Decentralized Stabilization for a Class of Continuous-Time Nonlinear Interconnected Systems Using Online Learning Optimal Control Approach

Decentralized Stabilization for a Class of Continuous-Time Nonlinear Interconnected Systems Using Online Learning Optimal Control Approach (http://blogs.cuit.columbia.edu/zp2130/files/2019/02/Decentralized-Stabilization-for-a-Class-of-Continuous-Time-Nonlinear-Interconnected-Systems-Using-Online-Learning-Optimal-Control-Approach.pdf)

Neural-network-based Online Learning Optimal Control

Decentralized Control Strategy

1. **Cost functions (critic neural networks) – local optimal controllers**
2. **Feedback gains to the optimal control policies – decentralized control strategy**

Optimal Control Problem (Stabilization)

Hamilton-Jacobi-Bellman (HJB) Equations

- **Apply Online Policy Iteration Algorithm (construct and train critic neural networks) to solve HJB Equations.**

The **decentralized control** has been a control of choice for large-scale systems because it is computationally efficient to formulate control law that use only **locally available subsystem states** or outputs.

Though **dynamic programming** is a useful technique to solve the optimization and optimal control problems, in may cases, it is computationally difficult to apply it because of the **curse of dimensionality**.

Considering the effectiveness of ADP and **reinforcement learning** techniques in solving the **nonlinear optimal control problem**, the **decentralized control approach** established is natural and convenient.

Notation

$i = 1, 2, ..., N.$: ith subsystem.

$x_i(t) \in \mathbb{R}^{n_i}$: **state** vector of the ***i***th subsystem.

$x_1, x_2, ..., x_N$: **local states**.

$\bar{u}_i(x_i(t)) \in \mathbb{R}^{m_i}$: **control** vector of the ith subsystem.

$\bar{u}_1(x_1), \bar{u}_2(x_2), ..., \bar{u}_N(x_N)$: **local controls**.

$u_i(x_i), i = 1, 2, ..., N$: **control policies**.

$f_i(x_i)$: **nonlinear internal dynamics**.

$g_i(x_i)$: **input gain matrix**.

$g_i(x_i)\bar{Z}_i(x)$: **interconnected term.** ***$Z_i(x)$***'s ***x*** has no ***i .***

$R_i \in \mathbb{R}^{m_i \times m_i}, i = 1, 2, ..., N.$: **symmetric positive definite matrices**.

ρ : nonnegative constants.

$h_{ij}(x_j)$: **positive semidefinite function.**

$Q_i(x_i), i = 1, 2, ..., N.$: **positive definite functions** satisfying $h_i(x_i) \le Q_i(x_i), i = 1, 2, ..., N.$

$\mu_i(x_i)$: **control policy**.

Ω_i : $f_i + g_i u_i$ is Lipshcitz continuous on a set Ω_i in $\mathbb{R}^{n_i}$ containing the origin, and the subsystem is controllable in the sense that there exists a continuous control policy on Ω_i that asymptotically stabilizes the subsystem.

Decentralized Control Problem of the Large-Scale System

Paper studies a class of continuous-time nonlinear large-scale systems: composed of **N interconnected subsystems** described by

$$\dot{x}_i(t) = f_i(x_i(t)) + g_i(x_i(t))\left(\bar{u}_i(x_i(t)) + \bar{Z}_i(x(t))\right) \quad (1)$$
$$i = 1, 2, ..., N$$

