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

Lecture 11

Incorporating Dynamics into IQCs

Lecturer: Bin Hu, Date:02/21/2019

In the last lecture, we talked about how to use advanced supply rate conditions to construct Lure-Postnikov Lyapunov functions for Nesterov's method and SAG. The constructions of these advanced supply rates require using the information of A and B. Therefore, we need to combine the information of G and Δ to construct such type of "coupling" constraints. In this lecture, we look at an alternative approach – the dynamic integral quadratic constraint (IQC) approach.

11.1 Incorporating Dynamics into IQCs

In Lecture 7, we briefly introduced the concept of IQCs. The example given in Lecture 7 has the following simple form:

$$\sum_{k=0}^{N} \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}^\mathsf{T} M \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix} \leq 0. \tag{11.1}$$

This type of IQCs does not involve dynamics and can be viewed as straightforward extensions of pointwise quadratic constraints. In this lecture, we discuss IQCs in more general forms. Specifically, we allow "dynamic" IQCs.

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

$$\begin{aligned} \psi_{k+1} &= A_\psi \psi_k + B_{\psi 1} v_k + B_{\psi 2} w_k \\ r_k &= C_\psi \psi_k + D_{\psi 1} v_k + D_{\psi 2} w_k \end{aligned} \tag{11.2}$$

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:

$$\begin{aligned} \psi^* &= A_\psi \psi^* + B_{\psi 1} v^* + B_{\psi 2} w^* \\ r^* &= C_\psi \psi^* + D_{\psi 1} v^* + D_{\psi 2} w^* \end{aligned} \tag{11.3}$$

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

$$\sum_{k=0}^{N} (r_k - r^*)^\mathsf{T} M (r_k - r^*) \leq 0 \tag{11.4}$$

where r is the output of the state-space model (11.2) with inputs (v, w) and an initial condition $\psi_0 = \psi^*$.

Many control papers use "$\geq$" in (11.4). Here we want to interpret (11.4) as a supply rate condition and hence use "$\leq$" instead.

Graphical interpretation. Time domain IQCs yield a graphical interpretation as shown in Figure 11.1. Let the input and output signals of Δ be filtered through Ψ with the initial condition $\psi_0 = \psi^*$. The IQC condition (11.4) just enforces a quadratic inequality on the filtered signal r.

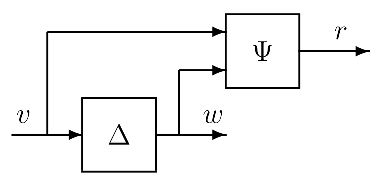

Figure 11.1. Graphical Interpretation for Time Domain IQCs

Relationship between static IQCs and general hard IQCs. If we ignore the dynamics part ($A_{\psi 1}$, $B_{\psi 1}$, $B_{\psi 2}$, and C_ψ are all gone), then we can set $D_{\psi 1} = \begin{bmatrix} I \\ 0 \end{bmatrix}$ and $D_{\psi 2} = \begin{bmatrix} 0 \\ I \end{bmatrix}$ to recover the static IQC (11.1) as a special case of (11.4). Equivalently, in Figure 11.1, the filter corresponding to a static IQC condition (11.1) is just a static identity mapping. As mentioned in the previous lectures, the tightness of the dissipation inequality analysis is determined by the approximating power of the quadratic constraints we use to describe Δ. By incorporating dynamics into IQCs, we introduce a larger family of quadratic constraints that can be used to approximate the relation $v = \Delta(w)$ better and tighter. This will lead to less conservative analysis results in many situations.

11.2 Zames-Falb IQCs

Various IQCs for many different perturbation operators have been developed in the robust control field. For our purposes of analyzing optimization methods, we focus on the IQCs developed for $\Delta = \nabla f$. We will look at one example in this section.

Suppose Δ maps v to w as $w_k = \nabla f(v_k)$ where f is a convex differentiable function. In Lecture 7, we talked about a static pointwise quadratic constraint (passivity) for such Δ:

$$\begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}^\mathsf{T} \begin{bmatrix} 0 & -I \\ -I & 0 \end{bmatrix} \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix} \leq 0, \forall k \tag{11.5}$$

Why is (11.5) **conservative?** The function f is not time-varying but the constraint (11.5) does not reflect this. Just imagine $w_k = \nabla f_k(v_k)$ where f_k ia 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 (11.5) still holds. Therefore, the constraint (11.5) does not exploit the fact that the function f is not changing over time.

How to fix the above issue? There is a large family of so-called Zames-Falb IQCs that do exploit the fact that f is a static function that does not change over time. Let's look at the simplest Zames-Falb IQC that is equivalent to the following inequality:

$$\sum_{k=0}^{N} w_k^{\mathsf{T}}(v_k - v_{k-1}) \geq 0, \ \forall N \tag{11.6}$$

where v_{-1} is defined to be v^* satisfying $\nabla f(v^*) = 0$.

