bmatrix} -\rho^2 I \\ 0 \end{bmatrix} \psi_k + \begin{bmatrix} LI \\ -mI \end{bmatrix} v_k + \begin{bmatrix} -I \\ I \end{bmatrix} w_k \end{aligned}$$

with the initial condition $\psi_0 = Lv_{-1} - w_{-1} = Lv^*$. It is straightforward to verify that the fixed point of the filter is given by $\psi^* = Lv^*$ and $r^* = \begin{bmatrix} (1-\rho^2)Lv^* \\ -mv^* \end{bmatrix}$ due to the fact $w^* = 0$. Therefore, we can rewrite (12.7) as an ρ-hard IQC by choosing $A_\psi = 0$, $B_{\psi 1} = LI$, $B_{\psi 2} = -I$, $C_\psi = \begin{bmatrix} -\rho^2 I \\ 0 \end{bmatrix}$, $D_{\psi 1} = \begin{bmatrix} LI \\ -mI \end{bmatrix}$, and $D_{\psi 2} = \begin{bmatrix} -I \\ I \end{bmatrix}$. You can use these matrices to formulate the LMI for Problem 2 in HW1!

Comparison with the Lure Postnikov Lyapunov function approach. In Lecture 10, we introduce the Lure Postnikov Lyapunov function approach for analyzing Nesterov's method. The resultant LMI is 3×3, and the size of P is 2×2. The above ρ-hard IQC will leads to a 4×4 LMI, and the size of P becomes 3×3. This makes analytical analysis more difficult. However, the analysis result from the ρ-hard IQC does improve the result from the Lure Postnikov Lyapunov function approach by a constant factor. You will see this in the homework.

Other Zames-Falb IQCs. There is a general routine that tailors Zames-Falb IQCs for convergence rate analysis, and (12.7) is only one example. Since (12.7) looks one step back and uses the information of (v_{k-1}, w_{k-1}), it is also named "weighted off-by-one IQC." Basically the off-by-one IQC is weighted by the rate ρ. Similarly, off-by-τ IQCs can be tailored as weighted off-by-τ IQCs.