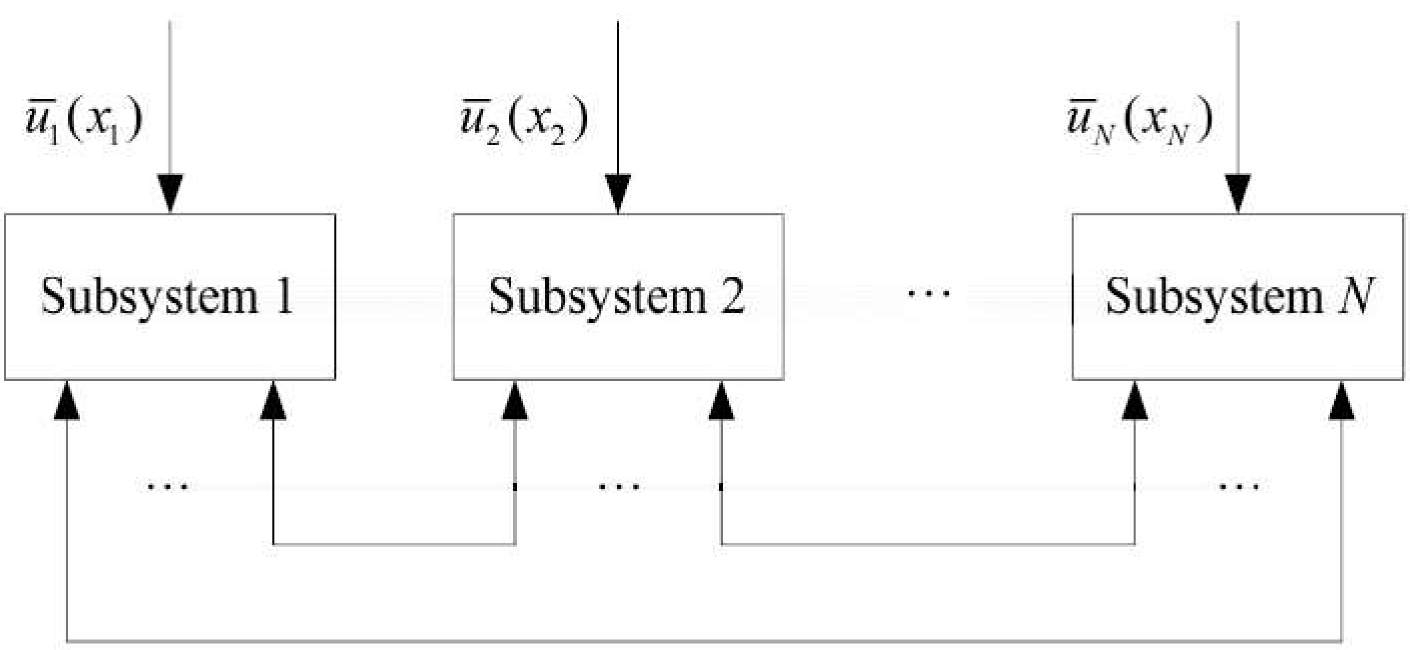

Fig. 1. Structural diagram of the decentralized control problem of the interconnected system.

$x_i(0) = x_{i0}$: initial state of the ith subsystem,

Assumption 1: When $x_i = 0$, ith subsystem is **equilibrium**.

Assumption 2: $f_i(x_i)$ and $g_i(x_i)$ are differentiable in arguments with $f_i(0) = 0$.

Assumption 3: When $x_i = 0$, the **feedback control** vector $\bar{u}_i(x_i) = 0$.

$$Z_i(x) = R_i^{1/2}\bar{Z}_i(x)$$

where

$R_i \in \mathbb{R}^{m_i \times m_i}, i = 1, 2, ..., N.$: **symmetric positive definite matrices**.

$$Z_i(x) \in \mathbb{R}^{m_i}, i = 1, 2, ..., N.$$

are bounded as follows:

$$\|Z_i(x)\| \leq \sum_{j=1}^{N} \rho_{ij} h_{ij}(x_j), \quad (2)$$
$$i = 1, 2, ..., N.$$

Define

$$h_i(x_i) = max\{h_{1i}(x_i), h_{2i}(x_i), ..., h_{Ni}(x_i)\}$$

then (2) can be formulated as

$$\|Z_i(x)\| \leq \sum_{j=1}^{N} \lambda_{ij} h_j(x_j), \ \ i = 1, 2, ..., N.$$
$$where$$
$$\lambda_{ij} \geq \frac{\rho_{ij} h_{ij}(x_j)}{h_j(x_j)}$$

C1 – Optimal Control of Isolated Subsystems (Framework of HJB Equations)

C2 – Decentralized Control Strategy

Consider the N isolated subsystems corresponding to (1)

$$\dot{x}_i(t) = f_i(x_i(t)) + g_i(x_i(t))(u_i(x_i(t))) \quad (4)$$
$$i = 1, 2, ..., N$$

Find the **control policies** $u_i(x_i), i = 1, 2, ..., N$ which **minimize** the **local cost functions**

$$J_i(x_{i0}) = \int_0^\infty \left\{ Q_i^2(x_i(\tau)) + u_i^T(x_i(\tau)) R_i u_i(x_i(\tau)) \right\} d\tau \quad (5)$$
$$i = 1, 2, ..., N$$

(**How to get the equation 5 ?** Should ***Q*** = ***Q*** and ***R*** = ***P***, (***Q*** and ***P*** ∈ **Lyapunov Equation**) **?**)

to deal with the infinite horizon **optimal control problem**.

where

$Q_i(x_i), i = 1, 2, ..., N.$: **positive definite functions** satisfying

$$h_i(x_i) \leq Q_i(x_i), i = 1, 2, ..., N. \quad (6)$$

Based on optimal control theory, feedback controls (**control policies**) must be **admissible** , i.e., stabilize the subsystmes on Ω_i , **guarantee cost function (5) are finite**.

