Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions

Setup
Assumptions
Optimal Gain

Model-based PO
Outer loop
Stabilization and Convergence

Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

On the Robustness and Convergence of Policy Optimization in Continuous-Time Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ Stochastic Control

Lekan Molu

Microsoft Research

New York City, NY 10012

Presented by **Lekan Molu** (Lay-con Mo-lu)

April 8, 2025

Talk Outline and Overview

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview

Risk-sensitive control

Contributions

Setup

Assumptions

Optimal Gain

Model-based PO

Outer loop

Stabilization and Convergence

Sampling-based PO

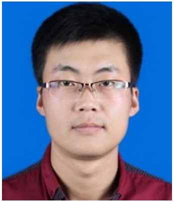
Discrete-time system

Sampling-based nonlinear system

- Policy Optimization and Stochastic Linear Control
  - Connections to risk-sensitive control;
  - Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control theory.
- The case for convergence analysis in stochastic PO.
  - Kleinman's algorithm, *redux.*
  - Kleiman's algorithm in an iterative best response setting;
  - PO Convergence in best response settings.
- Robustness margins in model- and sampling- settings.
  - PO as a discrete-time nonlinear system;
  - Kleiman and input-to-state-stability;
  - Robust policy optimization as a small-input stable state optimization algorithm

Credits

Leilei Cui



Postdoc, MIT

Zhong-Ping Jiang

Professor, NYU

Research Significance

Continuous-Time Stochastic Policy Optimization
Lekan Molu
Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- (Deep) RL and modern AI
  - Robotic manipulation (Levine et al., 2016), text-to-visual processing (DALL-E), Atari games (**?**), e.t.c.
  - Policy optimization (PO) is fundamental to modern AI algorithms' success.
  - Major success story: functional mapping of observations to policies.
  - But how does it work?

Policy Optimization – General Framework

- PO encapsulates policy gradients (**?**) or PG, actor-critic methods (Vrabie and Lewis, 2011), trust region PO **?**, and proximal PO methods (**?**).
- PG particularly suitable for complex systems.

$$\min J(K) \quad \text{subject to } K \in \mathcal{K} \tag{1}$$

where $\mathcal{K} = \{K_1, K_2, \cdots, K_n\}$.

- $J(K)$ could be tracking error, safety assurance, goal-reaching measure of performance e.t.c. required to be satisfied.

Continuous-time RL control applications

Continuous-Time Stochastic Policy Optimization
Lekan Molu
Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- A little randomness in a system's mathematical model coefficients?
  - Population growth model: $dN/dt = a(t)N(t), \; N(0) = N_0$; growth rate $a(t)$ subject to random effects e.g. $a(t) = r(t) +$ "noise".
  - We only know the distribution of "noise".
- Filtering and state estimation problems where the nature of the noise is unknown, but it is observed via sensor measurements.
  - Kalman + Bucy Filters – aerospace (Apollo, Mariner etc.).

Continuous-time RL control applications

- Semielliptic P.D.E.s with Dirichlet boundary value problems e.g. slender flexible rods, Cosserat dynamics etc: $\Delta q = \sum_{i=1}^{n} \dfrac{\partial^2 q}{\partial \xi_i^2} = 0 \in \Omega, \;\; q = q_{\rightarrow} \text{ on } \partial\Omega, \;\; \Omega \subset \mathbb{R}^n$
- An economic portfolio problem where the price, $p(t)$, of a stock satisfies a stochastic differential equation e.g. $dp/dt = (a + \alpha \cdot \text{``}noise\text{''})p$ for $a > 0$, $\alpha \in reline$.
- Call options pricing: The *Black-Scholes option price formula*.

Policy Optimization – Open questions

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- Gradient-based data-driven methods: prone to divergence from true system gradients.
  - Challenge I: Optimization occurs in non-convex objective landscapes.
    - Get performance certificates as a mainstay for control design: Coerciveness property (**?**).
  - Challenge II: Taming PG's characteristic high-variance gradient estimates (REINFORCE, NPG, Zeroth-order approx.).
    - Hello, (linear) robust ($\mathcal{H}_\infty$-synthesis) control!

Policy Optimization – Open questions

- Challenge III: Under what circumstances do we have convergence to a desired equilibrium in RL settings?
- Challenge IV: Stochastic control, not deterministic control settings.
  - models involving round-off error computations in floating point arithmetic calculations; the stock market; protein kinetics.
- Challenge V: Continuous-time RL control.
  - Very little theory. Lots of potential applications encompassing rigid and soft robotics, aerospace or finance engineering, protein kinetics.

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

$\mathcal{H}_\infty$-Control Under Model Mismatch

$$
\mathrm{d}x(t) = Ax(t)\mathrm{d}t + Bu(t)\mathrm{d}t + D\mathrm{d}w(t),
$$
$$
z(t) = Cx(t) + Eu(t), \;\; \alpha > 0;
$$

Algorithm 1 Search for the closed-loop $\mathcal{H}_\infty$-norm

```
Given a user-defined step size η > 0
Set the initial upper bound on γ as γ_ub = ∞.
Initialize a buffer for possible H∞ norms for each K_1 to be found, Γ_buf = {}.
Initialize ordered poles P = {p_i ∈ Re(s) < 0 | i = 1, 2, }          ▷ p_1 < p_2 < ···
for p_i ∈ P do
    Place p_i on (2);                                  ▷ (Tits and Yang, 1996)
    Compute stabilizing K_1^{p_i}
    Find lower bound γ_lb for H(γ, K_1^{p_i});         ▷ using (22)
    Γ_buf(i) = get_hinf_norm(T_zw, γ_lb, K_1^{p_i}).
end for
function get_hinf_norm(T_zw, γ_lb, K_1^{p_i})
    while γ_ub = ∞ do
        γ := (1 + 2η) γ_lb;
        Get λ_i(H(γ, K_1^{p_i}))                       ▷ c.f. (14)
        if Re(Λ) ≠ ∅ for Λ = {λ_1, ··· λ_n} then
            Set γ_ub = γ; exit
        else
            Set buffer Γ_lb = {}
            for λ_k ∈ {Imag(Λ)_{:p-1}} do               ▷ k = 1 to K
                Set m_k = ½(ω_k + ω_{k+1})
                Set Γ_lb(k) = max{σ [T_zw(j m_k)]};
            end for
            γ_lb = max(Γ_lb)
        end if
        Set γ_ub = ½(γ_lb + γ_ub).
    end while
    return γ_ub
end function
```

Tools: Complexity, Convergence, Robustness.

Continuous-Time Stochastic Policy Optimization
Lekan Molu
Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- Risk-sensitive $\mathcal{H}_\infty$-control (Glover, 1989) and discrete- and continuous-time mixed $\mathcal{H}_2/\mathcal{H}_\infty$ design (Khargonekar et al., 1988; **?**):
  - min. upper bound on $\mathcal{H}_2$ cost subject to satisfying a set of risk-sensitive (often $\mathcal{H}_\infty$) constraints (**?**):

$$min_{K \in \mathcal{K}} J(K) := Tr(P_K D D^\top) \tag{2}$$
$$\text{subject to } \mathcal{K} := \{K | \rho(A - BK) < 1, \|T_{zw}(K)\|_\infty < \gamma\}$$

  - P_K: solution to the generalized algebraic Riccati equation (GARE);
  - A, B, D, K: standard closed-loop system matrices;
  - $\|T_{zw}(K)\|_\infty$: $\mathcal{H}_\infty$-norm of the closed-loop transfer function from a disturbance input w to output z.

Tools: Complexity, Convergence, Robustness.

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

Infinite-horizon

- discrete-time deterministic LQR settings (Fazel et al., 2018):

$$\min_{K \in \mathcal{K}} \mathbb{E} \sum_{t=0}^{\infty} (x_t^\top Q x_t + u_t^\top R u_t) \text{ s.t. } x_{t+1} = A x_t + B u_t, x_0 \sim \mathcal{P}_0$$

- discrete-time LQ problems under multiplicative noise (**?**): $\min_{\pi \in \Pi} \mathbb{E}_{x_0, \{\delta_i\}, \{\gamma_i\}\}} \sum_{t=0}^{\infty} (x_t^\top Q x_t + u_t^\top R u_t)$ subject to $x_{t+1} = (A + \sum_{i=1}^{p} \delta_{ti} A_i) x_t + (B + \sum_{i=1}^{q} \gamma_{ti} B_i) u_t$;

(Non-exhaustive) Lit. Landscape on PO Theory

Literature landscape	Cont. time (Kalman '61, Luenberger '63)	Stochastic. LQR (Kalman '60)	Cont. Phase	LEQG or Mixed H_2/H_∞	Finite/Infinite Horizon
Fazel (2018)	No	No	Yes	No	Finite-horizon
Mohammadi (TAC -- 2020)	Yes	No	Yes	No	Finite-Horizon
Zhang (2019)	Yes	Yes (Gaussian)	Yes	Yes	Inf-horizon
Gravell (2021)	No	Multiplicative	Yes	No	Inf-horizon
Zhang (2020)	No	No	Yes	Yes	Rand-horizon
Molu (2022)	Yes	Yes (Brownian)	Yes	Yes	Inf-Horizon
Cui & Molu (2023)	Yes	Yes (Brownian)	Yes	Yes	Inf-Horizon

Mainstay

- Continuous-time infinite-dimensional linear systems.
  - Disturbances enter additively as random stochastic Wiener processes.
  - Many natural systems admit uncertain additive Brownian noise as diffusion processes.
    - Theoretical analysis machinery: Îto's stochastic calculus.
- Goal: keep controlled process, z, small i.e.

$$\|z\|_2 = \left(\int |z(t)|^2 dt \right)^{1/2},$$

  - Under a minimizing $u(x(t)) \in \mathcal{U}$ in spite of unforeseen $w(t) \in \mathcal{W} \subseteq \mathbb{R}^q$.

Minimization Objective and Risk-Sensitive Control

Continuous-Time Stochastic Policy Optimization
Lekan Molu
Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- Risk-sensitive linear exponential quadratic Gaussian objective functional (Jacobson, 1973):

$$\min_{u \in \mathcal{U}} \mathcal{J}_{exp}(x_0, u, w) = \mathbb{E}\Big|_{x_0 \in \mathcal{P}_0} \exp\left[\frac{\alpha}{2}\int_0^\infty z^\top(t) z(t) \mathrm{d}t\right],$$
$$\text{subject to } \mathrm{d}x(t) = Ax(t)\mathrm{d}t + Bu(t)\mathrm{d}t + D\mathrm{d}w(t),$$
$$z(t) = Cx(t) + Eu(t), \;\; \alpha > 0; \tag{3}$$

- where $dw/dt = \mathcal{N}(0, W)$, $x_0 = \mathcal{N}(0, \mu)$, and $(x_0, w(t)) \subseteq (\Omega, \mathcal{F}, \mathcal{P})$.

Minimization Objective and Risk-Sensitive Control

- A Taylor series expansion of (3) reveals:

$$\mathcal{J}_{exp}(x_0, u, w) = \lim_{T\to\infty} \mathbb{E}\Big|_{x_0\in\mathcal{P}_0} \left[\frac{\alpha}{2}\sum_{t=0}^{T} z^\top(t) z(t)\right] + \frac{\alpha^2}{4} var\left[\sum_{t=0}^{T} z^\top(t) z(t)\right]. \tag{4}$$

- Consider the variance term $\frac{\alpha^2}{4} var\left[\sum_{t=0}^{T} z^\top(t) z(t)\right] \to \epsilon$.
  - α a measure of risk-propensity if $\alpha > 0$;
  - α a measure of risk-aversion if $\alpha < 0$;
  - $\alpha = 0$ implies solving a classic LQP.

RL PO as a Risk-Sensitive Control Problem

- RL (via PG) computes high-variance gradient estimates from Monte-Carlo trajectory roll-outs and bootstrapping.
- If we set $\alpha > 0$ in the LEQG problem (3), we have a controlled setting where we can study the theoretical properties of RL-based PO.
- Framework: an ADP policy iteration (PI) in a continuous PO setting.
- LEQG also interprets as a risk-attenuation algorithm.

Contributions

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- A two-loop iterative alternating best-response procedure for computing the optimal mixed-design policy;
- Rigorous convergence analyses follow for the model-based loop updates;
- In the absence of exact system models, we provide an input-to-state-stable hybrid robust stabilization scheme.

Transition Slide

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions

Setup
Assumptions
Optimal Gain

Model-based PO
Outer loop
Stabilization and Convergence

Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

This page is left blank intentionally.

Problem Setup

For $\alpha > 0$, the cost $\mathcal{J}_{exp}(x_0, u) = \mathbb{E}\Big|_{x_0 \in \mathcal{P}_0} \exp\left[\frac{\alpha}{2}\int_0^\infty z^\top(t) z(t) \mathrm{d}t\right]$, becomes

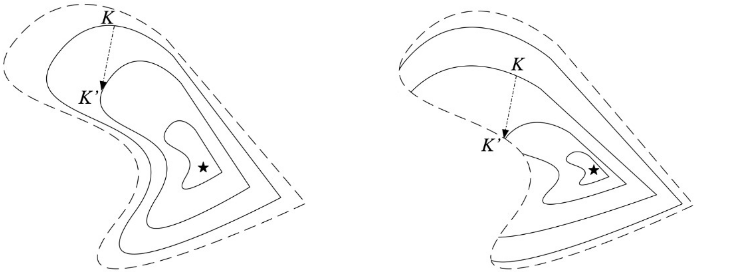
$$\mathbb{E}\Big|_{x_0 \in \mathcal{P}_0} \exp\left\{\frac{\alpha}{2}\int_0^\infty \left[x^\top(t) Q x(t) + u^\top(t) R u(t)\right] \mathrm{d}t\right\}, \tag{5}$$

with the associated closed loop transfer function,

$$T_{zw}(K) = (C - EK)(sI - A + BK)^{-1} D. \tag{6}$$

Nonconvexity and Coercivity in PG

- Coercivity: iterates remain feasible and strictly separated from the infeasible set as the cost decreases.



(a) Landscape of LQR

(b) Landscape of Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ Control

Figure: Coercivity property of PG on LQR and in mixed-design settings. Credit: (Zhang et al., 2019).

Assumptions

- $C^\top C = Q \succ 0,\ E^T(C,\ E) = (0,\ R)$ for some $R \succ 0$.
- Coercivity satisfaction: (A, B) is stabilizable;
- Optimization satisfaction: $(\sqrt{Q}, A)$ is detectable.

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

Transition Slide

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions

Setup
Assumptions
Optimal Gain

Model-based PO
Outer loop
Stabilization and Convergence

Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

This page is left blank intentionally.

