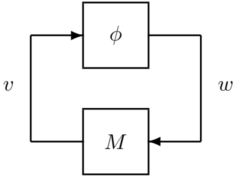
Review: Basic Ideas of Quadratic Constraints



- **Goal:** Analyze the following set of coupled sequences $\{\xi(k), w(k), v(k)\}$

$$\{(\xi, w, v) : \xi(k+1) = A\xi(k) + Bw(k),\ v(k) = C\xi(k)\} \cap \{(\xi, w, v) : w(k) = \phi(v(k))\}$$

- **Key idea: Quadratic constraints!** Replace the graph of ϕ with a relation on (v, w) captured by a quadratic constraint:

$$\{(v, w) : w(k) = \phi(v(k))\} \subset \left\{(v, w) : \begin{bmatrix} v(k) \\ w(k) \end{bmatrix}^\mathsf{T} \Lambda \begin{bmatrix} v(k) \\ w(k) \end{bmatrix} \geq 0 \right\},$$

  where Λ is constructed from the property of ϕ.

- We only need to analyze the stability of the following set:

$$\left\{(\xi, w, v) : \xi(k+1) = A\xi(k) + Bw(k),\ v(k) = C\xi(k), \begin{bmatrix} v(k) \\ w(k) \end{bmatrix}^\mathsf{T} \Lambda \begin{bmatrix} v(k) \\ w(k) \end{bmatrix} \geq 0 \right\}.$$

Review: Basic Ideas of Quadratic Constraints

Now we are analyzing the sequence from the following set:

$$\left\{(\xi, w, v) : \xi(k+1) = A\xi(k) + Bw(k),\ v(k) = C\xi(k),\ \begin{bmatrix} v(k) \\ w(k) \end{bmatrix}^\mathsf{T} \Lambda \begin{bmatrix} v(k) \\ w(k) \end{bmatrix} \geq 0 \right\}.$$

Theorem

If there exists a positive definite matrix P and $0 < \rho < 1$ s.t.

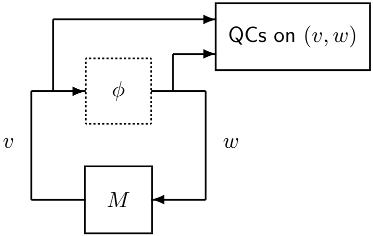
$$\begin{bmatrix} A^\mathsf{T} P A - \rho^2 P & A^\mathsf{T} P B \\ B^\mathsf{T} P A & B^\mathsf{T} P B \end{bmatrix} \preceq - \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix}^\mathsf{T} \Lambda \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix}$$

then $\xi(k+1)^\mathsf{T} P \xi(k+1) \leq \rho^2 \xi(k)^\mathsf{T} P \xi(k)$ *and* $\lim_{k\to\infty} \xi(k) = 0$.

$$\underbrace{\begin{bmatrix} \xi(k) \\ w(k) \end{bmatrix}^\mathsf{T} \begin{bmatrix} A^\mathsf{T} P A - \rho^2 P & A^\mathsf{T} P B \\ B^\mathsf{T} P A & B^\mathsf{T} P B \end{bmatrix} \begin{bmatrix} \xi(k) \\ w(k) \end{bmatrix}}_{\xi(k+1)^\mathsf{T} P \xi(k+1) - \rho^2 \xi(k)^\mathsf{T} P \xi(k)} \leq \underbrace{- \begin{bmatrix} \xi(k) \\ w(k) \end{bmatrix}^\mathsf{T} \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix}^\mathsf{T} \Lambda \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \xi(k) \\ w(k) \end{bmatrix}}_{-\begin{bmatrix} v(k) \\ w(k) \end{bmatrix}^\mathsf{T} \Lambda \begin{bmatrix} v(k) \\ w(k) \end{bmatrix} \leq 0}$$

This condition is a semidefinite program (SDP) problem!