Proof of the Zames-Falb IQC (11.6)**.** We use the fact $f(x) \geq f(y) + \nabla f(y)^{\mathsf{T}}(x-y)$ in a repeated manner. We can obtain

$$\begin{aligned}
f(v_{k-1}) &\geq f(v_k) + \nabla f(v_k)^{\mathsf{T}}(v_{k-1} - v_k) \\
f(v_{k-2}) &\geq f(v_{k-1}) + \nabla f(v_{k-1})^{\mathsf{T}}(v_{k-2} - v_{k-1}) \\
&\vdots \\
f(v^0) &\geq f(v_1) + \nabla f(v_1)^{\mathsf{T}}(v_0 - v_1) \\
f(v^*) &\geq f(v_0) + \nabla f(v_0)^{\mathsf{T}}(v^* - v_0)
\end{aligned}$$

Summing the above inequalities up leads to $\sum_{k=0}^{N} w_k^{\mathsf{T}}(v_k - v_{k-1}) \geq f(v_k) - f(v^*) \geq 0$.

How to rewrite (11.6) **in a filter form?** If we choose $r_k = \begin{bmatrix} v_k - v_{k-1} \\ w_k \end{bmatrix}$ and $M = \begin{bmatrix} 0 & -I \\ -I & 0 \end{bmatrix}$, then (11.4) and (11.6) are just the same. The question becomes how to generate $r_k = \begin{bmatrix} v_k - v_{k-1} \\ w_k \end{bmatrix}$. 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}, we need to use ψ_k to memorize v_{k-1}. We will have $\psi_k = v_{k-1}$ and hence $\psi_{k+1} = v_k$. Therefore, we have the following filter dynamics:

$$\begin{aligned}
\psi_{k+1} &= v_k \\
r_k &= \begin{bmatrix} -I \\ 0 \end{bmatrix} \psi_k + \begin{bmatrix} I \\ 0 \end{bmatrix} v_k + \begin{bmatrix} 0 \\ I \end{bmatrix} w_k
\end{aligned}$$

with the initial condition $\psi_0 = v_{-1} = v^*$. It is straightforward to verify that the fixed point of the filter is given by $\psi^* = v^*$ and $r^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Therefore, we can rewrite (11.6) as an IQC by choosing $A_\psi = 0$, $B_{\psi 1} = I$, $B_{\psi 2} = 0$, $C_\psi = \begin{bmatrix} -I \\ 0 \end{bmatrix}$, $D_{\psi 1} = \begin{bmatrix} I \\ 0 \end{bmatrix}$, and $D_{\psi 2} = \begin{bmatrix} 0 \\ I \end{bmatrix}$.

Other Zames-Falb IQCs. There are many Zames-Falb IQCs, and (11.6) is only one of them. Since (11.6) looks one step back and uses the information of v_{k-1}, it is also named "off-by-one IQC." Similarly, the off-by-two IQC will look two steps back and is equivalent to the inequality

$$\sum_{k=0}^{N} w_k^\mathsf{T}(v_k - v_{k-2}) \geq 0, \ \forall N \tag{11.7}$$

where $v_{-1} = v_{-2} = v^*$. Clearly, A_ψ for the off-by-two IQC is $(2p) \times (2p)$ where p is the dimension of v_k. Looking back for more steps leads to a higher dimension of A_ψ. Off-by-τ IQCs can be defined similarly. One can also use more general filter dynamics to define Zames-Falb IQCs and the details are omitted here. We will modify the off-by-one IQC for analysis of Nesterov's method in next lecture.

11.3 Dissipation Inequality with General IQCs

Recall that to use a static IQC (11.1) for analysis of a feedback interconnection $F_u(G, \Delta)$, one just replaces $v = \Delta(w)$ with the supply rate condition and then focuses on G that maps w to v. If there exists a positive definite matrix P such that

$$\begin{bmatrix} A^\mathsf{T}PA - \rho^2 P & A^\mathsf{T}PB \\ B^\mathsf{T}PA & B^\mathsf{T}PB \end{bmatrix} \leq \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix}^\mathsf{T} M \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix}$$

then one has $V(\xi_{k+1}) - V(\xi_k) \leq S(\xi_k, w_k)$ holds with $V(\xi_k) = (\xi_k - \xi^*)^\mathsf{T} P(\xi_k - \xi^*)$ and $S(\xi_k, w_k) = \begin{bmatrix} \xi_k - \xi^* \\ w_k - w^* \end{bmatrix}^\mathsf{T} \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix}^\mathsf{T} M \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \xi_k - \xi^* \\ w_k - w^* \end{bmatrix}$. Based on the supply rate condition $\sum_{k=0}^N S(\xi_k, w_k) \leq 0$, one has $V(\xi_{N+1}) \leq V(\xi_0)$ for all N.