PO and Dynamic Games: Finite-horizon Gain

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- Coercivity: feasibility set of optimization iterates

$$\mathcal{K} = \{ K : \lambda_i(A - B_1 K) < 0, \; \| T_{zw}(K) \|_\infty < \gamma \}. \tag{7}$$

- Finite-horizon optimization $u^\star(t) = -K^\star_{leqg} \hat{x}(t)$.
- $K^\star_{leqg} = R^{-1} B^\top P_\tau$, and P_τ is the unique, symmetric, positive definite solution to the algebraic Riccati equation (ARE)

$$A^\top P_\tau + P_\tau A - P_\tau (B R^{-1} B^\top - \alpha^{-2} D D^\top) P_\tau = -Q. \tag{8}$$

(**?**, Proposition I), (Duncan, 2013) .

- ∞-horizon case: $P^\star \triangleq P_\infty = \lim_{\tau \to \infty} P_\tau$, and $K^\star_{leqg} \triangleq K_\infty = \lim_{\tau \to \infty} K_\tau$[Theorem on limit of monotonic operators (**?**)].

Solving the LEQG Problem

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- Directly solving the LEQG problem (3) in policy-gradient frameworks incurs biased gradient estimates during iterations;
- Affects risk-sensitivity preservation in infinite-horizon LTI settings (see (**?**Zhang et al., 2019));
- Workaround: an equivalent dynamic game formulation to the stochastic LQ PO problem.

Two-Player Zero-Sum Game and LEQG

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- An equivalent closed-loop two-player game connection (**?**, Lemma 1):

$$\begin{aligned}&\min_{u \in \mathcal{U}} \max_{\xi \in W} \bar{\mathcal{J}}_\gamma(x_0, u, \xi) \\ &\quad \text{subject to } \mathrm{d}x(t) = Ax(t)\mathrm{d}t + Bu(t)\mathrm{d}t + D\mathrm{d}w(t), \\ &\qquad\qquad z(t) = Cx(t) + Eu(t) \qquad (9)\end{aligned}$$

$$\bar{\mathcal{J}}_\gamma(x_0, u, \xi) = \mathbb{E}_{x_0 \sim \mathcal{P}_0, \xi(t)} \int_0^\infty \left[x^\top(t) Q x(t) + u^\top(t) R u(t)\right] \mathrm{d}t - \mathbb{E}_{x_0 \sim \mathcal{P}_0, \xi(t)} \int_0^\infty \left[\gamma^2 \xi^\top(t) \xi(t)\right] \mathrm{d}t$$

, $\xi (\equiv dw) \sim \mathcal{N}(0, \Sigma)$, and $\gamma \equiv \alpha$.

Proof Sketch (?, Lemma 1)

- If a non-negative definite (n.n.d) GARE (8)'s solution exists, then a minimal realization $P^\star$ must exist.
  - Existence: the bounded real Lemma (Zhou et al., 1996).
- If $(A, Q^{\frac{1}{2}})$ is observable, then every n.n.d solution of (8), *i.e.* $P^\star$, is positive definite.
- For a n.n.d $P^\star$, we essentially have a Nash (equivalently a Saddle) equilibrium with $\bar{\mathcal{J}}_\gamma = \underline{\mathcal{J}}_\gamma$.

Proof Sketch (?, Lemma 1)

- If $\bar{\mathcal{J}}_\gamma$ is finite for some $\gamma = \hat{\gamma} > 0$, then $\bar{\mathcal{J}}_\gamma$ is bounded (if and only if the pair (A, B) is stabilizable).
- For a bounded $\bar{\mathcal{J}}_\gamma$ for some $\gamma = \hat{\gamma}$ and for optimal $K^\star = R^{-1}B^\top P_{K,L}$, $L^\star = \gamma^{-2}D^\top P_{K,L}$ and all $\gamma > \hat{\gamma}$, $\bar{\mathcal{J}}_\gamma$ admits the closed-loop matrices

$$A_K^\star = A - BK^\star,\ A_{K,L}^\star = A_K^\star + DL^\star. \tag{10}$$

- Whence, the saddle-point optimal controllers are

$$u^\star(x(t)) = -K^\star x(t),\ \xi^\star(x(t)) = L^\star x(t). \tag{11}$$

Transition Slide

Continuous-Time Stochastic Policy Optimization
Lekan Molu
Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

This page is left blank intentionally.

Model-based PO

- Define $\{p, q\}_{p=1,q=1}^{\bar{p},\bar{q}}$ where $(\bar{p}, \bar{q}) \in \mathbb{N}_+$ as nested iteration indices for a gain K_p (in an outer loop) and an alternating gain $L_q(K_p)$ (in an inner-loop).

Problem 1 (Model-Based Policy Iteration)

Given system matrices A, B, C, D, E, find the optimal controller gains K_p, $L_q(K_p)$ that robustly stabilizes (3) such that the controller gains do not leave the set of all suboptimal controllers denoted by

$$\check{\mathcal{K}} = \{(K_p, L_q(K_p)) : \lambda_i(A_K^p) < 0, \lambda_i(A_{K,L}^{p,q}) < 0, \\ \|T_{zw}(K_p, L_q(K_p))\|_\infty < \gamma \textit{ for all } (p, q) \in \mathbb{N}\}. \quad (12)$$

Model-based Policy Optimization

- Further, define the following closed-loop matrix identities

$$A_K^p = A - BK_p, \quad A_{K,L}^{p,q} = A_K^p + DL_q(K_p),$$
$$Q_K^p = Q + K_p^\top R K_p, \; A_K^\gamma = A_K^p + \gamma^{-2} DD^\top P_K^p. \tag{13}$$

- Equation (13) informs the value iterations of the Riccati equations for the outer and inner loops.

$$A_K^{p\top} P_K^p + P_K^p A_K^p + Q_K^p + \gamma^{-2} P_K^p DD^\top P_K^p = 0, \tag{14a}$$
$$K_{p+1} = R^{-1} B^\top P_K^p. \tag{14b}$$

$$A_{K,L}^{(p,q)^\top} P_{K,L}^{p,q} + P_{K,L}^{p,q} A_{K,L}^{p,q} + Q_K^p - \gamma^2 L_q^\top(K_p) L_q(K_p) = 0 \tag{15a}$$
$$K_{p+1} = R^{-1} B^\top P_K^{p,q}, \; L_{q+1}(K_p) = \gamma^{-2} D^\top P_{K,L}^{p,q}. \tag{15b}$$

Kleinman's Algorithm

- An iterative algorithm for solving infinite-time Riccati equations (Kleinman, 1968).
- Based on a successive substitution method.
- For a *deterministic LTI system's* cost matrix P_d, the value iterations of P_d^k are monotonically convergent to $P_d^\star$.
- Kleinman's algorithm as policy iteration
  - Choose a stabilizing control gain K_0, and let $p = 0$.
  - (Policy evaluation) Evaluate the performance of K_p from the GARE's solution.
  - (Policy improvement) Improve the policy: $K_p = -R^{-1}B^\top P_d^p$.
  - Advance iteration $p \leftarrow p + 1$.

Model-based Policy Iteration

Algorithm 1: (Model-Based) PO via Policy Iteration

Input: Max. outer iteration $\bar{p}$, $q = 0$, and an $\epsilon > 0$;
Input: Desired risk attenuation level $\gamma > 0$;
Input: Minimizing player's control matrix $R \succ 0$.

Compute $(K_0, L_0) \in \mathcal{K}$; ▷ From [24, Alg. 1];
Set $P_{K,L}^{0,0} = Q_K^0$; ▷ See equation (9);
for $p = 0, \ldots, \bar{p}$ **do**
  Compute Q_K^p and A_K^p ▷ See equation (9);
  Obtain P_K^p by evaluating K_p on (10);
  while $\|P_K^p - P_{K,L}^{p,q}\|_F \leq \epsilon$ **do**
    Compute $L_{q+1}(K_p) := \gamma^{-2} D^\top P_{K,L}^{p,q}$;
    Solve (11) until $\|P_K^p - P_{K,L}^{p,q}\|_F \leq \epsilon$;
    $\bar{q} \leftarrow q + 1$
  end
  Compute $K_{p+1} = R^{-1} B^\top P_{K,L}^{p,\bar{q}}$ ▷ See (11b) ;
end

Transition Slide

Continuous-Time Stochastic Policy Optimization
Lekan Molu
Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

This page is left blank intentionally.

Convergence Analyses: Outer Loops

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview

Risk-sensitive control

Contributions

Setup

Assumptions

Optimal Gain

Model-based PO

Outer loop

Stabilization and Convergence

Sampling-based PO

Discrete-time system

Sampling-based nonlinear system

Lemma 1

Under our assumptions and for the ARE (14), if $K_0 \in \mathcal{K}$, then for any $p \in \mathbb{N}_+$, we must have the following conditions for the optimal $K^\star$ and $P^\star$,

(1) $K_p \in \mathcal{K}$;

(2) $P_K^0 \succeq P_K^1 \succeq \cdots P_K^p \succeq \cdots \succeq P^\star$;

(3) $\lim_{p\to\infty}\|K_p - K^*\|_F = 0$, $\lim_{p\to\infty}\|P_K^p - P^*\|_F = 0$.

Proof Sketch: The Bounded Real Lemma

Under our standard stabilizability and observability assumptions, for a stabilizing gain K, the following conditions are equivalent

- $$\|\mathcal{T}(K)\|_\infty < \gamma;$$
- The Riccati equation
$$A_K^\top P_K + P_K A_K + C^\top C + K^\top RK + \gamma^{-2} P_K DD^\top P_K = 0, \tag{16}$$
admits a unique positive definite solution $P_K \succeq 0$ for a Hurwitz matrix $(A_K + \gamma^{-2}DD^\top P_K)$;
- There exists $P_K \succ 0$ such that
$$A_K^\top P_K + P_K A_K + Q + K^\top RK + \gamma^{-2} P_K DD^\top P_K \prec 0. \tag{17}$$

Stabilizing Proof Sketch

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions

Setup
Assumptions
Optimal Gain

Model-based PO
Outer loop
Stabilization and Convergence

Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- At an iteration 0, find a K_0 that is stabilizing (**?**, Alg. 1), so that $K_0 \in \mathcal{K}$ by the bounded real Lemma.
- For $p > 0$, set $Q_K^{p+1} = C^\top C + K_{p+1}^\top R K_{p+1}$, the outer loop GARE is

$$A_K^{(p+1)^\top} P_K^p + P_K^p A_K^{(p+1)} + \gamma^{-2} P_K^p D D^\top P_K^p + C^\top C + K_{p+1}^\top R K_{p+1} + (K_{p+1} - K_p)^\top R (K_{p+1} - K_p) = 0. \quad \text{(A.2)}$$

Thus, for a stabilizing $K_{p+1} (\neq K_p)$ we must have $(K_{p+1} - K_p)^\top R (K_{p+1} - K_p) \succ 0$ so that

$$A_K^{(p+1)^\top} P_K^p + P_K^p A_K^{(p+1)} + \gamma^{-2} P_K^p D D^\top P_K^p + Q_K^{p+1} \prec 0. \quad \text{(A.3)}$$

- For $p > 1$, $K_p \in \mathcal{K}$. Rest: completion of squares, the bounded real Lemma, and the theorem on the "limit of monotonic operators." (**?**).

Convergence Analysis

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions

Setup
Assumptions
Optimal Gain

Model-based PO
Outer loop
Stabilization and Convergence

Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- In (Zhang et al., 2019, Theorem A.7 and A.8), the authors showed that this controller update in the outer-loop has a global sub-linear and local quadratic convergence rates.
- We now show that the outer-loop iteration has a global linear convergence rate.

Transition Slide

Continuous-Time Stochastic Policy Optimization
Lekan Molu
Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

This page is left blank intentionally.

Convergence Analysis: Outer Loop

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions

Setup
Assumptions
Optimal Gain

Model-based PO
Outer loop
Stabilization and Convergence

Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

Lemma 2

Let $\Psi = (K_{p+1} - K_p)^\top R(K_{p+1} - K_p)$; *and* $\Psi = \Psi^\top \succeq 0$. *Furthermore, let* $\Phi \in \mathbb{R}^{n \times n}$ *be Hurwitz so that* $\Theta = \int_0^\infty e^{(\Phi^\top t)} \Psi e^{(\Phi t)} dt$ *and define* $c(\Phi) = \log(5/4)\|\Phi\|^{-1}$. *Then,* $\|\Theta\| \geq \frac{1}{2} c(\Phi) \|\Psi\|$.

Convergence Analysis: Outer Loop

Remark 1

For $A_K = A - BK$, we know from the bounded real Lemma (Zhang et al., 2019, Lemma A.1) that the Riccati equation

$$A_K^\top P_K + P_K A_K + Q_K + \gamma^{-2} P_K D D^\top P_K = 0 \tag{18}$$

admits a unique positive definite solution $P_K \succ 0$ with a Hurwitz $(A_K + \gamma^{-2} D D^\top P_K)$.

Transition Slide

This page is left blank intentionally.

Optimality of the Iteration

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview

Risk-sensitive control

Contributions

Setup

Assumptions

Optimal Gain

Model-based PO

Outer loop

Stabilization and Convergence

Sampling-based PO

Discrete-time system

Sampling-based nonlinear system

Lemma 3 (Optimality of the iteration)

Consider any $K \in \mathcal{K}$, let $K' = R^{-1}B^\top P_K$ (where P_K is the solution to (18)*, and $\Psi_K = (K - K')^\top R(K - K')$. If $\Psi_K = 0$, then $K = K^\star$.*

Proof.

