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

Lecture 8

Supply Rate Constructions and Quadratic Constraints, Part II

Lecturer: Bin Hu, Date:02/12/2019

In the last lecture, we have talked about a few commonly-used quadratic constraints including small gain bound, passivity, and sector bound. In today's lecture, let's look at several tricks for selecting and manipulating quadratic constraints.

8.1 Redundancy in Quadratic Constraints

It is OK to allow some redundancy when choosing the quadratic constraints. We will illustrate this by an example. Recall that the sector bound condition gives the following quadratic inequality with $L \geq m$:

$$\begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}^\mathsf{T} \begin{bmatrix} 2mLI & -(m+L)I \\ -(m+L)I & 2I \end{bmatrix} \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix} \leq 0. \tag{8.1}$$

First, we discuss the connections between sector bound and other conditions.

1. If we let $m = 0$, we obtain the constraint

$$\begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}^\mathsf{T} \begin{bmatrix} 0 & -LI \\ -LI & 2I \end{bmatrix} \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix} \leq 0. \tag{8.2}$$

2. If we let $L \to \infty$, we have $\frac{m}{L} \to 0$ and (8.1) reduces to

$$\begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}^\mathsf{T} \begin{bmatrix} 2mI & -I \\ -I & 0 \end{bmatrix} \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix} \leq 0. \tag{8.3}$$

3. If we let $L \to \infty$ and $m = 0$, we recover the passivity condition

$$\begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}^\mathsf{T} \begin{bmatrix} 0 & -I \\ -I & 0 \end{bmatrix} \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix} \leq 0. \tag{8.4}$$

4. If we let $m = -L$, (8.1) reduces to the small gain bound:

$$\begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}^\mathsf{T} \begin{bmatrix} -L^2 I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix} \leq 0. \tag{8.5}$$

An important fact. Given two symmetric matrices X_1 and X_2, if we can find $\lambda \geq 0$ such that $X_2 \leq \lambda X_1$, then the quadratic constraint $\begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}^\mathsf{T} X_1 \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix} \leq 0$ will directly guarantee the other constraint $\begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}^\mathsf{T} X_2 \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix} \leq 0$. This is one version of the famous S**-procedure**. Based on this procedure, if (v, w) satisfies the sector bound (8.1) with $L \geq m$, then (v, w) will also satisfy the bound

$$\begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}^\mathsf{T} \begin{bmatrix} 2m_0 L_0 I & -(m_0 + L_0)I \\ -(m_0 + L_0)I & 2I \end{bmatrix} \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix} \leq 0. \tag{8.6}$$

for any $m_0 \leq m$ and $L_0 \geq L$. To prove this, we use the key relationship

$$\begin{bmatrix} 2mL_0 I & -(m + L_0)I \\ -(m + L_0)I & 2I \end{bmatrix} = \frac{L_0 - m}{L - m} \begin{bmatrix} 2mLI & -(m + L)I \\ -(m + L)I & 2I \end{bmatrix} - \frac{L_0 - L}{L - m} \begin{bmatrix} 2m^2 I & -2mI \\ -2mI & 2I \end{bmatrix}$$

Therefore, if (8.1) holds and $L_0 \geq L \geq m$, we have

$$\begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}^\mathsf{T} \begin{bmatrix} 2mL_0 I & -(m + L_0)I \\ -(m + L_0)I & 2I \end{bmatrix} \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix} \leq 0.$$

We apply the same trick again and will get (8.6).

A consequence of the above fact is that if we know $L \geq m \geq 0$, the sector bound (8.1) directly guarantees other constraints including (8.2), (8.3), (8.4), and (8.5). This leads to an important question: **Which quadratic constraint shall we use to construct the dissipation inequality?** Intuitively, (8.1) is the general one and should be used. However, a combined use of (8.2), (8.3), and (8.4) may actually simplify the convergence rate proofs. It seems that (8.4) is a redundant constraint here, but sometimes adding this constraint can simplify the **analytical** proof.