Conservatism in QCs for ReLU Networks



- To reduce conservatism, replace troublesome ϕ with multiple QCs:

$$\{(v,w) : w(k) = \phi(v(k))\} \subset \bigcap_{\Lambda \in \mathcal{M}} \left\{ (v,w) : \begin{bmatrix} v(k) \\ w(k) \end{bmatrix}^{\mathsf{T}} \Lambda \begin{bmatrix} v(k) \\ w(k) \end{bmatrix} \geq 0 \right\},$$

- Issue: If the set on the right side is much **larger** than the set on the left side, then the analysis can become conservative!
- Finding $\mathcal{M}$ is not an easy task. The controls community has made huge efforts in this.

Illustrative Example: Gradient Descent Method

- Rewrite the gradient method $x_{k+1} = x_k - \alpha \nabla f(x_k)$ as:

$$\underbrace{(x_{k+1} - x^\star)}_{\xi_{k+1}} = \underbrace{(x_k - x^\star)}_{\xi_k} - \alpha \underbrace{\nabla f(x_k)}_{w_k}$$

- If f is L-smooth and m-strongly convex, then by co-coercivity:

$$\begin{bmatrix} x - x^\star \\ \nabla f(x) \end{bmatrix}^\mathsf{T} \underbrace{\begin{bmatrix} 2mLI & -(m+L)I \\ -(m+L)I & 2I \end{bmatrix}}_{\Lambda} \begin{bmatrix} x - x^\star \\ \nabla f(x) \end{bmatrix} \leq 0$$

- We have $A = I$, $B = -\alpha I$, $C = I$, and the following SDP

$$\begin{bmatrix} A^\mathsf{T} P A - \rho^2 P & A^\mathsf{T} P B \\ B^\mathsf{T} P A & B^\mathsf{T} P B \end{bmatrix} \preceq \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix}^\mathsf{T} \Lambda \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix}$$

- This leads to $\left(\begin{bmatrix} (1-\rho^2)p & -\alpha p \\ -\alpha p & \alpha^2 p \end{bmatrix} + \begin{bmatrix} -2mL & m+L \\ m+L & -2 \end{bmatrix} \right) \otimes I \preceq 0$
- Choose (α, ρ, p) to be $(\frac{1}{L}, 1 - \frac{m}{L}, L^2)$ or $(\frac{2}{L+m}, \frac{L-m}{L+m}, \frac{1}{2}(L+m)^2)$ to recover standard rates, i.e. $\|x_{k+1} - x^\star\| \leq (1 - m/L)\|x_k - x^\star\|$
- For this proof, is strong convexity really needed? No! Regularity condition!

Illustrative Example: Gradient Descent Method

- We have shown $\|x_{k+1} - x^\star\| \le (1 - m/L)\|x_k - x^\star\|$
- Is it a contraction, i.e. $\|x_{k+1} - x'_{k+1}\| \le (1 - m/L)\|x_k - x'_k\|$?
- $\underbrace{(x_{k+1} - x'_{k+1})}_{\xi_{k+1}} = \underbrace{(x_k - x'_k)}_{\xi_k} - \alpha \underbrace{(\nabla f(x_k) - \nabla f(x'_k))}_{w_k}$
- If f is L-smooth and m-strongly convex, then by co-coercivity:

$$\begin{bmatrix} x - x' \\ \nabla f(x) - \nabla f(x') \end{bmatrix}^\mathsf{T} \underbrace{\begin{bmatrix} 2mLI & -(m+L)I \\ -(m+L)I & 2I \end{bmatrix}}_{\Lambda} \begin{bmatrix} x - x' \\ \nabla f(x) - \nabla f(x') \end{bmatrix} \le 0$$

- We have $A = I$, $B = -\alpha I$, $C = I$, and the same SDP

$$\left(\begin{bmatrix} (1-\rho^2)p & -\alpha p \\ -\alpha p & \alpha^2 p \end{bmatrix} + \begin{bmatrix} -2mL & m+L \\ m+L & -2 \end{bmatrix} \right) \otimes I \preceq 0$$

- Choose (α, ρ, p) to be $(\frac{1}{L}, 1 - \frac{m}{L}, L^2)$ or $(\frac{2}{L+m}, \frac{L-m}{L+m}, \frac{1}{2}(L+m)^2)$ to give the contraction result!
- For this proof, is strong convexity really needed? Yes!