Since $R \succ 0$, $\Psi_K = 0$ implies $K = K'$. Therefore at $\Psi_K = 0$, we must have $K = K'$ which implies that $P_K = P'_K$. If $K = K'$ and $P_K = P'_K$, it suffices to conclude that $K' = K \triangleq K^\star$ where $K^\star = R^{-1}B^\top P^\star$. Hence, $\Psi_K = 0$ is tantamount to $P_K = P^\star$ and $K = K^\star$. □

Bound on Cost Difference Matrix

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview

Risk-sensitive control

Contributions

Setup

Assumptions

Optimal Gain

Model-based PO

Outer loop

Stabilization and Convergence

Sampling-based PO

Discrete-time system

Sampling-based nonlinear system

Lemma 4 (Bound on Cost Difference Matrix)

For any $h > 0$*, define* $\mathcal{K}_h := \{K \in \mathcal{K} | Tr(P_K^p - P^\star) \le h\}$*. For any* $K \in \mathcal{K}_h$*, let* $K' := R^{-1}B^\top P_K^p$*, where* P_K^p *is the p'th iterate's solution to (18), and* $\Psi_{K_p} = (K_p - K_p')^\top R(K_p - K_p')$*. Then, there exists* $b(h) > 0$*, such that* $\|P_K^p - P^\star\|_F \le b(h)\|\Psi_{K_p}\|_F$*.*

Bound on Cost Difference Matrix

Continuous-Time Stochastic Policy Optimization
Lekan Molu
Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- For $A^\star = A - BR^{-1}B^\top P^\star + \gamma^{-2}DD^\top P^\star$, rewrite the closed-loop Riccati equation as

$$\begin{aligned} &A^{\star\top} P_K^p + P_K^p A^\star + Q_{K_p} + (K^\star - K_p)^\top R K_p' \\ &+ K_p'^\top R(K^\star - K_p) - \gamma^{-2} P^\star DD^\top P_K^p - \gamma^{-2} P_K^p DD^\top P^\star \\ &+ \gamma^{-2} P_K^p DD^\top P_K^p = 0. \end{aligned} \tag{19}$$

- Then do completion of squares so that

$$\begin{aligned} &A^{\star\top}(P_K^p - P^\star) + (P_K^p - P^\star)A^\star + \Psi_{K_p} \\ &+ \gamma^{-2}(P_K^p - P^\star)DD^\top(P_K^p - P^\star) \\ &- (K_p' - K^\star)^\top R(K_p' - K^\star) = 0. \end{aligned} \tag{20}$$

Proof

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- Implicit function theorem: $P_K^p = f(K_p \in \mathcal{K})$, $f(\cdot) \in \mathcal{C}^n$.
- There exists a ball $\mathcal{B}_\delta(K^\star) := \{K \in \mathcal{K} | \|K - K^\star\|_F \le \delta\}$, such that $\mathcal{A}(K)$ is invertible for any $K \in \mathcal{K}_h \cap \mathcal{B}_\delta(K^\star)$.
  - $\mathcal{A}(K_p) = I_n \otimes A^{\star\top} + (A - BR^{-1}B^\top P_K^p + \gamma^{-2}DD^\top P_K^p)^\top \otimes I_n$.
- Therefore, for any $K \in \mathcal{K}_h \cap \mathcal{B}_\delta(K^\star)$,
  - $\|\tilde{P}_K^p\|_F \le \underline{\sigma}^{-1}(\mathcal{A}(K_p))\|\Psi_{K_p}\|_F$.
- Similarly, for any $K \in \mathcal{K}_h \cap \mathcal{B}_\delta^c(K^\star)$, where $\mathcal{B}^c$ is a complement of $\mathcal{B}$, $\Psi_{K_p} \neq 0$ and there exists a constant $b_1 > 0$ such that $\|\Psi_{K_p}\| \ge b_1$.
- Set $b_2 = \max_{K \in \mathcal{K}_h \cap \mathcal{B}_\delta(K^\star)} \underline{\sigma}^{-1}(\mathcal{A}(K))$ and $b(h) = \max\{b_2, \frac{h + Tr(P^\star)}{b_1}\}$, then the proof follows immediately.

Transition Slide

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

This page is left blank intentionally.

Outer Loop Convergence: Exponential Stability of P_K^p

Theorem 2

For any $h > 0$ and $K_0 \in \mathcal{K}_h$, there exists $\alpha(h) \in \mathbb{R}$ such that $Tr(P_K^{p+1} - P^\star) \leq \alpha(h) Tr(P_K^p - P^\star)$. That is, $P^\star$ is an exponentially stable equilibrium.

Transition Slide

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

This page is left blank intentionally.

Convergence Analysis: Inner Loop

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions

Setup
Assumptions
Optimal Gain

Model-based PO
Outer loop
Stabilization and Convergence

Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- Now, we analyze the monotonic convergence rate of the inner loop.
- Given arbitrary gains $K_p \in \mathcal{K}$ and $L_q(K_p) \in \mathcal{L}$, and $P_{K,L}^{p,q} \succ 0$ solution of the inner-loop Lyapunov equation, the cost matrix $P_{K,L}^{p,q}$ monotonically converges to the solution of (15).

$$A_{K,L}^{(p,q)^\top} P_{K,L}^{p,q} + P_{K,L}^{p,q} A_{K,L}^{p,q} + Q_K^p - \gamma^2 L_q^\top(K_p) L_q(K_p) = 0 \tag{21a}$$

$$K_{p+1} = R^{-1} B^\top P_K^{p,q}, \; L_{q+1}(K_p) = \gamma^{-2} D^\top P_{K,L}^{p,q}. \tag{21b}$$

Convergence Analysis: Inner Loop I

Lemma 5

Suppose that $L_0(K_0)$ is stabilizing, then for any $q \in \mathbb{N}_+$ (with $P_{K,L}^{p,\bar{q}}$ as the solution to (15)), i.e.

$$A_{K,L}^{(p,q)^\top} P_{K,L}^{p,q} + P_{K,L}^{p,q} A_{K,L}^{p,q} + Q_K^p - \gamma^2 L_q^\top(K_p) L_q(K_p) = 0 \quad (22a)$$

$$K_{p+1} = R^{-1} B^\top P_K^{p,q}, \; L_{q+1}(K_p) = \gamma^{-2} D^\top P_{K,L}^{p,q}. \quad (22b)$$

Then, the following statements hold

1. $A_{K,L}^{p,q}$ *is Hurwitz;*
2. $P_{K,L}^{p,\bar{q}} \succeq \cdots \succeq P_K^{(p,q+1)} \succeq P_K^{p,q} \succeq \cdots \succeq P_{K,L}^{p,0}$; *and*
3. $\lim_{q \to \infty} \| P_{K,L}^{p,q} - P_{K,L}^{p,\bar{q}} \|_F = 0.$

Convergence Rate – Inner Loop

Lemma 6 (Monotonic Convergence of the Inner-Loop)

For any $K \in \mathcal{K}$, let $L(K)$ be the control gain for the player w such that $A_K + DL(K)$ is Hurwitz. Let P_K^L be the solution of

$$\left(A_K + DL(K)\right)^\top P_K^L + P_K^L \left(A_K + DL(K)\right) + Q_K - \gamma^2 L(K)^\top L(K) = 0. \tag{23}$$

Let $L'(K) = \gamma^{-2} D^\top P_K^L$ and $\Psi_K^L = \gamma^{-2}(L'(K) - L(K))^\top (L'(K) - L(K))$. Then, for a $c(K) = Tr\left(\int_0^\infty e^{(A_K + DL(K^\star))t} e^{(A_K+DL(K^\star))^\top t} \mathrm{d}t\right)$, the following inequality holds $Tr(P_K - P_K^L) \leq \|\Psi_K^L\| c(K)$.

Convergence of the Inner Loop Iteration

Theorem 3

For a $K \in \check{\mathcal{K}}$, and for any $(p, q) \in \mathbb{N}_+$, there exists $\beta(K) \in \mathbb{R}$ such that

$$Tr(P_K^p - P_{K,L}^{p,q+1}) \leq \beta(K) Tr(P_K^p - P_{K,L}^{p,q}). \qquad (24)$$

Remark 2

*As seen from Lemma 5, $P_K^p - P_{K,L}^{p,q} \succeq 0$. By the norm on a matrix trace (**?**, Lemma 13) and the result of Theorem 3, we have $\|P_K - P_{K,L}^{p,q}\|_F \leq Tr(P_K - P_{K,L}^{p,q}) \leq \beta(K) Tr(P_K)$, i.e. $P_{K,L}^{p,q}$ exponentially converges to P_K in the Frobenius norm.*

Algorithm as a Policy Iteration Scheme

- Choosing a stabilizing K_p we first evaluate u's performance by solving (14).
  - This is the policy evaluation step in PI.
- The policy is then improved in a following iteration by solving for the cost matrix in (15b);
  - This is the policy improvement step.
- Essentially, a policy iteration algorithm whereupon
  - Performance of an initial control gain K_p is first evaluated against a cost function.
  - A newer evaluation of the cost matrix $P_{K,L}^{p,q}$ is then used to improve the controller gain K_{p+1} in the outer loop.

Transition Slide

This page is left blank intentionally.

Sampling-based PO Scheme

- A, B, C, D, E are often unavailable so that the policy evaluation step will result in biased estimates.
- There is the possibility for a divergence from the stability-robustness feasibility set $\check{\mathcal{K}}$:
  - When errors are present from I/O or state data;
  - Residuals from early termination of numerically solving Riccati equations;
  - Using an approximate cost function owing to inexact values of Q and R;
  - Since the inner loop is computed in a finite number of steps;
  - In a data sampling scheme, we must guarantee the stability and robustness of the closed-loop system.

Sampling-based PO: Statement of the Problem

Problem 4 (Sampling-based Policy Optimization)

If A, B, C, D, E, P are all replaced by approximate matrices $\hat{A}, \hat{B}, \hat{C}, \hat{D}, \hat{E}, \hat{P}$, under what conditions will the sequences $\{\hat{P}_{K,L}^{p,q}\}_{(p,q)=1}^{(p,q)=\infty}$, $\{\hat{K}_p\}_{p=0}^{\infty}$, $\{\hat{L}_q\}_{q=0}^{\infty}$ converge to a small neighborhood of the optimal values $\{P_{K,L}^{\star}\}_{(p,q)=0}^{(p,q)=\infty}$, $\{K_p^{\star}\}_{p=0}^{\infty}$, and $\{L_q^{\star}\}_{q=0}^{\infty}$?

Discrete-Time Nonlinear System Interpretation

Continuous-Time Stochastic Policy Optimization
Lekan Molu
Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- From assumptions, a $P_K^0 \in \mathbb{S}^n$ exists such that when applied to find a K_0 such a K_0 will be stabilizing.
- Approximation errors between the nested iteration steps yield a hybrid of a continuous-time policy gain pair $(\hat{K}_p, \hat{L}_q(\hat{K}_p))$ and a learning scheme.
  - This learning scheme is essentially a discrete sampled data from a nonlinear system (owing to errors from various sources).
- Task: under inexact loop updates, lump iterates of gain errors into system inputs to the online PO scheme;

Transition Slide

This page is left blank intentionally.

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions

Setup
Assumptions
Optimal Gain

Model-based PO
Outer loop
Stabilization and Convergence

Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- How do we converge to the optimal solution and preserve closed-loop dynamic stability?
- What does input-to-state stability (ISS) Sontag (2008) have to do with it?

Online Model-free Reparameterization

- Suppose that $\hat{P}_K^0 \in \mathbb{S}^n$ is chosen following the controllability and stabilizability assumptions.
  - Then $\hat{K}_k^1 = R^{-1}B^\top \hat{P}_K^0$ will be stabilizing since $\tilde{K}_k^1 = \hat{K}_k^1 - K_k^1 \triangleq 0$.
- Ditto argument for L_1.

Problem 5

For $(p, q) > 0$, show that for $\tilde{K}_k^p = \hat{K}_k^p - K_k^p \triangleq 0$ so that the sequence $\{P_{K,L}^{p,q}\}_{(p,q)=0}^{\infty}$ converges to the locally exponentially stable $\hat{P}_{K,L}^\star$.

Transition Slide

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions

Setup
Assumptions
Optimal Gain

Model-based PO
Outer loop
Stabilization and Convergence

Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

This page is left blank intentionally.

Hybrid System Reparameterization

- Lump estimate errors as an input into the gain terms to be computed in the PO algorithm.
- With inexact outer loop update, K_{p+1} becomes biased so that the inexact outer-loop GARE value iteration involves the recursions

$$\hat{A}_K^{p\top} \hat{P}_K^p + \hat{P}_K^p \hat{A}_K^p + \hat{Q}_K^p + \gamma^{-2} \hat{P}_K^p D D^\top \hat{P}_K^p = 0, \tag{25a}$$

$$\hat{K}_{p+1} = R^{-1} B^\top \hat{P}_K^p + \tilde{K}_{p+1} \triangleq \bar{K}_{p+1} + \tilde{K}_{p+1}, \tag{25b}$$

  - NB: $\hat{A}_K^p = A - B\hat{K}_p$ and $\hat{Q}_K^p = Q + \hat{K}_p^\top R \hat{K}_p$.

Discrete-Time System Closed-loop System

- Same argument for the inner-loop inexact GARE value iteration updates:

$$\hat{A}_{K,L}^{p,q\top}\hat{P}_{K,L}^{p,q} + \hat{P}_{K,L}^{p,q}\hat{A}_{K,L}^{p,q} + \hat{Q}_K^p - \gamma^2\hat{L}_q^\top\hat{L}_q(\hat{K}_p) = 0 \quad (26a)$$

$$\hat{K}_{p+1} = R^{-1}B^\top\hat{P}_K^{p,q} + \tilde{K}_p, \quad (26b)$$

$$\hat{L}_{q+1}(\hat{K}_p) = \gamma^{-2}D^\top\hat{P}_{K,L}^{p,q} + \tilde{L}_{q+1}(\tilde{K}_p) \quad (26c)$$

$$\triangleq \bar{L}_{q+1}(\bar{K}_p) + \tilde{L}_{q+1}(\tilde{K}_p). \quad (26d)$$

- Rewrite the infinite-dimensional stochastic differential equation as the discrete-time system (for iterates $(p,q) > 0$):

$$\mathrm{d}x = [\hat{A}_{K,L}^{p,q}x + B(\hat{K}_p x - D\hat{L}_q(K_p) + u)]\mathrm{d}t + D\mathrm{d}w. \quad (27)$$

Transition Slide

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

This page is left blank intentionally.

System Trajectories from HJB Interpretation

- On a time interval $[s, s+\delta s]$, it follows from Itô's stochastic calculus and the Hamilton-Jacobi-Bellman equation that

$$d\left[x^\top(s+\delta s)\hat{P}^{p,q}_{K,L}x(s+\delta s) - x^\top(s)\hat{P}^{p,q}_{K,L}x(s)\right] = (\mathrm{d}x)^\top \hat{P}^{p,q}_{K,L}x + x^\top \hat{P}^{p,q}_{K,L}\mathrm{d}x + (\mathrm{d}x)^\top \hat{P}^{p,q}_{K,L}(\mathrm{d}x). \tag{28}$$

- Along the trajectories of equation (27) and using the gains in (15), *i.e.*

$$K_{p+1} = R^{-1}B^\top P^{p,q}_K, \;\; L_{q+1}(K_p) = \gamma^{-2}D^\top P^{p,q}_{K,L}.$$

System Trajectories

- The r.h.s. in (28) becomes

$$x^\top \left[\hat{A}_{K,L}^{p,q\top} \hat{P}_{K,L}^{p,q} + \hat{P}_{K,L}^{p,q} \hat{A}_{K,L}^{p,q}\right] x\mathrm{d}t + 2x^\top \hat{P}_{K,L}^{p,q} D\mathrm{d}w \tag{29}$$

$$+ 2x^\top \hat{P}_{K,L}^{p,q} B(K_p x - D\hat{L}_q(K_p) + u)dt + Tr(D^\top PD),$$

$$= -x^\top \hat{Q}_K^p x\mathrm{d}t - \gamma^{-2} x^\top \hat{P}_{K,L}^{p,q} DD^\top \hat{P}_{K,L}^{p,q} x dt + Tr(D^\top \hat{P}_{K,L}^{p,q} D) + 2x^\top \hat{P}_{K,L}^{p,q} B\left[\hat{K}_p x - D\hat{L}_q(K_p) + u\right] \mathrm{d}t + 2x^\top \hat{P}_{K,L}^{p,q} D\mathrm{d}w \tag{30}$$

System Trajectories via HJB Expansions

- So that

$$
\begin{aligned}
&x^\top(s+\delta s)\hat{P}^{p,q}_{K,L}(s+\delta s) - x^\top(s)\hat{P}^{p,q}_{K,L}x(s) \\
&= \int_s^{s+\delta s}\left[(-x^\top \hat{Q}^p_K x - \gamma^2 w^\top w)dt + 2\gamma^2 x^\top \hat{L}^\top_{q+1}(K_p)dw\right] \\
&+ \int_s^{s+\delta s} 2x^\top \hat{K}^\top_{p+1} R\left[\hat{K}_p x - D\hat{L}_q(\hat{K}_p) + u\right] dt \\
&\quad + \int_s^{s+\delta s} Tr(D^\top \hat{P}^{p,q}_{K,L} D)dt. \qquad (31)
\end{aligned}
$$

Transition Slide

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

This page is left blank intentionally.