Admissible Control

Definition 1

Consider the isolated subsystem i,

$$\mu_i \in \Psi_i(\Omega_i)$$
$$\mu_i(0) = 0$$
$$u_i(x_i) = \mu_i(x_i)$$

For any set of admissible **control policies** $\mu_i \in \Psi_i(\Omega_i), i = 1, 2, ..., N,$ if the associated **cost functions**

$$V_i(x_{i0}) = \int_0^\infty \left\{ Q_i^2(x_i(\tau)) + \mu_i^T(x_i(\tau)) R_i \mu_i(x_i(\tau)) \right\} d\tau$$
$$i = 1, 2, ..., N.$$

(7)

are continuously differentiable, then the infinitesimal versions of (7) are the so-called **nonlinear Lyapunov equations**

$$0 = Q_i^2(x_i) + \mu_i^T(x_i) R_i \mu_i(x_i) + (\bigtriangledown V_i(x_i))^T (f_i(x_i)) + g_i(x_i)\mu_i(x_i)) \quad (8)$$

(**How to get the equation 8 ?** Should ***Q*** = ***Q*** and ***R*** = ***P***, (***Q*** and ***P*** ∈ **Lyapunov Equation**) **?**)

where

$$V_i(0) = 0$$

$$\nabla V_i(x_i) = \frac{\partial V_i(x_i)}{\partial x_i}$$

$$i = 1, 2, ..., N.$$

Lyapunov Equation

Linear Quadratic Lyapunov Theory (https://blogs.cuit.columbia.edu/zp2130/files/2019/02/lq-lyap.pdf)

Linear Quadratic Lyapunov Theory Notes (http://blogs.cuit.columbia.edu/zp2130/files/2019/02/lq-lyap-notes.pdf)

Lyapunov Equation

We assume $A \in \mathbb{R}^{n \times n}, P = P^T \in \mathbb{R}^{n \times n}$. It follows that $Q = Q^T \in \mathbb{R}^{n \times n}$.
Continuous-time linear systems: $for\ \dot{x} = Ax, V(z) = z^T Pz, we\ have\ \dot{V}(z) = -z^T Qz$ where ***P, Q*** satisfy (continuous-time) **Lyapunov Equation:** $A^T P + PA + Q = 0$
If ***P>0, Q>0,*** then system is (globally asymptotically) **stable.**
If ***P>0, Q≥0,*** and (***Q,A***) **observable,** then system is (globally asymptotically) **stable.**

$$A^T P + PA + Q = 0$$

where ***A***, ***P***, ***Q*** ∈ $R^{n \times n}$, and ***P***, ***Q*** are **symmetric**

interpretation: for linear system

$$\dot{x} = Ax$$

if

$$V(z) = z^T Pz$$

$V(z) = z^T Pz$

then

$$\dot{V}(z) = (Az)^T Pz + z^T P(Az) = -z^T Qz$$

$\dot{V}(z) = (Az)^T Pz + z^T P(Az) = -z^T Qz$

i.e., if $z^T Pz$ is the (generalized) ***energy***, then $z^T Qz$ is the associated (generalized) ***dissipation***

Lyapunov Integral

If ***A*** is **stable** there is an explicit formula for **solution** of **Lyapunov equation**:

$$P = \int_0^\infty e^{tA^T} Q e^{tA} dt$$

to see this, we note that

$$\begin{aligned} A^T P + PA &= \int_0^\infty \left(A^T e^{tA^T} Q e^{tA} + e^{tA^T} Q e^{tA} A \right) \\ &= \int_0^\infty \left(\frac{\mathrm{d}}{\mathrm{d}t} e^{tA^T} Q e^{tA} \right) dt \\ &= e^{tA^T} Q e^{tA} \mid_0^\infty \\ &= -Q \end{aligned}$$

Interpretation as cost-to-go

If ***A*** is **stable,** and P is (unique) solution of

$$A^T P + PA + Q = 0$$

, then

Invalid Equation

thus ***V(z)*** is **cost-to-go from point z (with no input)** and **integral quadratic cost function** with **matrix Q**