Let's look at the analysis of the gradient method again. Suppose f is L-smooth and m-strongly convex. If we use (8.1), the resultant LMI condition is

$$\begin{bmatrix} 1 - \rho^2 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix} - \lambda_1 \begin{bmatrix} 2mL & -(m + L) \\ -(m + L) & 2 \end{bmatrix} \leq 0 \tag{8.7}$$

where $\lambda_1 \geq 0$ is the only decision variable. If we combine (8.2), (8.3), and (8.4), the LMI condition becomes

$$\begin{bmatrix} 1 - \rho^2 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix} - \left(\lambda_1 \begin{bmatrix} 0 & -L \\ -L & 2 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2m & 1 \\ 1 & 0 \end{bmatrix} + \lambda_3 \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \right) \leq 0 \tag{8.8}$$

where non-negative scalers $(\lambda_1, \lambda_2, \lambda_3)$ are all decision variables. In (8.7), we only have one decision variable λ_1. It is more difficult to figure out which negative semidefinite matrix we should set the left side of (8.7) to. On the other hand, (8.8) has three variables and actually we can set the left side of (8.8) to be a diagonal matrix whenever $\alpha \leq \frac{1}{L}$. Given any α and $\rho^2 = (1-m\alpha)^2$, we just set $\lambda_1 = \alpha^2$, $\lambda_2 = \alpha$, and $\lambda_3 = \alpha - L\alpha^2$, and the left side of (8.8) just becomes $\begin{bmatrix} -m^2\alpha^2 & 0 \\ 0 & -\alpha^2 \end{bmatrix} \leq 0$. This result can also be obtained by solving (8.7). However, how to set up the left side of (8.7) is a little bit trickier.

Key message. Notice λ_3 in (8.8) is just set up to cancel the off-diagonal terms of the resultant 2×2 matrix. We can clearly see that adding the redundant constraint (8.4) just helps us to cancel the off-diagonal terms and simplify the proof a little bit.

8.2 The Feedback Representation is Not Unique!

The feedback representation for an optimization method is not unique. Different feedback formulations lead to different LMIs that require different supply rates. Some feedback representations may yield simpler convergence proofs than the others. We will use the gradient method as an example to illustrate this point.

In the previous lectures, we modeled the gradient method as $F_u(G, \Delta)$ where $\Delta = \nabla f$, and G is governed by an LTI model with $(A, B, C) = (I, -\alpha I, I)$. The matrix X in the supply rate is $\begin{bmatrix} 2mLI & -(m+L)I \\ -(m+L)I & 2I \end{bmatrix}$, and the resultant LMI condition is (8.7). We have to handle the non-zero off-diagonal term when choosing λ.

Alternatively, we can model the gradient method as the following feedback model:

$$\begin{aligned} \xi_{k+1} &= w_k \\ v_k &= \xi_k \\ w_k &= v_k - \alpha\nabla(v_k) \end{aligned}$$

In this case, G is described by an LTI model with $A = 0$, $B = I$, and $C = I$. The perturbation operator Δ maps v to w as $w_k = v_k - \alpha\nabla(v_k)$. Since $A = 0$, we have $A^\mathsf{T}PB = B^\mathsf{T}PA = 0$. Therefore, we can formulate the following new LMI condition:

$$\begin{bmatrix} -\rho^2 & 0 \\ 0 & 1 \end{bmatrix} \leq \lambda X \tag{8.9}$$

where X is a 2×2 symmetric matrix such that

$$\begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}^\mathsf{T} (X \otimes I) \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix} \leq 0. \tag{8.10}$$

How to obtain X from existing quadratic constraints on ∇f? When f is L-smooth and m-strongly convex, we know the following quadratic constraint holds

$$\begin{bmatrix} v_k - v^* \\ \nabla f(v_k) - \nabla f(v^*) \end{bmatrix}^\mathsf{T} \begin{bmatrix} 2mLI & -(m+L)I \\ -(m+L)I & 2I \end{bmatrix} \begin{bmatrix} v_k - v^* \\ \nabla f(v_k) - \nabla f(v^*) \end{bmatrix} \leq 0. \tag{8.11}$$