Input To State System Interpretation

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- System matrices $\hat{A}_{K,L}^{p,q}, B, C, D$ now embedded within input and state terms: $\hat{Q}_K^p$, $\hat{K}_{p+1}$, and $\hat{L}_{q+1}$;
- Retrievable via online measurements.
- We essentially end up with an input-to-state system!
- The price that we pay is that the noise feedthrough matrix D must be known precisely.
  - No marvel: in many linear stochastic system with Brownian motion, D is identity (**??**).

Sampling-based Scheme

Continuous-Time Stochastic Policy Optimization
Lekan Molu
Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- Explore system model until we achieve exact equality in $\hat{A}^{p,q}_{K,L} \equiv A^{p,q}_{K,L}$, $\hat{P}^{p,q}_{K,L}$, $\hat{K}_{p+1} \equiv K_{p+1}$, and $\hat{L}_{q+1}(K_p) \equiv L_{q+1}(K_p)$.
  - Choose $u = -K_0 x + \eta_p$ and $w = -L_0 x + \eta_q$ where (η_p, η_q) is drawn uniformly at random over matrices with a Frobenium norm r similar to (**?**Fazel et al., 2018).

Sampled System Parameterization

Continuous-Time Stochastic Policy Optimization
Lekan Molu
Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- Consider the identities

$$
\begin{aligned}
x^\top \hat{Q}_K^p x &= (x^\top \otimes x^\top)\,\texttt{vec}(\hat{Q}_K^p), \\
\gamma^2 w^\top w &= \gamma^2 (w^\top \otimes w^\top)\ \texttt{vec}(I_v), \\
2\gamma^2 x^\top \hat{L}_{q+1}^\top(\hat{K}_p) dw &= 2\gamma^2 (I_n \otimes x^\top) dw\ \texttt{vec}(\hat{L}_{q+1}^\top(\hat{K}_p)), \\
2x^\top \hat{K}_{p+1}^\top R \hat{K}_p x &= 2(x^\top \otimes x^\top)(I_n \otimes \hat{K}_p^\top)\,\texttt{vec}(\hat{K}_{p+1}^\top R), \\
2x^\top \hat{K}_{p+1}^\top R D \hat{L}_q(\hat{K}_p) &= 2(\hat{L}_q^\top(\hat{K}_p) D^\top \otimes x^\top)\,\texttt{vec}(\hat{K}_{p+1}^\top R), \\
2x^\top \hat{K}_{p+1}^\top R u &= 2(u^\top \otimes x^\top)\,\texttt{vec}(\hat{K}_{p+1}^\top R), \\
Tr(D^\top \hat{P}_{K,L}^{p,q} D) &= \texttt{vec}^\top(D)\,\texttt{vec}(\hat{P}_{K,L}^{p,q} D).
\end{aligned} \tag{32}
$$

Sampled System Parameterization I

- Let $\Delta_{xx} \in \mathbb{R}^{\frac{n(n+1)}{2} l}$, $\Delta_{ww} \in \mathbb{R}^{\frac{v(v+1)}{2} l}$, $I_{xx} \in \mathbb{R}^{l \times n^2}$, and $I_{ux} \in \mathbb{R}^{l \times mn}$ for $l \in \mathbb{N}_+$
- It follows that

$$\Delta_{xx} = \left[\texttt{vecv}(x_1), \ldots, \texttt{vecv}(x_l)\right]^\top, \; x_l = x_{l+1} - x_l,$$

$$\Delta_{ww} = \left[\texttt{vecv}(w_1), \ldots, \texttt{vecv}(w_l)\right]^\top, \; w_l = w_{l+1} - w_l,$$

$$I_{xx} = \left[\int_{s_0}^{s_1} x \otimes x \, dt, \ldots, \int_{s_{l-1}}^{s_l} x \otimes x \, dt\right]^\top,$$

Transition Slide

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

This page is left blank intentionally.

Sampled System Parameterization

$$
\begin{aligned}
I_{xw} &= \left[\int_{s_0}^{s_1} (I_n \otimes x)\mathrm{d}w, \ldots, \int_{s_{l-1}}^{s_l} (I_n \otimes x)\mathrm{d}w\right]^\top, \\
I_{ux} &= \left[\int_{s_0}^{s_1} u \otimes x \,\mathrm{d}t, \ldots, \int_{s_{l-1}}^{s_l} u \otimes x \,\mathrm{d}t\right]^\top.
\end{aligned} \tag{33}
$$

Next, set

$$
\begin{aligned}
\Theta_{K,L}^{p,q} &= \Big[\Delta_{xx}, -2I_{xx}(I_n \otimes \hat{K}_p^\top) + 2(\hat{L}_q^\top(\hat{K}_p)D^\top \otimes x^\top) \\
&\quad -2I_{ux}, -2\gamma^2 I_{xw}, -\mathtt{vec}^\top(D)\mathtt{vec}(\hat{P}_{K,L}^{p,q} D)\Big], \qquad \text{(34a)} \\
\Upsilon_{K,L}^{p,q} &= \left[-I_{xx}\mathtt{vec}(\hat{Q}_K^p), \; -\gamma^2 I_{ww}\mathtt{vec}(I_v)\right]. \qquad \text{(34b)}
\end{aligned}
$$

Sampled System Parameterization

Continuous-Time Stochastic Policy Optimization
Lekan Molu
Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

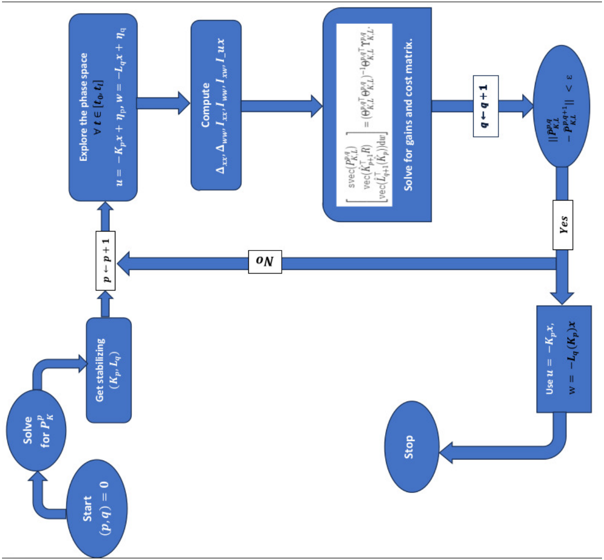
Define 1_{q^2} as a one-vector with dimension q^2. Thus,

$$\Theta_{K,L}^{p,q} \left[\texttt{svec}(P_{K,L}^{p,q}) \quad \texttt{vec}(\hat{K}_{p+1}^\top R) \quad \texttt{vec}(\hat{L}_{q+1}^\top(\hat{K}_p)) \quad 1_{q^2} \right]^\top = \Upsilon_{K,L}^{p,q}. \tag{35}$$

Suppose that $\Theta_{K,L}^{p,q}$ is of full rank, then we can retrieve the unknown matrices via least squares estimation *i.e.*

$$\begin{bmatrix} \texttt{svec}(P_{K,L}^{p,q}) \\ \texttt{vec}(\hat{K}_{p+1}^\top R) \\ \texttt{vec}(\hat{L}_{q+1}^\top(\hat{K}_p))\mathrm{d}w \\ 1_{q^2} \end{bmatrix} = (\Theta_{K,L}^{p,q\top} \Theta_{K,L}^{p,q})^{-1} \Theta_{K,L}^{p,q\top} \Upsilon_{K,L}^{p,q}. \tag{36}$$

Sampling-based Algorithm



Transition Slide

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions

Setup
Assumptions
Optimal Gain

Model-based PO
Outer loop
Stabilization and Convergence

Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

This page is left blank intentionally.

Robustness Analyses

- Define $\tilde{P} = P_K - \hat{P}_K$ and $\tilde{K} = K - \hat{K}$.
- Keep $|\tilde{K}| < \epsilon$, start with a $K \in \mathcal{K}$: iterates stay in $\mathcal{K}$.

Lemma 7 (Lemma 10, C&M, '23)

For any $K \in \mathcal{K}$, there exists an $e(K) > 0$ such that for a perturbation $\tilde{K}$, $K + \tilde{K} \in \mathcal{K}$, as long as $\|\tilde{K}\| < e(K)$.

Theorem 6

The inexact outer loop is small-disturbance ISS. That is, for any $h > 0$ and $\hat{K}_0 \in \mathcal{K}_h$, if $\|\tilde{K}\| < f(h)$, there exist a $\mathcal{KL}$-function $\beta_1(\cdot,\cdot)$ and a $\mathcal{K}_\infty$-function $\gamma_1(\cdot)$ such that

$$\|P^p_{\hat{K}} - P^\star\| \le \beta_1(\|P^0_{\hat{K}} - P^*\|, p) + \gamma_1(\|\tilde{K}\|). \tag{37}$$

ISS Outer Loop Robustness Proof

- Prelim result (Lemma 12, C&M, '23): For any $h > 0$ and $K \in \mathcal{K}_h$, let $K' = R^{-1}B^\top P_K$, where P_K is the solution of (18), and $\hat{K}' = K' + \tilde{K}$. Then, there exists $f(h) > 0$, such that $\hat{K}' \in \mathcal{K}_h$ as long as $\|\tilde{K}\| < f(h)$.
- Therefore, $\hat{K}_K^p \in \mathcal{K}_h$ for any $p \in \mathbb{N}_+$.
- Let

$$f_1(\hat{K}') = \frac{\log(5/4)b(h)}{2n\|A^\star_{\hat{K}'}\|}, f_2(\hat{K}') = Tr\left(\int_0^\infty e^{A^{\star\top}_{\hat{K}'}t} e^{A^{\star t}_{\hat{K}'}} \mathrm{d}t\right).$$

ISS Outer Loop Robustness Proof

- $$\underline{f}_1(h) = \inf_{\hat{K}' \in \mathcal{K}_h} f_1(\hat{K}') > 0, \bar{f}_2(h) = \sup_{\hat{K}' \in \mathcal{K}_h} f_2(\hat{K}') < \infty. \tag{38}$$

- This implies
$$Tr(P_{\hat{K}}^p - P^\star) \leq [1 - \underline{f}_1(h)]\, Tr(P_{\hat{K}}^{p-1} - P^\star) + \bar{f}_2(h)\|R\|\|\tilde{K}_K^p\|^2. \tag{39}$$

- Repeating (39) for $p, p-1, \cdots, 1$,
$$Tr[P_{\hat{K}}^p - P^\star] \leq (1 - \underline{f}_1)^p\, Tr(P_{\hat{K}}^1 - P^\star) + \frac{\bar{f}_2\|R\|\|\tilde{K}\|_\infty^2}{\underline{f}_1(h)}. \tag{40}$$

Outer Loop Robustness

It follows from (40) and (Mori, 1988, Theorem 2) that

$$\|P_{\hat{K}}^{p} - P^{\star}\|_F \le (1 - \underline{f}_1)^p \sqrt{n} \|P_{\hat{K}}^{1} - P^{\star}\|_F + \frac{\bar{f}_2 \|R\| \|\tilde{K}\|_\infty^2}{\underline{f}_1}. \tag{41}$$

As $p \to \infty$, $P_{\hat{K}}^{p} \to P^{\star}$. Whence, a radius of $P^{\star}$'s neighbor is proportional to $\|\tilde{K}\|_\infty^2$.

Inner Loop Robustness

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions

Setup
Assumptions
Optimal Gain

Model-based PO
Outer loop
Stabilization and Convergence

Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

The perturbed inner-loop iteration (26) has inexact matrix $\hat{A}^{p,q}_{K,L}$, and sequences $\{\hat{L}_{q+1}(K_p)\}_{q=0}^{\infty}$, and $\{\hat{P}^{p,q}_{K,L}\}_{q=0}^{\infty}$.

Lemma 8 (Stability of the Inner-Loop's System Matrix)

Given $K \in \check{\mathcal{K}}$, there exists a $g \in \mathbb{R}_+$, such that if $\|\tilde{L}_{q+1}(K_p)\|_F \leq g$, $\hat{A}^{p,q}_{K,L}$ is Hurwitz for all $q \in \mathbb{N}_+$.

Inner Loop Robustness

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

Theorem 7

Assume $\|\tilde{L}_q(K_p)\| < e$ *for all* $q \in \mathbb{N}_+$. *There exists* $\hat{\beta}(K) \in [0, 1)$, *and* $\lambda(\cdot) \in \check{\mathcal{K}}_\infty$, *such that*

$$\|\hat{P}^{p,q}_{K,L} - P^{p,q}_{K,L}\|_F \leq \hat{\beta}^{q-1}(K) Tr(P^{p,q}_{K,L}) + \lambda(\|\tilde{L}\|_\infty). \tag{42}$$

- From Theorem 7, as $q \to \infty$, $\hat{P}^{p,q}_{K,L}$ approaches the solution P_K and enters the ball centered at $P^{p,q}_{K,L}$ with radius proportional to $\|\tilde{L}\|_\infty$.
- The proposed inner-loop iterative algorithm well approximates $P^{p,q}_{K,L}$.

Transition Slide

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

This page is left blank intentionally.