Now let's look at how to modify this framework for general IQCs. The idea is quite similar. We just replace Δ with some quadratic constraints on (v, w). Now the dynamics of Ψ has to be taken into accounts. A graphical interpretation is shown in Figure 11.2. After replacing Δ with the IQC condition, the pair (v, w) still satisfies $v = G(w)$. In addition, let $r = \Psi(v, w) = \Psi(G(w), w)$. Then r must satisfy the constraint (11.4). Eventually we only need to analyze a composite system $\Psi(G(w), w)$ with input w and the output r. And we know r has to satisfy (11.4) as long as $\psi_0 = \psi^*$ and $v = \Delta(w)$. This is the main difference. For static IQCs, we only need to look at $G(w)$. For general IQCs, we need to look at $\Psi(G(w), w)$. Eventually we will write $\Psi(G(w), w)$ as a state space model

$$\begin{aligned} \eta_{k+1} &= \hat{A}\eta_k + \hat{B}w_k \\ r_k &= \hat{C}\eta_k + \hat{D}w_k \end{aligned} \tag{11.8}$$

which maps w to r, and the state ξ_k should be included as a part of η_k. If there exists a positive definite matrix P such that

$$\begin{bmatrix} \hat{A}^\mathsf{T} P \hat{A} - P & \hat{A}^\mathsf{T} P \hat{B} \\ \hat{B}^\mathsf{T} P \hat{A} & \hat{B}^\mathsf{T} P \hat{B} \end{bmatrix} \leq \begin{bmatrix} \hat{C} & \hat{D} \\ 0 & I \end{bmatrix}^\mathsf{T} M \begin{bmatrix} \hat{C} & \hat{D} \\ 0 & I \end{bmatrix} \tag{11.9}$$

then one has $V(\eta_{k+1}) - V(\eta_k) \leq S(\eta_k, w_k)$ holds with $V(\eta_k) = (\eta_k - \eta^*)^\mathsf{T} P (\eta_k - \eta^*)$ and $S(\eta_k, w_k) = r_k^\mathsf{T} M r_k$. Based on the IQC condition (11.4), one have $V(\eta_{N+1}) \leq V(\eta_0)\ \forall N$.

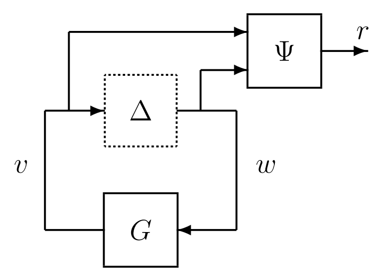

Figure 11.2. System G Extended to Include Filter Ψ

What is this composite system $\Psi(G(w), w)$? In the controls field, there are many formulas for manipulating state-space models. For illustrative purposes, let's derive a state-space model for $\Psi(G(w), w)$. Suppose G is LTI and governed by

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

Using the fact $v_k = C\xi_k$, we can show

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

Therefore, we can augment states and obtain the following model for $\Psi(G(w), w)$ whose input is w and output is r:

$$\begin{bmatrix} \xi_{k+1} \\ \psi_{k+1} \end{bmatrix} = \begin{bmatrix} A & 0 \\ B_{\psi 1} C & A_\psi \end{bmatrix} \begin{bmatrix} \xi_k \\ \psi_k \end{bmatrix} + \begin{bmatrix} B_{\psi 1} \\ B_{\psi 2} \end{bmatrix} w_k \tag{11.10}$$

$$r_k = \begin{bmatrix} D_{\psi 1} C & C_\psi \end{bmatrix} \begin{bmatrix} \xi_k \\ \psi_k \end{bmatrix} + D_{\psi 2} w_k \tag{11.11}$$

Now we can obtain the model (11.8) by choosing $\eta_k = \begin{bmatrix} \xi_k \\ \psi_k \end{bmatrix}$, $\hat{A} = \begin{bmatrix} A & 0 \\ B_{\psi 1} C & A_\psi \end{bmatrix}$, $\hat{B} = \begin{bmatrix} B_{\psi 1} \\ B_{\psi 2} \end{bmatrix}$, $\hat{C} = \begin{bmatrix} D_{\psi 1} C & C_\psi \end{bmatrix}$, and $\hat{D} = D_{\psi 2}$. Clearly, if we can show η_k is bounded for all k, then ξ_k is also bounded for all k.

Initial condition of Ψ. One has to initialize Ψ from $\psi_0 = \psi^*$ such that the IQC condition (11.4) holds. Therefore, the initial condition of the composite system $\Psi(G(w), w)$ is set up as $\eta_0 = \begin{bmatrix} \xi_0 \\ \psi_0 \end{bmatrix} = \begin{bmatrix} \xi_0 \\ \psi^* \end{bmatrix}$. This is OK since we can initialize ξ_0 from any arbitrary point. The initial condition of $F_u(G, \Delta)$ is typically only embedded in ξ_0.

The separation property. Notice the IQC condition (11.4) is just a property of Δ. The construction of (11.4) is completely independent of G. Therefore, the IQC framework gives a better way to study G and Δ separately. However, since Ψ will increase the state dimension of the overall system, the size of the resultant LMI will also be larger and this may cause trouble when the goal is to analytically solve the LMI.

IQCs for convergence rate analysis. In the next lecture, we will discuss how to modify Zames-Falb IQCs for convergence rate analysis.