However, now we have $w_k = v_k - \alpha\nabla f(v_k)$. Can we just manipulate the above quadratic constraint to describe the relationship between v and w? The answer is yes! Just notice $\nabla f(v_k) = (v_k - w_k)/\alpha$ (this is equivalent to $w_k = v_k - \alpha\nabla f(v_k)$). Therefore, there is a linear mapping from $\begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}$ to $\begin{bmatrix} v_k - v^* \\ \nabla f(v_k) - \nabla f(v^*) \end{bmatrix}$:

$$\begin{bmatrix} v_k - v^* \\ \nabla f(v_k) - \nabla f(v^*) \end{bmatrix} = \begin{bmatrix} I & 0 \\ \frac{1}{\alpha} I & -\frac{1}{\alpha} I \end{bmatrix} \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix} \tag{8.12}$$

where $w^* = v^* = x^*$. All we need to do is to substitute the above equation into (8.11) and obtain X as

$$\begin{aligned} X &= \begin{bmatrix} I & 0 \\ \frac{1}{\alpha} I & -\frac{1}{\alpha} I \end{bmatrix}^\mathsf{T} \begin{bmatrix} 2mLI & -(m+L)I \\ -(m+L)I & 2I \end{bmatrix} \begin{bmatrix} I & 0 \\ \frac{1}{\alpha} I & -\frac{1}{\alpha} I \end{bmatrix} \\ &= \frac{1}{\alpha^2} \begin{bmatrix} 2(L\alpha - 1)(m\alpha - 1) & (m+L)\alpha - 2 \\ (m+L)\alpha - 2 & 2 \end{bmatrix} \end{aligned}$$

Consequently, the LMI (8.9) becomes

$$\begin{bmatrix} -\rho^2 & 0 \\ 0 & 1 \end{bmatrix} \le \frac{\lambda}{\alpha^2} \begin{bmatrix} 2(L\alpha - 1)(m\alpha - 1) & (m+L)\alpha - 2 \\ (m+L)\alpha - 2 & 2 \end{bmatrix} \tag{8.13}$$

This LMI leads to simpler convergence rate proofs of the gradient method for the following two stepsize choices.

- Case 1: For $\alpha = \frac{2}{m+L}$, the off-diagonal term in (8.13) just becomes 0, and we only need to look at the diagonal terms. Setting $\lambda = \frac{\alpha^2}{2}$ leads to $\rho = \frac{L-m}{L+m}$.
- Case 2: For $\alpha = \frac{1}{L}$, the LMI condition becomes

  $$\begin{bmatrix} -\rho^2 & 0 \\ 0 & 1 \end{bmatrix} \le L^2\lambda \begin{bmatrix} 0 & \frac{m}{L} - 1 \\ \frac{m}{L} - 1 & 2 \end{bmatrix}$$

  We can simply choose $\lambda = \frac{1}{L^2}$ and $\rho = 1 - \frac{m}{L}$ to satisfy the above LMI. Although we have non-zero off-diagonal terms here, the first entry of the LMI depends on ρ^2 and is independent of λ. This makes the analytical proof simpler.

Key message. From the above example, we can see that the feedback representations for an optimization method are not unique and some of them may lead to simpler convergence rate proofs. Although the feedback representation can be different, one can still obtain quadratic constraints for the new Δ by manipulating known quadratic constraints.

8.3 Manipulating IQCs via Linear Mapping

The example in the last section actually demonstrates an important trick. Suppose we have some IQC (notice IQCs are more general than pointwise quadratic constraints) to couple h_k and u_k, i.e.

$$\sum_{k=0}^{N} \begin{bmatrix} h_k - h^* \\ u_k - u^* \end{bmatrix}^\mathsf{T} M \begin{bmatrix} h_k - h^* \\ u_k - u^* \end{bmatrix} \leq 0. \tag{8.14}$$

If we have the following linear mapping

$$\begin{bmatrix} h_k - h^* \\ u_k - u^* \end{bmatrix} = H \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}, \tag{8.15}$$

then we can immediately obtain an IQC for v and w:

$$\sum_{k=0}^{N} \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix}^\mathsf{T} (H^\mathsf{T} M H) \begin{bmatrix} v_k - v^* \\ w_k - w^* \end{bmatrix} \leq 0. \tag{8.16}$$