Numerical Results – Car Cruise Control System

- (?, §3.1):

$$m\frac{dv}{dt} = \alpha_n u \tau(\alpha_n v) - mgC_r sgn(u) - \frac{1}{2}\rho C_d A|v|v - mg\sin\theta \tag{43}$$

- $u(x(t)) = [u_1(t), u_2(t)]$ must maintain a constant velocity v (the state), whilst automatically adjusting the car's throttle, $u_1(t), t \in [0, T]$
  - despite disturbances characterized by road slope changes ($u_3 = \theta$),
  - rolling friction (F_r), and
  - aerodynamic drag forces (F_d).

Numerical Results – Car Cruise Control System

Continuous-Time Stochastic Policy Optimization

Lekan Molu

Outline and Overview
Risk-sensitive control
Contributions

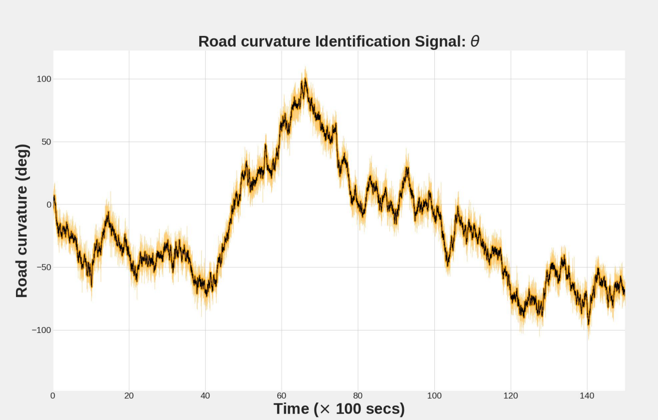
Setup
Assumptions
Optimal Gain

Model-based PO
Outer loop
Stabilization and Convergence

Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

- Well-suited to our robust control formulation because
  - the disturbances and state variables are separable and can be lumped into the form of the stochastic differential equations;
  - it is a multiple-input (throttle, gear, vehicle speed) single-output (vehicle acceleration) system that introduces modeling challenges;
  - the entire operating range of the system is nonlinear though there is a reasonable linear bandwidth that characterize the input/output (I/O) system as we will see shortly.

Road (Disturbance) Profile



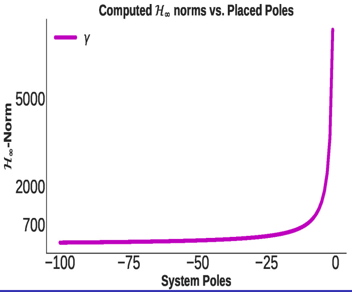
Search for initial stabilizing gain and $\mathcal{H}_\infty$-norm bound.

Proposition 1

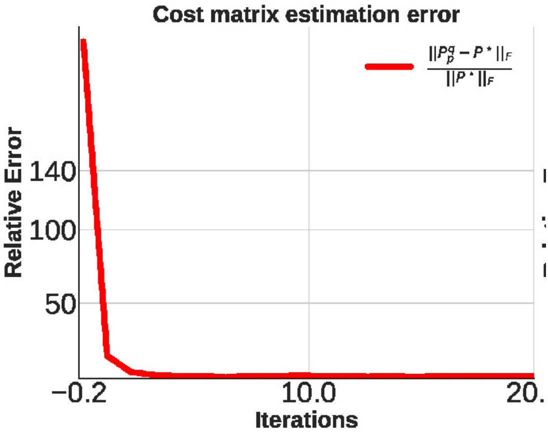
(**?**) *For all $\omega_p \in \mathbb{R}$, we have that $j\omega_p$ is an eigenvalue of the Hamiltonian $H(\gamma_1)$ if and only if γ_1 is a singular value of $T_{zw}(j\omega_p)$.*

Algorithm 1 Search for the closed-loop $\mathcal{H}_\infty$-norm

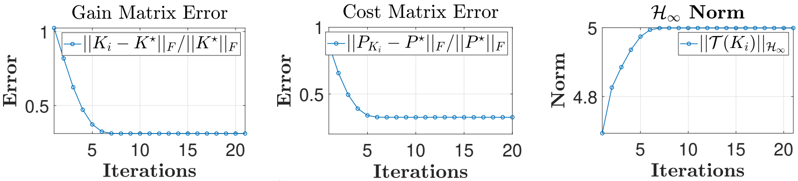
Given a user-defined step size $\eta > 0$
Set the initial upper bound on γ as $\gamma_{ub} = \infty$.
Initialize a buffer for possible $\mathcal{H}_\infty$ norms for each K_1 to be found, $\Gamma_{buf} = \{\}$.
Initialize ordered `poles` $\mathcal{P} = \{p_i \in Re(s) < 0 \,|\, i = 1, 2, \}$ ▷ $p_1 < p_2 < \cdots$
for $p_i \in \mathcal{P}$ **do**
  Place p_i on (2); ▷ (Tits and Yang, 1996)
  Compute stabilizing $K_1^{p_i}$
  Find lower bound γ_{lb} for $H(\gamma, K_1^{p_i})$; ▷ using (22)
  $\Gamma_{buf}(i) =$ `get_hinf_norm`$(T_{zw}, \gamma_{lb}, K_1^{p_i})$.
end for
function `get_hinf_norm`$(T_{zw}, \gamma_{lb}, K_1^{p_i})$
  while $\gamma_{ub} = \infty$ **do**
    $\gamma := (1 + 2\eta)\,\gamma_{lb}$;
    Get $\lambda_i(H(\gamma, K_1^{p_i}))$ ▷ c.f. (14)
    if `Re`$(\Lambda) \neq \emptyset$ for $\Lambda = \{\lambda_1, \cdots \lambda_n\}$ **then**
      Set $\gamma_{ub} = \gamma$; exit
    else
      Set buffer $\Gamma_{lb} = \{\}$
      for $\lambda_k \in \{$`Imag`$(\Lambda)_{:p-1}\}$ **do** ▷ $k = 1$ to K
        Set $m_k = \frac{1}{2}(\omega_k + \omega_{k+1})$
        Set $\Gamma_{lb}(k) = \max\{\sigma\,[T_{zw}(jm_k)]\}$;
      end for
      $\gamma_{lb} = \max(\Gamma_{lb})$



Cost Matrix and Gains Convergence

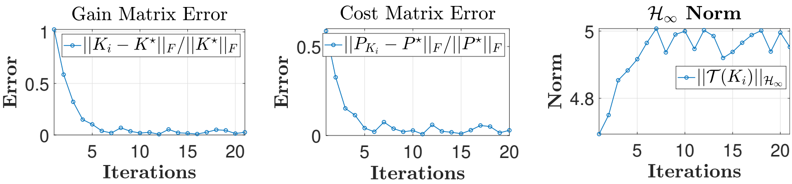


Pendulums Experiment – Comparison to NPG



Model-free design: $\|\tilde{K}\|_\infty = 0.15$.

Pendulums Experiment – Comparison to NPG



Model-based design: $\|\tilde{K}\|_\infty = 0.15$.

Double Pendulum and Acrobot Experiment – Comparison to NPG

Table: Computational Time: Model-based PO vs. Model-free PO vs. NPG.

Policy Optimization Computational time (secs)					
Double Inverted Pendulum			Triple Inverted Pendulum		
Model-based	Model-free	NPG	Model-based	Model-free	NPG
0.0901	0.3061	2.1649	0.1455	0.7829	2.3209

References I

Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-End Training of Deep Visuomotor Policies. *The Journal of Machine Learning Research*, 17(1):1334–1373, 2016.

Draguna Vrabie and Frank Lewis. Adaptive dynamic programming for online solution of a zero-sum differential game. *J. Contr. Theory Appl.*, 9:353–360, 08 2011. doi: 10.1007/s11768-011-0166-4.

K. Glover. Minimum entropy and risk-sensitive control: the continuous time case. In *Proceedings of the 28th IEEE Conference on Decision and Control,*, pages 388–391 vol.1, 1989.

P.P. Khargonekar, I.R. Petersen, and M.A. Rotea. $\mathcal{H}_\infty$ optimal control with state-feedback. *IEEE Transactions on Automatic Control*, 33(8):786–788, 1988. doi: 10.1109/9.1301.

Maryam Fazel, Rong Ge, Sham Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1467–1476. PMLR, 10–15 Jul 2018.

D. Jacobson. Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games. *IEEE Transactions on Automatic Control*, 18(2):124–131, 1973. doi: 10.1109/TAC.1973.1100265.

Kaiqing Zhang, Bin Hu, and Tamer Başar. Policy Optimization for $\mathcal{H}_2$ Linear Control with $\mathcal{H}_\infty$ Robustness Guarantee: Implicit Regularization and Global Convergence. *arXiv e-prints*, art. arXiv:1910.09496, October 2019.

Tyrone E. Duncan. Linear-Exponential-Quadratic Gaussian control. *IEEE Transactions on Automatic Control*, 58(11):2910–2911, 2013. doi: 10.1109/TAC.2013.2257610.

Kemin Zhou, John Comstock Doyle, and Keith Glover. *Robust and Optimal Control*. Prentice hall Upper Saddle River, NJ, 1996.

David Z. Kleinman. On an iterative technique for riccati equation computations. *IEEE Transactions on Automatic Control*, 13:114–115, 1968.

References II

Outline and Overview
Risk-sensitive control
Contributions
Setup
Assumptions
Optimal Gain
Model-based PO
Outer loop
Stabilization and Convergence
Sampling-based PO
Discrete-time system
Sampling-based nonlinear system

Eduardo D. Sontag. *Input to State Stability: Basic Concepts and Results*, pages 163–220. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

T. Mori. Comments on "a matrix inequality associated with bounds on solutions of algebraic Riccati and Lyapunov equation" by J. M. Saniuk and I.B. Rhodes. *IEEE Transactions on Automatic Control*, 33 (11):1088–, 1988. doi: 10.1109/9.14428.

Towards Adaptive Soft Robots with Improved Motion Strategies: Strides in Modeling and Control

Lekan Molu

Microsoft Research

New York City, NY 10012

Presented by **Lekan Molu** (Lay-con Mo-lu)

April 8, 2025

Talk Overview

- The principle of morphological computation in nature
  - Morphology: shape, geometry, and mechanical properties.
  - Computation: sensorimotor information transmission among geometrical components.
- Morphology and computation in artificial robots
  - Cosserat Continua and reduced soft robot models.
  - **Reductions**: Structural Lagrangian properties and control.
- Towards real-time strain regulation and control
  - **Simplexity**: Hierarchical and fast versatile control with reduced variables.

Credits

Shaoru Chen

Postdoc, MSR

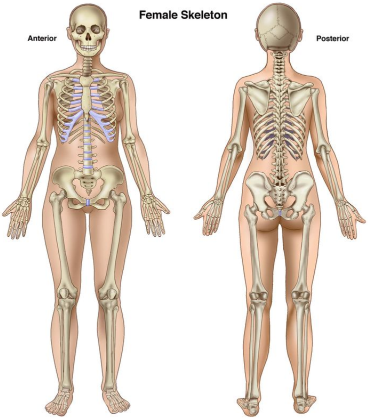
Lekan Molu

Senior Researcher, MSR

Morphology and computation

- Morphology: Emergent behaviors of natural organisms from complex sensorimotor nonlinear mechanical feedback from the environment.
  - Shape affecting behavioral response.
  - Geometrical Arrangement of motors such that processing and perception affect computational characteristics.
  - Mechanical properties that allow the engineering of emergent behaviors via adaptive environmental interaction.
- Computation: The information transformation among the system geometrical units, upon environmental perception, that effect morphological changes in shape and material properties.

MC in vertebrates – a case for soft designs



An adult human skeleton $\approxeq$ 11% of the body mass. ©Brittanica

- The arrangement and compliance of body parts, perception, and computation creates emergence of complex interactive behavior.
- Soft bodies seem critical to the emergence of adaptive natural behaviors.
- Morphological computation is crucial in the design of robots that execute adaptive natural behavior.

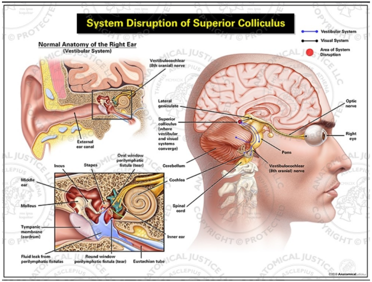
Simplexity in Morphological Computation

- Simplexity: Exploiting structure for effective control.
  - The geometrical tuning of the morphology and neural circuitry in the brain of mammals that simplify the perception and control of complex natural phenomena.
  - Not exactly simplified models or reduced complexity.
  - But rather, sparse connections and finite variables to execute adaptive sensorimotor strategies!
- Example: Saccades (focused eye movements) are controlled by (small) Superior Colliculus in the human brain.
  - Plug: Complex neural circuitry; simple control systems!

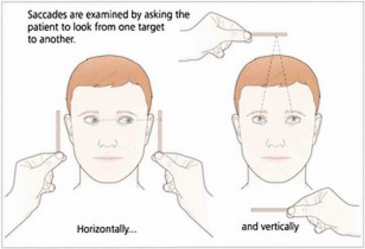
Simplexity: The Central Pattern Generator

- A neural mechanism (in vertebrates) that generates motor control with minimal parameters.
- CPG: Neurons and synapses couple to generate effective motor activation for rhythmic environmental motion.
  - In Lampreys, only two signals trigger swimming motion, for example!
  - This CPG enables indirect use of brain computational power via nonlinear feedback from stretch receptor neurons on Lamprey's skin.

Saccades and the Superior Colliculus



©Anatomical Justice.



Credit: Vision and Learning Center.

Morphing in Invertebrates: Cephalopods

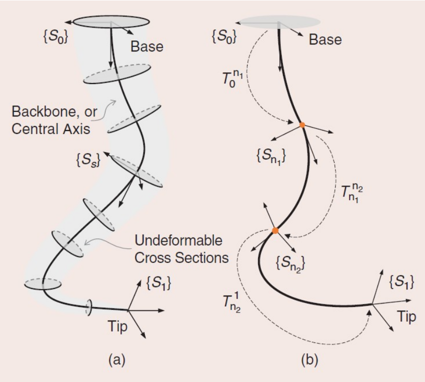
Cuttlefish. ©Monterey Bay Museum

Octopus. ©Smithsonian Magazine

The Octopus and Cuttlefish

- No exoskeleton, or spinal cord.
- A muscular hydrostat: transversal, longitudinal, and oblique muscles along richly innervated arms and mechanoreceptors:
  - Allows for bending, stretching, stiffening, and retraction.
  - Diverse compliance across eight arms imply sophisticated motion strategies in the wild!
- Simplexity enhanced by a peripheral nervous system and a central nervous system.

Soft Robot Mechanism in Focus

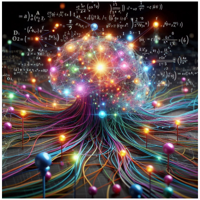


A continuum soft robot whose mechanics can be well-described with Cosserat rod theory. Reprinted from (Della Santina et al. (2023))

- One dimension is quintessentially longer than the other two.
- Characterized by a central axis with undeformable discs that characterize deformable cross-sectional segments.
- Strain and deformation, via e.g. Cosserat rod theory, enables precise finite-dimensional mathematical models.