If *A* is **stable** and Q>0, then for each t, $e^{tA^T} Q e^{tA} > 0$, so

$$P = \int_0^\infty e^{tA^T} Q e^{tA} dt > 0$$

meaning: if ***A*** is **stable,**

- we can choose ***any* positive definite** quadratic form z^TQz as the dissipation, i.e., $-\dot{V} = z^TQz$
- then solve a set of linear equations to find the (unique) quadratic form $\dot{V} = z^TQz$
- ***V*** will be positive definite, so it is a **Lyapunov function** that proves ***A*** is **stable**.

```
In particular: a linear system is stable if an only if there is a quadratic Layapunov function that proves it.
```

Evaluating Quadratic Integrals

Suppose $\dot{x} = Ax$ is **stable**, and define

$$J = \int_0^\infty x(t)^T Q x(t) dt$$

to find ***J***, we solve **Lyapunov equation**

$$A^TP + PA + Q = 0$$

for ***P*** then,

$$J = x(0)^T P x(0)$$

```
In other words: we can evaluate quadratic integral exactly, by solving a set of linear equations, without even computing
a matrix exponential.
```

Online Policy Iteration Algorithm (Critic Networks)

Solve HJB Equations

edit (https://blogs.cuit.columbia.edu/zp2130/wp-admin/post.php?post=4809&action=edit)

Author: Z Pei (https://blogs.cuit.columbia.edu/zp2130/author/zp2130/) on February 27, 2019

Categories: AI (https://blogs.cuit.columbia.edu/zp2130/category/ai/), Control (https://blogs.cuit.columbia.edu/zp2130/category/control/), Reinforcement Learning (https://blogs.cuit.columbia.edu/zp2130/category/reinforcement-learning/), RL (https://blogs.cuit.columbia.edu/zp2130/category/rl/)

Tags: AI (https://blogs.cuit.columbia.edu/zp2130/tag/ai/), Control (https://blogs.cuit.columbia.edu/zp2130/tag/control/), Reinforcement Learning (https://blogs.cuit.columbia.edu/zp2130/tag/reinforcement-learning/), RL (https://blogs.cuit.columbia.edu/zp2130/tag/rl/)

Other posts

Hierarchical Policy Gradient Algorithms (https://blogs.cuit.columbia.edu/zp2130/hierarchical_policy_gradient_algorithms/) «» Reinforcement Learning is Direct Adaptive Optimal Control (https://blogs.cuit.columbia.edu/zp2130/reinforcement_learning_is_direct_adaptive_optimal_control/)

Last posts

- Symbolic Netlist to Innovus-friendly Netlist (https://blogs.cuit.columbia.edu/zp2130/symbolic_netlist_to_innovus-friendly_netlist/)
- Finite-Sample Convergence Rates for Q-Learning and Indirect Algorithms (https://blogs.cuit.columbia.edu/zp2130/finite-sample_convergence_rates_for_q-learning_and_indirect_algorithms/)
- Solving H-horizon, Stationary Markov Decision Problems In Time Proportional To Log(H) (https://blogs.cuit.columbia.edu/zp2130/paul_tseng_1990/)
- Randomized Linear Programming Solves the Discounted Markov Decision Problem In Nearly-Linear (Sometimes Sublinear) Run Time (https://blogs.cuit.columbia.edu/zp2130/randomized_linear_programming_solves_the_discounted_markov_decision_problem_in_nearly-linear_sometimes_sublinear_run_time/)
- KL Divergence (https://blogs.cuit.columbia.edu/zp2130/kl_divergence/)
- The Asymptotic Convergence-Rate of Q-learning (https://blogs.cuit.columbia.edu/zp2130/the_asymptotic_convergence-rate_of_q-learning/)
- Hierarchical Apprenticeship Learning, with Application to Quadruped Locomotion (https://blogs.cuit.columbia.edu/zp2130/hierarchical_apprenticeship_learning_with_application_to_quadruped_locomotion/)
- Policy Gradient Methods (https://blogs.cuit.columbia.edu/zp2130/policy_gradient_methods/)
- Actor-Critic Algorithms for Hierarchical Markov Decision Processes (https://blogs.cuit.columbia.edu/zp2130/actor-critic_algorithms_for_hierarchical_markov_decision_processes/)

- Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation (https://blogs.cuit.columbia.edu/zp2130/hierarchical_deep_reinforcement_learning_integrating_temporal_abstraction_and_intrinsic_motivation/)