A Finite and Reliable Model

- A soft robot's usefulness is informed by control system that melds its body deformation with internal actuators.
- By design, this calls for a high-fidelity model or a delicate balancing of complex morphology and data-driven methods.



- Non-interpretable; non-reliable.
- $\times$Continuous coupled interaction between the material, actuators, and external affordances.

The case for model-based control

- Soft robots are infinite degrees-of-freedom continua i.e., PDEs are the main tools for analysis.
- Nonlinear PDE theory is tedious and computationally intensive.
- Notable strides in reduced-order, finite-dimensional mathematical models that induce tractability in continuum models.

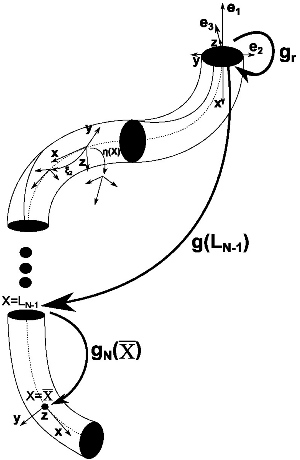
Tractable reduced-order models

- Morphoelastic filament theory: Moulton et al. (2020); Kaczmarski et al. (2023); Gazzola et al. (2018);
- Generalized Cosserat rod theory: Rubin (2000); Cosserat and Cosserat (1909);
- The constant curvature model: Godage et al. (2011);
- The piecewise constant curvature model: Webster and Jones (2010); Qiu et al. (2023); and
- Ordinary differential equations-based discrete Cosserat model: Renda et al. (2016, 2018).

Cosserat-based piecewise constant strain model

- A discrete Cosserat model: Renda et al. (2018).
  - Shapes defined by a finite-dimensional functional space, parameterized by a curve, $X : [0, L]$..
  - Assumes constant strains between finite nodal points on robot's body.
  - Strain-parameterized dynamics on a reduced special Euclidean-3 group (SE(3)).

The piecewise constant strain model



Credit: Renda et al. (2018).

- C-space: $g(X) : X \to \mathbb{SE}(3) = \begin{pmatrix} \mathrm{R}(X) & \mathrm{p}(X) \\ 0^\top & 1 \end{pmatrix}$.
- Strain and twist vectors: $\{\boldsymbol{\eta}, \boldsymbol{\xi}\} \in \mathbb{R}^6$.
  - $\{\boldsymbol{\eta}, \boldsymbol{\xi}\} := \{\mathrm{q}, \dot{\mathrm{q}}\}$
- Strain field: $\breve{\eta}(X) = g^{-1}\partial g/\partial X$.
- Twist field: $\breve{\xi}(X) = g^{-1}\partial g/\partial t$.

Dynamic equations

From the continuum equations for a cable-driven soft arm [Renda et al. (2014)], we can derive the following dynamic equation [Renda et al. (2018)]:

$$\underbrace{\left[\int_0^{L_N} J^T \mathcal{M}_a \mathrm{J} dX\right]}_{\mathrm{M(q)}} \ddot{\mathrm{q}} + \underbrace{\left[\int_0^{L_N} \mathrm{J}^T \mathrm{ad}^{\star}_{\mathrm{J\dot{q}}} \mathrm{M}_a \mathrm{J} dX\right]}_{\mathrm{C}_1(\mathrm{q},\dot{\mathrm{q}})} \dot{\mathrm{q}} + \underbrace{\left[\int_0^{L_N} J^T \mathcal{M}_a \dot{\mathrm{J}} dX\right]}_{\mathrm{C}_2(\mathrm{q},\dot{\mathrm{q}})} \dot{\mathrm{q}}$$

$$+ \underbrace{\left[\int_0^{L_N} \mathrm{J}^T \mathrm{D} \mathrm{J} \|\mathrm{J}\dot{\mathrm{q}}\|_p dX\right]}_{D(\mathrm{q},\dot{\mathrm{q}})} \dot{\mathrm{q}} - \underbrace{(1 - \rho_f/\rho) \left[\int_0^{L_N} \mathrm{J}^T \mathrm{M} \mathrm{Ad}_{\mathrm{g}}^{-1} dX\right]}_{\mathrm{N(q)}} \mathrm{Ad}_{\mathrm{g}_r}^{-1} \mathrm{G}$$

$$- \underbrace{\mathrm{J}(\bar{X})^T \mathrm{F}_p}_{\mathrm{F(q)}} - \underbrace{\int_0^{L_N} \mathrm{J}^T \left[\nabla_x \mathrm{F}_i - \nabla_x \mathrm{F}_a + \mathrm{ad}^{\star}_{\xi_n} (\mathrm{F}_i - \mathrm{F}_a)\right] dX}_{\boldsymbol{\tau}(\mathrm{q})} = 0, \qquad (1)$$

Structural properties – mass inertia operator

$$M(q)\ddot{q} + [C_1(q, \dot{q}) + C_2(q, \dot{q})]\,\dot{q} = F(q) + N(q)Ad_{g_r}^{-1}\mathcal{G} + \tau(q) - D(q, \dot{q})\dot{q}. \quad (2)$$

Property 1 (Boundedness of the Mass Matrix)

The mass inertial matrix $M(q)$ *is uniformly bounded from below by* mI *where* m *is a positive constant and* I *is the identity matrix.*

Proof of Property 1.

This is a restatement of the lower boundedness of $M(q)$ for fully actuated n-degrees of freedom manipulators [Romero et al. (2014)]. □

Structural properties – parameters Identification

Property 2 (Linearity-in-the-parameters)

There exists a constant vector $\boldsymbol{\Theta} \in \mathbb{R}^{l}$ *and a regressor function* $Y(\mathsf{q}, \dot{\mathsf{q}}, \ddot{\mathsf{q}}) \in \mathbb{R}^{N \times l}$ *such that*

$$\begin{aligned} \mathsf{M}(\mathsf{q})\ddot{(}\mathsf{q}) + \left[\mathsf{C}_1(\mathsf{q}, \dot{\mathsf{q}}) + \mathsf{C}_2(\mathsf{q}, \dot{\mathsf{q}}) + \mathsf{D}(\mathsf{q}, \dot{\mathsf{q}})\right] \dot{\mathsf{q}} - \mathsf{F}(\mathsf{q})\mathsf{N}(\mathsf{q}) Ad_{g_r}^{-1} \mathcal{G} \\ = \mathsf{Y}(\mathsf{q}, \dot{\mathsf{q}}, \ddot{\mathsf{q}})\Theta. \end{aligned} \tag{3}$$

Structural properties – skew symmetry of system inertial forces

Property 3 (Skew symmetric property)

The matrix $\dot{M}(q) - 2\left[C_1(q, \dot{q}) + C_2(q, \dot{q})\right]$ *is skew-symmetric.*

Skew-symmetric of robot's mass and Coriolis forces

By Leibniz's rule, we have

$$\dot{\mathrm{M}}(\mathrm{q}) = \frac{d}{dt}\left(\int_0^{L_N} \mathrm{J}^T \mathrm{M}_a \mathrm{J} dX\right) = \int_0^{L_N} \frac{\partial}{\partial t}\left(\mathrm{J}^T \mathrm{M}_a \mathrm{J}\right) dX$$
$$\triangleq \int_0^{L_N} \left(\dot{\mathrm{J}}^T \mathrm{M}_a \mathrm{J} + \mathrm{J}^T \dot{\mathrm{M}}_a \mathrm{J} + \mathrm{J}^T \mathrm{M}_a \dot{\mathrm{J}}\right) dX. \tag{4}$$

Therefore, $\dot{\mathrm{M}}(\mathrm{q}) - 2\left[C_1(\mathrm{q}, \dot{\mathrm{q}}) + C_2(\mathrm{q}, \dot{\mathrm{q}})\right]$ becomes

$$\int_0^{L_N} \left(\dot{\mathrm{J}}^\top \mathrm{M}_a \mathrm{J} + \mathrm{J}^\top \dot{\mathrm{M}}_a \mathrm{J} + \mathrm{J}^\top \mathrm{M}_a \dot{\mathrm{J}}\right) dX - 2\int_0^{L_N} \left(\mathrm{J}^\top \mathrm{ad}^\star_{\mathrm{J}\dot{\mathrm{q}}} \mathrm{M}_a \mathrm{J} + \mathrm{J}^\top \mathrm{M}_a \dot{\mathrm{J}}\right) dX \tag{5}$$

$$\triangleq \int_0^{L_N} \left(\dot{\mathrm{J}}^\top \mathrm{M}_a \mathrm{J} + \mathrm{J}^\top \dot{\mathrm{M}}_a \mathrm{J} - \mathrm{J}^\top \mathrm{M}_a \dot{\mathrm{J}}\right) dX - 2\int_0^{L_N} \mathrm{J}^\top \mathrm{ad}^\star_{\mathrm{J}\dot{\mathrm{q}}} \mathrm{M}_a \mathrm{J} dX. \tag{6}$$

Skew-Symmetric Property Proof

Similarly, $-\left[\dot{M}(q) - 2\left[C_1(q,\dot{q}) + C_2(q,\dot{q})\right]\right]^\top$ expands as

$$
\begin{aligned}
&-\dot{M}^\top(q) + 2\left[C_1^\top(q,\dot{q}) + C_2^\top(q,\dot{q})\right] = \\
&\int_0^{L_N} dX^\top \left(-J^\top M_a \dot{J} - J^\top \dot{M}_a J - \dot{J}^\top M_a J\right) + 2\int_0^{L_N} dX^\top \left(J^\top M_a \mathrm{ad}_{J\dot{q}} J + \dot{J}^\top M_a J\right) \\
&\triangleq \int_0^{L_N} \left(J^\top M_a \dot{J} - \dot{J}^\top M_a J - J^\top \dot{M}_a J\right) dX - 2\int_0^{L_N} J^\top \mathrm{ad}^\star_{J\dot{q}} M_a J dX
\end{aligned} \tag{7}
$$

which satisfies the identity:

$$
\begin{aligned}
&\dot{M}(q) - 2\left[C_1(q,\dot{q}) + C_2(q,\dot{q})\right] = \\
&\qquad -\left[\dot{M}(q) - 2\left[C_1(q,\dot{q}) + C_2(q,\dot{q})\right]\right]^\top .
\end{aligned} \tag{8}
$$

A fortiori, the skew symmetric property follows.

MC Takeaways: Simplexity

- Simplexity: Reliance on a few parameters to model an infinite-DoF system:

$$M(q)\ddot{q} + [C_1(q, \dot{q}) + C_2(q, \dot{q})]\,\dot{q} = F(q) + N(q)\mathrm{Ad}_{g_r}^{-1}G + \tau(q) - D(q, \dot{q})\dot{q}.$$

- Simplexity: From PDE to ODE, i.e. inifinite-dimensional analysis (Continuum PDE) to finite-dimensional ODE!

Control exploiting structural properties

Regarding the generalized torque $\boldsymbol{\tau}(\mathrm{q})$ as a control input, $\mathrm{u}(\mathrm{q}, \dot{\mathrm{q}})$, feedback laws are sufficient for attaining a desired soft body configuration.

Theorem 1 (Cable-driven Actuation)

For positive definite diagonal matrix gains K_D and K_p, without gravity/buoyancy compensation, the control law

$$\mathrm{u}(\mathrm{q}, \dot{\mathrm{q}}) = -\mathrm{K}_p \tilde{\mathrm{q}} - \mathrm{K}_D \dot{\mathrm{q}} - \mathrm{F}(\mathrm{q}) \tag{9}$$

under a cable-driven actuation globally asymptotically stabilizes system (2), where $\tilde{\mathrm{q}}(t) = \mathrm{q}(t) - \mathrm{q}^d$ is the joint error vector for a desired equilibrium point q^d.

Computational Control exploiting structural properties

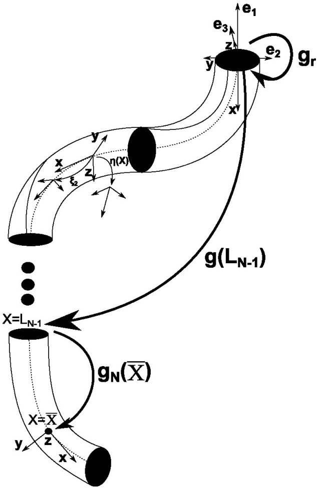
Corollary 2 (Fluid-driven actuation)

If the robot is operated without cables, and is driven with a dense medium such as pressurized air or water, then the term $F(\mathrm{q}) = 0$ *so that the control law* $\mathrm{u}(\mathrm{q}, \dot{\mathrm{q}}) = -\mathrm{K}_p\tilde{\mathrm{q}} - \mathrm{K}_D\dot{\mathrm{q}}$ *globally asymptotically stabilizes the system.*

Proof.

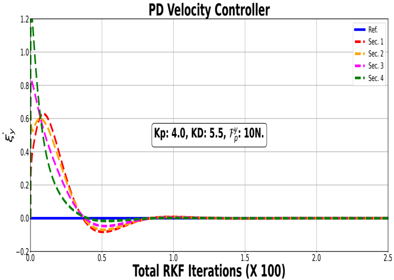
Proofs in Section V of Molu and Chen (2024). □

Robot parameters

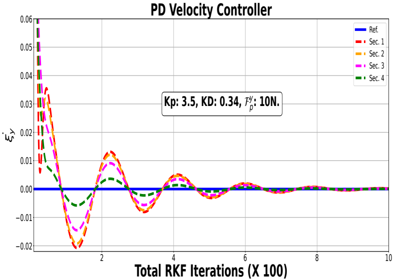


- Tip load in the $+y$ direction in the robot's base frame.
- Poisson ratio: 0.45; $\mathcal{M} = \rho[I_x, I_y, I_z, A, A, A]$ with $\rho = 2,000 kgm^{-3}$;
- $\mathrm{D} = -\rho_w \nu^T \nu \check{\mathrm{D}} \nu / |\nu|$.
- $X \in [0, L]$ discretized into 41 segments.

Computational Control exploiting structural properties

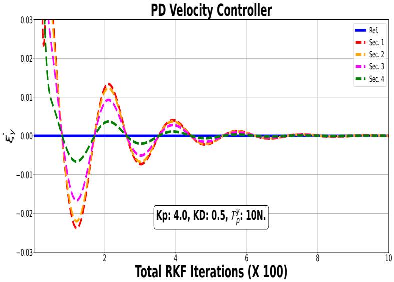


Cable-driven, strain twist setpoint terrestrial control.

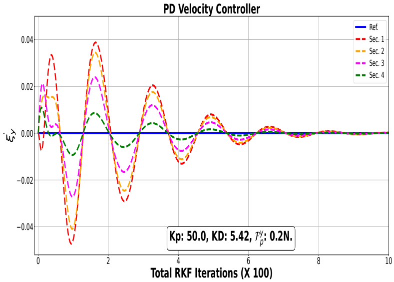


Fluid-actuated, strain twist setpoint terrestrial control.

Computational Control exploiting structural properties

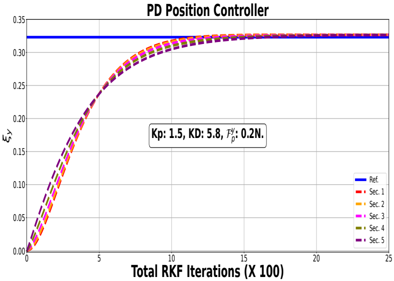


Fluid-actuated, strain twist setpoint underwater control.

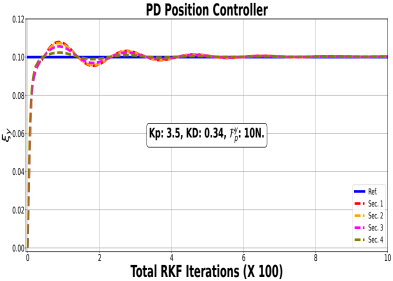


Cable-driven, strain twist setpoint regulation.

Computational Control exploiting structural properties



Cable-based position control with a small tip load, 0.2N.



Terrestrial position control.

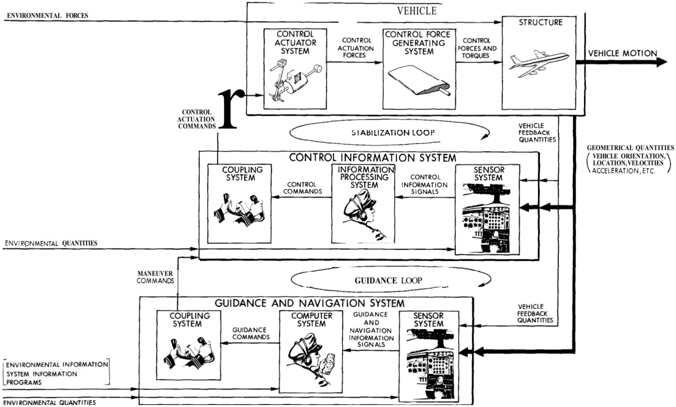
Exploiting Mechanical Nonlinearity for Feedback!

This page is left blank intentionally.

Hierarchical Dynamics and Control

- Reaching steps towards the real-time strain control of multiphysics, multiscale continuum soft robots.
- Separate subdynamics — aided by a perturbing time-scale separation parameter.
- Respective stabilizing nonlinear backstepping controllers.
- Stability of the interconnected singularly perturbed. system.
- Fast numerical results on a single arm of the Octopus robot arm.

A case for layered control



©C. Draper, "Guidance and Navigation, MIT, 1965.

Layered control architecture: Singularly Perturbed Dynamics

- Essentially a layered multirate control scheme (Matni et al. (2024)) of the various interconnected physics components of a soft robot prototype.
- Informed by a standard two-time-scale singularly perturbed system.

$$\dot{z}_1 = f(z_1, z_2, \epsilon, u_s, t), \; z_1(t_0) = z_1(0), \; z_1 \in \mathbb{R}^{6N}, \tag{10a}$$

$$\epsilon \dot{z}_2 = g(z_1, z_2, \epsilon, u_f, t), \; z_2(t_0) = z_2(0), \; z_2 \in \mathbb{R}^{6N} \tag{10b}$$

Framework: Singularly Perturbed Dynamics

- f and g are $C^n (n \gg 0)$ differentiable functions of their arguments;
- $\epsilon > 0$ denotes all small parameters to be ignored.
- u_s is the slow sub-dynamics' control law, and
- u_f is the fast sub-dynamics' controller.

Isolated Equilibrium Manifold Justification

Assumption 1 (Real and distinct root)

Equation (10) has the unique and distinct root $z_2 = \phi(z_1, t)$ *(for a sufficiently smooth* ϕ*) so that*

$$0 = g(z_1, \phi(z_1, t), 0, 0, t) \triangleq \bar{g}(z_1, 0, t), \; z_1(t_0) = z_1(0). \tag{11}$$

The slow subsystem therefore becomes

$$\dot{z}_1 = f(z_1, \phi(z_1, t), 0, u_s, t) \triangleq f_s(z_1, u_s, t). \tag{12}$$

Framework: Slow Dynamics Extraction

- Assumption: the fast feedback law is asymptotically stable;
  - It does not modify the open-loop equilibrium manifold of the fast dynamics.
- With $\epsilon = 0$ we have,

$$\dot{z}_1 = f(z_1, z_2, 0, u_s, t), \; z_1(t_0) = z_1(0), \tag{13a}$$

$$0 = g(z_1, z_2, 0, 0, t). \tag{13b}$$

Framework: Fast Dynamics Extraction

Introduce the time scale $T = t/\epsilon$, and write the deviation of z_2 from its isolated equilibrium manifold, $\phi(z_1, t)$ as $\tilde{z}_2 = z_2 - \phi(z_1, t)$. Then, (10) becomes

$$\frac{dz_1}{dT} = \epsilon f(z_1, \tilde{z}_2 + \phi(z_1, t), \epsilon, u_s, t), \tag{14a}$$

$$\frac{d\tilde{z}_2}{dT} = \epsilon \frac{dz_2}{dt} - \epsilon \frac{\partial \phi}{\partial z_1}\dot{z}_1, \tag{14b}$$

$$= g(z_1, \tilde{z}_2 + \phi(z_1, t), \epsilon, u_f, t) - \epsilon \frac{\partial \phi(z_1, t)}{\partial z_1}\dot{z}_1. \tag{14c}$$

Framework for singularly perturbed dynamics

Setting $\epsilon = 0$, we obtain the algebraic equation

$$\frac{d\tilde{z}_2}{dT} = \mathrm{g}(\mathrm{z}_1, \tilde{\mathrm{z}}_2 + \phi(z_1, t), 0, \mathrm{u}_f, t) \tag{15}$$

with z_1 frozen to its initial values.

Decomposition of SoRo Rod Dynamics

This page is left blank intentionally

Decomposition of SoRo Rod Dynamics

- $\mathcal{M}_i^{\text{core}}$: composite mass distribution as a result of microsolid i's barycenter motion;
- $\mathcal{M}^{\text{pert}}$: motions relative to $\mathcal{M}_i^{\text{core}}$, considered as a perturbation;
- $\mathcal{M} = \mathcal{M}^{\text{pert}} \cup \mathcal{M}^{\text{core}}$.
- Introduce the transformation: $[\mathsf{q}, \dot{\mathsf{q}}] = [\mathsf{q}, \mathsf{z}]$, rewrite (2):

$$\mathsf{M}(\mathsf{q})\dot{\mathsf{z}} + [\mathsf{C}_1(\mathsf{q},\mathsf{z}) + \mathsf{C}_2(\mathsf{q},\mathsf{z}) + \mathsf{D}(\mathsf{q},\mathsf{z})]\,\mathsf{z} - \mathsf{F}(\mathsf{q}) - \mathsf{N}(\mathsf{q})\mathsf{Ad}_{\mathsf{g}_r}^{-1}\mathsf{G} = \tau(\mathsf{q})$$

Dynamics separation

Suppose that $M^p = \int_{L^p_{\text{min}}}^{L^p_{\text{max}}} J^\top \boldsymbol{\mathcal{M}}^{pert} J dX$, and $M^c = \int_{L^c_{\text{min}}}^{L^c_{\text{max}}} J^\top M^{core} J dX$, then,

$$M(q) = (M^c + M^p)(q), \; N = (N^c + N^p)(q), \tag{16a}$$

$$F(q) = (F^c + F^p)(q), \quad D(q) = (D^c + D^p)(q) \tag{16b}$$

$$C_1(q, \dot{q}) = (C_1^c + C_1^p)(q, \dot{q}), \tag{16c}$$

$$C_2(q, \dot{q}) = (C_2^c + C_2^p)(q, \dot{q}). \tag{16d}$$

Dynamics Separation

Furthermore, let

$$M = \underbrace{\begin{bmatrix} \mathcal{H} & 0 \\ 0 & 0 \end{bmatrix}}_{M^c(q)} + \underbrace{\begin{bmatrix} 0 & \mathcal{H}_{\text{slow}}^{\text{fast}} \\ \mathcal{H}_{\text{slow}}^{\text{fast}\,\top} & \mathcal{H}_{\text{slow}} \end{bmatrix}}_{M^p(q)}, \tag{17}$$

where $\mathcal{H}_{\text{slow}}^{\text{fast}}$ denotes the decomposed mass of the perturbed sections of the robot relative to the core sections.

- Let robot's state, $x = [q^\top, z^\top]^\top$ decompose as $q = [q_{\text{fast}}^\top, q_{\text{slow}}^\top]^\top$ and $z = [z_{\text{fast}}^\top, z_{\text{slow}}^\top]^\top$,
- Define $\bar{M}^p = M^p/\epsilon$, and let $u = [u_{\text{fast}}^\top, u_{\text{slow}}^\top]^\top$ be the applied torque.

SoRo Dynamics Separation

$$(M^c + \epsilon \bar{M}^p)\dot{z} = s + u, \tag{18}$$

where

$$s = \begin{bmatrix} s_{\text{fast}} \\ s_{\text{slow}} \end{bmatrix} = \begin{bmatrix} F^c + N^c \text{Ad}_{g_r}^{-1} \boldsymbol{\mathcal{G}} - [C_1^c + C_2^c + D^c] z_{\text{fast}} \\ F^p + N^p \text{Ad}_{g_r}^{-1} \boldsymbol{\mathcal{G}} - [C_1^p + C_2^p + D^p] z_{\text{slow}} \end{bmatrix}. \tag{19}$$

- Since $\boldsymbol{\mathcal{H}}_{\text{fast}}$ is invertible, let

$$\bar{M}^p = \begin{bmatrix} \bar{M}_{11}^p & \bar{M}_{12}^p \\ \bar{M}_{21}^p & \bar{M}_{22}^p \end{bmatrix} \text{ and } \boldsymbol{\Delta} = \begin{bmatrix} 0 & 0 \\ \bar{M}_{21}^p \boldsymbol{\mathcal{H}}_{\text{fast}}^{-1} & 0 \end{bmatrix}. \tag{20}$$

SoRo Dynamics Separation

Premultiplying both sides by $\mathrm{I} - \epsilon\boldsymbol{\Delta}$, it can be verified that

$$\begin{bmatrix} \boldsymbol{\mathcal{H}}_{\text{fast}} & \bar{\mathrm{M}}_{12}^{p} \\ 0 & \bar{\mathrm{M}}_{22}^{p} \end{bmatrix} \begin{bmatrix} \dot{\mathrm{z}}_{\text{fast}} \\ \epsilon\dot{\mathrm{z}}_{\text{slow}} \end{bmatrix} = \begin{bmatrix} \mathrm{s}_{\text{fast}} \\ \mathrm{s}_{\text{slow}} - \epsilon\bar{\mathrm{M}}_{21}^{p}\boldsymbol{\mathcal{H}}_{\text{fast}}^{-1}\mathrm{s}_{\text{fast}} \end{bmatrix} + \begin{bmatrix} \mathrm{u}_{\text{fast}} \\ \mathrm{u}_{\text{slow}} - \epsilon\bar{\mathrm{M}}_{21}^{p}\boldsymbol{\mathcal{H}}_{\text{fast}}^{-1}\mathrm{u}_{\text{fast}} \end{bmatrix} \tag{21}$$

which is in the standard singularly perturbed form (10):

$$\dot{\mathrm{z}}_1 = \mathrm{f}(\mathrm{z}_1, \mathrm{z}_2, \epsilon, \mathrm{u}_s, t), \;\; \mathrm{z}_1(t_0) = \mathrm{z}_1(0), \;\; \mathrm{z}_1 \in \mathbb{R}^{6N}, \tag{22a}$$

$$\epsilon\dot{\mathrm{z}}_2 = \mathrm{g}(\mathrm{z}_1, \mathrm{z}_2, \epsilon, \mathrm{u}_f, t), \;\; \mathrm{z}_2(t_0) = \mathrm{z}_2(0), \;\; \mathrm{z}_2 \in \mathbb{R}^{6N} \tag{22b}$$

SoRo Fast Subsystem Extraction

On the fast time scale $T = t/\epsilon$, with $dT/dt = 1/\epsilon$ so that,

$$\dot{z}_{\text{fast}} = \frac{dz_{\text{fast}}}{dt} \equiv \frac{1}{\epsilon}\frac{dz_{\text{fast}}}{dT} \triangleq \frac{1}{\epsilon} z'_{\text{fast}}$$

; and

$$\epsilon \dot{z}_{\text{slow}} = z'_{\text{slow}}.$$

Fast subdynamics:

$$z'_{\text{fast}} = \epsilon \mathcal{H}_{\text{fast}}^{-1}(s_{\text{fast}} + u_{\text{fast}}) - \mathcal{H}_{\text{fast}}^{-1} \mathcal{H}_{\text{slow}}^{\text{fast}} z'_{\text{slow}}, \tag{23a}$$

$$z'_{\text{slow}} = \mathcal{H}_{\text{slow}}^{-1}(s_{\text{slow}} - u_{\text{slow}}) - \mathcal{H}_{\text{fast}}^{-1}(s_{\text{fast}} - u_{\text{fast}}) \tag{23b}$$

where the slow variables are frozen on this fast time scale.

SoRo Slow Subsystem Extraction

- We let $\epsilon \to 0$ in (21), so that what is left, i.e.,

$$\dot{z}_{\text{slow}} = \mathcal{H}_{\text{slow}}^{-1}(s_{\text{slow}} + u_{\text{slow}}) \tag{24}$$

constitutes the system's slow dynamics; where the fast components are frozen on this slow time scale.

This page is left blank intentionally

Control of the Fast Strain Subdynamics

- Consider the transformation: $\begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\phi} \end{bmatrix} = \begin{bmatrix} \mathrm{q}_{\text{fast}} \\ \mathrm{z}_{\text{fast}} \end{bmatrix}$ so that $\boldsymbol{\theta}' = \epsilon \mathrm{z}_{\text{fast}} \triangleq \boldsymbol{\nu} :=$ A virtual input.
- Let $\{\mathrm{q}^d_{\text{fast}}, \dot{\mathrm{q}}^d_{\text{fast}}\} = \{\boldsymbol{\xi}^d_1, \ldots, \boldsymbol{\xi}^d_{n_\xi}, \boldsymbol{\eta}^d_1, \ldots, \boldsymbol{\eta}^d_{n_\xi}\}_{\text{fast}}$ be the desired joint space configuration for the fast subsystem.

Theorem 3 (Molu (2024))

The control law

$$\mathrm{u}_{fpos} = \mathrm{q}^d_{fast}(t_f) - \mathrm{q}_{fast}(t_f) + \mathrm{q}'^d_{fast}(t_f)$$

is sufficient to guarantee an exponential stability of the origin of $\boldsymbol{\theta}' = \boldsymbol{\nu}$ *such that for all* $t_f \geq 0$, $\mathrm{q}_{fast}(t_f) \in S$ *for a compact set* $S \subset \mathbb{R}^{6N}$. *That is,* $\mathrm{q}_{fast}(t_f)$ *remains bounded as* $t_f \to \infty$.

Control of the Fast Strain Subdynamics

Proof Sketch 1 (Proof of Theorem 3)

$$e_1 = \boldsymbol{\theta} - q^d_{fast}, \implies e_1' = \boldsymbol{\theta}' - q'^d_{fast} \triangleq \boldsymbol{\nu} - q'^d_{fast}. \tag{25}$$

$$\textit{Choose } V_1(e_1) = \frac{1}{2} e_1^\top K_p e_1 \tag{26}$$

$$\textit{Then, } V_1' = e_1^\top K_p e_1' = e_1^\top K_p(\boldsymbol{\nu} - q'^d_{fast}). \tag{27}$$

For $\boldsymbol{\nu} = q'^d_{fast} - e_1$, $V_1' = -e_1 K_p e_1 \leq 2V_1$.

Stability Analysis of the Fast Velocity Subdynamics

Theorem 4 (Molu (2024))

Under the tracking error $e_2 = \phi - \nu$ *and matrices* $(K_p, K_q) = (K_p^\top, K_q^\top) > 0$, *the control input*

$$u_{fvel} = \frac{1}{\epsilon}\mathcal{H}_{fast}[q''^{d}_{fast} + e_1 - 2e_2 - K_q^\top (K_q K_q^\top)^{-1} K_p e_1] + \frac{1}{\epsilon}\mathcal{H}^{fast}_{slow} z'_{slow} - s_{fast} \tag{28}$$

exponentially stabilizes the fast subdynamics (23).

Stability Analysis of Fast Velocity Subdynamics

Proof Sketch 2 (Sketch Proof of Theorem 4)

Recall from the position dynamics controller:

$$e_1' = \boldsymbol{\theta}' - q_{fast}'^{d} \triangleq z_{fast} - q_{fast}'^{d} + (\boldsymbol{\nu} - \boldsymbol{\nu}) \tag{29a}$$

$$= (\boldsymbol{\phi} - \boldsymbol{\nu}) + (\boldsymbol{\nu} - q_{fast}'^{d}) \triangleq e_2 - e_1. \tag{29b}$$

It follows that

$$e_2' = \boldsymbol{\phi}' - \boldsymbol{\nu}' = z_{fast}' + e_1' - q_{fast}''^{d} \tag{30}$$

$$= \mathcal{H}_{fast}^{-1}\left[\epsilon u_{fast} + \epsilon s_{fast} - \mathcal{H}_{slow}^{fast} z_{slow}'\right] + (e_2 - e_1) - q_{fast}''^{d}.$$

Stability Analysis of the Fast Velocity Subdynamics

Proof Sketch 3 (Sketch Proof of Theorem 4)

For diagonal matrices K_p, K_q *with positive damping, let us choose the Lyapunov candidate function*

$$V_2(e_1, e_2) = V_1 + \frac{1}{2} e_2^\top K_q e_2 = \frac{1}{2}[e_1 \;\; e_2] \begin{bmatrix} K_p & 0 \\ 0 & K_q \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}.$$

If $\tilde{q}_{fast} = q_{fast} - q^d_{fast}$ *and* $\tilde{q}'_{fast} = q'_{fast} - q'^d_{fast}$*, then the controller*

$$u_{fvel} = \frac{1}{\epsilon} \mathcal{H}_{fast} [q''^d_{fast} - \tilde{q}_{fast} - 2\tilde{q}'_{fast} - K_q^\top (K_q K_q^\top)^{-1} K_p \tilde{q}_{fast}] + \frac{1}{\epsilon} \mathcal{H}^{fast}_{slow} z'_{slow} - s_{fast},$$

exponentially stabilizes the system;

Stability Analysis of the Fast Velocity Subdynamics

Proof Sketch 4 (Sketch Proof of Theorem 4)

since it can be verified that

$$\begin{aligned} V_2' &= e_1^\top K_p (e_2 - e_1) \\ &\quad - e_2^\top K_q \left(e_2 - K_q^\top (K_q K_q^\top)^{-1} K_p e_1 \right) && (31a) \\ &= -e_1^\top K_p e_1 - e_2^\top K_q e_2 && (31b) \\ &\triangleq -2V_2 \leq 0. && (31c) \end{aligned}$$

Stability analysis of the slow subdynamics

Set $e_3 = z_{slow} - \boldsymbol{\nu}$ so that $\dot{e}_3 = \dot{z}_{slow} - \dot{\boldsymbol{\nu}}$. Then,

$$\dot{e}_3 = \dot{z}_{slow} - \ddot{q}^d_{fast} + (e_2 - e_1), \tag{32a}$$

$$= \boldsymbol{\mathcal{H}}^{-1}_{slow}(s_{slow} + u_{slow}) - \ddot{q}^d_{fast} + (e_2 - e_1). \tag{32b}$$

Theorem 5

The control law

$$u_{slow} = \boldsymbol{\mathcal{H}}_{slow}(e_1 - e_2 - e_3 + \ddot{q}^d_{fast}) - s_{slow} \tag{33}$$

exponentially stabilizes the slow subdynamics.

Stability analysis of the slow subdynamics

Proof.

Consider the Lyapunov function candidate

$$V_3(e_3) = \frac{1}{2} e_3^\top K_r e_3 \text{ where } K_r = K_r^\top > 0. \tag{34}$$

It follows that

$$\dot{V}_3(e_3) = e_3^\top K_r \dot{e}_3 \tag{35a}$$

$$= e_3^\top K_r \left[\boldsymbol{\mathcal{H}}_{slow}^{-1} (s_{slow} + u_{slow}) - \ddot{q}^d_{fast} + e_2 - e_1 \right]. \tag{35b}$$

Substituting u_{slow} in (33), it can be verified that

$$\dot{V}_3(e_3) = e_3^\top K_r e_3 \triangleq -2V_3(e_3) \leq 0. \tag{36}$$

Hence, the controller (33) stabilizes the slow subsystem. □

Stability of the singularly perturbed interconnected system

Let $\varepsilon = (0, 1)$ and consider the composite Lyapunov function candidate $\Sigma(z_{fast}, z_{slow})$ as a weighted combination of V_2 and V_3 i.e. ,

$$\Sigma(z_{fast}, z_{slow}) = (1-\varepsilon)V_2(z_{fast}) + \varepsilon V_3(z_{slow}), \; 0 < \varepsilon < 1. \tag{37}$$

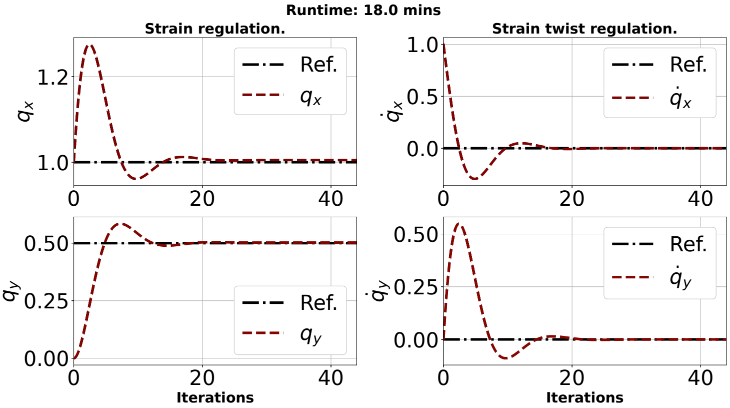
It follows that,

$$\begin{aligned}\dot{\Sigma}(z_{fast}, z_{slow}) &= (1-\varepsilon)[e_1^\top K_p \dot{e}_1 + e_2^\top K_q \dot{e}_2] + \varepsilon e_3^\top K_r \dot{e}_3, \\ &= -2(V_2 + V_3) + 2\varepsilon V_2 \leq 0\end{aligned} \tag{38}$$

which is clearly negative definite for any $\varepsilon \in (0, 1)$. Therefore, we conclude that the origin of the singularly perturbed system is asymptotically stable under the control laws.

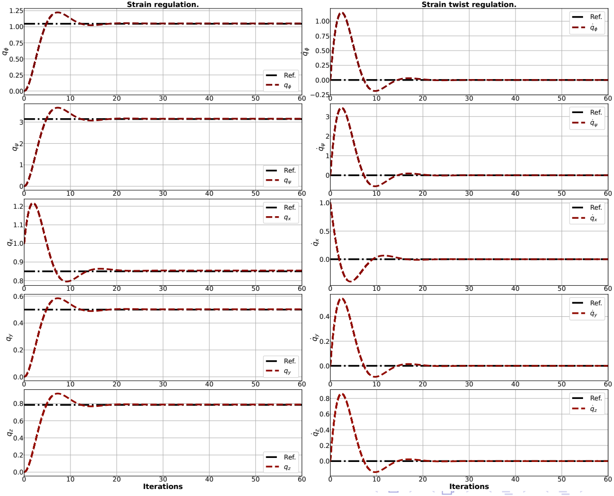
$$u(z_{fast}, z_{slow}) = (1-\varepsilon)u_{fast} + \varepsilon u_{slow}. \tag{39}$$

Asynchronous, time-separated control



Ten discretized PCS sections: 6 fast, 4 slow subsections. $\mathcal{F}_p^y = 10\,N$, with $K_p = 10$, $K_d = 2.0$ for $\boldsymbol{\eta}^d = [0, 0, 0, 1, 0.5, 0]^\top$ and $\boldsymbol{\xi}^d = 0_{6\times 1}$.

Five-axes control



Time Response Comparison with Non-hierarchical Controller

Pieces			Runtime (mins)	
Total	Fast	Slow	Hierarchical SPT (mins)	Single-layer PD control (hours)
6	4	2	18.01	51.46
8	5	3	30.87	68.29
10	7	3	32.39	107.43

Table: Time to Reach Steady State.

Contributions

- Layered singularly perturbed techniques for decomposing system dynamics to multiple timescales.
- Stabilizing nonlinear backstepping controllers were introduced to the respective subdynamics for fast strain regulation.

Discussions

- Leverage the *multiphysics of (often) heterogeneous soft material components*;
- Neat manipulation strategies for motion is a *multiscale problem* that requires imbuing geometric mathematical reasoning into the control strategies for desired movements.
- Challenge: Merging the long-term planning horizon of spatial perception tasks with the *fast time-constant* (typically milliseconds or microseconds) requirements of the precise control of soft, compliant pneumatic/mechanical systems across multiple time-scales;

Discussions

- Process spatial information (Lagrangian) often within a long-time horizon context (Eulerian) for the real-time control or planning across multiple time-scales.

Conclusion

- Email: lekanmolu@microsoft.com
- Thank you!

References I

Cosimo Della Santina, Christian Duriez, and Daniela Rus. Model-based control of soft robots: A survey of the state of the art and open challenges. *IEEE Control Systems Magazine*, 43(3):30–65, 2023. doi: 10.1109/MCS.2023.3253419.

Derek E Moulton, Thomas Lessinnes, and Alain Goriely. Morphoelastic Rods III: Differential Growth and Curvature Generation in Elastic Filaments. *Journal of the Mechanics and Physics of Solids*, 142:104022, 2020.

Bartosz Kaczmarski, Alain Goriely, Ellen Kuhl, and Derek E Moulton. A Simulation Tool for Physics-informed Control of Biomimetic Soft Robotic Arms. *IEEE Robotics and Automation Letters*, 2023.

Mattia Gazzola, LH Dudte, AG McCormick, and Lakshminarayanan Mahadevan. Forward and inverse problems in the mechanics of soft filaments. *Royal Society open science*, 5(6):171628, 2018.

M. B. Rubin. *Cosserat Theories: Shells, Rods, and Points*. Springer-Science+Business Medis, B.V., 2000.

Eugène Maurice Pierre Cosserat and François Cosserat. *Théorie des corps déformables*. A. Hermann et fils, 1909.

Isuru S Godage, David T Branson, Emanuele Guglielmino, Gustavo A Medrano-Cerda, and Darwin G Caldwell. Shape function-based kinematics and dynamics for variable length continuum robotic arms. In *2011 IEEE International Conference on Robotics and Automation*, pages 452–457. IEEE, 2011.

Robert J. III Webster and Bryan A. Jones. Design and kinematic modeling of constant curvature continuum robots: A review. *The International Journal of Robotics Research*, 29(13):1661–1683, 2010.

Ke Qiu, Jingyu Zhang, Danying Sun, Rong Xiong, Haojian Lu, and Yue Wang. An efficient multi-solution solver for the inverse kinematics of 3-section constant-curvature robots. *arXiv preprint arXiv:2305.01458*, 2023.

Federico Renda, Vito Cacucciolo, Jorge Dias, and Lakmal Seneviratne. Discrete cosserat approach for soft robot dynamics: A new piece-wise constant strain model with torsion and shears. *IEEE International Conference on Intelligent Robots and Systems*, 2016-Novem:5495–5502, 2016. ISSN 21530866.

References II

Federico Renda, Frédéric Boyer, Jorge Dias, and Lakmal Seneviratne. Discrete cosserat approach for multisection soft manipulator dynamics. *IEEE Transactions on Robotics*, 34(6):1518–1533, 2018.

Federico Renda, Michele Giorelli, Marcello Calisti, Matteo Cianchetti, and Cecilia Laschi. Dynamic model of a multibending soft robot arm driven by cables. *IEEE Transactions on Robotics*, 30(5):1109–1122, 2014.

José Guadalupe Romero, Romeo Ortega, and Ioannis Sarras. A globally exponentially stable tracking controller for mechanical systems using position feedback. *IEEE Transactions on Automatic Control*, 60(3):818–823, 2014.

Lekan Molu and Shaoru Chen. Lagrangian Properties and Control of Soft Robots Modeled with Discrete Cosserat Rods. In *IEEE International Conference on Decision and Control, Milan, Italy*. IEEE, 2024.

Nikolai Matni, Aaron D Ames, and John C Doyle. A quantitative framework for layered multirate control: Toward a theory of control architecture. *IEEE Control Systems Magazine*, 44(3):52–94, 2024.

Lekan Molu. Fast Whole-Body Strain Regulation in Continuum Robots. *(submitted to) American Control Conference*, 2024.