Enhancing Robustness in Reinforcement Learning

Peter Seiler University of Minnesota



November 16, 2018 Seminar at United Technologies Research Lab





Jordan Hoyt Parul Singh <u>Wind Energy</u>



Raghu Venkataraman <u>Safety Critical Systems</u>



Abhineet Gupta <u>Aeroelasticity</u>



Robust Control Design and Analysis

Chris Regan Curt Olson Brian Taylor Harish Venkataraman Jyot Buch

Past: What is the impact of model uncertainty and nonlinearities on feedback system?

Past: What is the impact of model uncertainty and nonlinearities on feedback system?

Key contributions

- 1. Theoretical connections between frequency domain and timedomain (dissipation inequality) analysis methods
- 2. Tools for uncertain time-varying and gain-scheduled systems
- 3. Applications to wind energy, UAVs, flex aircraft, hard disk drives
- Numerically reliable algorithms with transition to Matlab's Robust Control Toolbox

Past: What is the impact of model uncertainty and nonlinearities on feedback system?

Key contributions

- 1. Theoretical connections between frequency domain and timedomain (dissipation inequality) analysis methods
- 2. Tools for uncertain time-varying and gain-scheduled systems
- 3. Applications to wind energy, UAVs, flex aircraft, hard disk drives
- Numerically reliable algorithms with transition to Matlab's Robust Control Toolbox

Future: What is the impact of model uncertainty on control systems designed via data-driven methods?

Outline

- Flutter Suppression on Flexible Aircraft
- Robustness of Time-Varying Systems
- Robustness in Reinforcement Learning

Outline

- Flutter Suppression on Flexible Aircraft
- Robustness of Time-Varying Systems
- Robustness in Reinforcement Learning

Fuel Consumption of Commercial Airliners



Fuel Efficient Aircraft Design

Breguet Range Equation

Range =
$$V I_{sp} \frac{L}{D} \ln\left(\frac{m_{\text{takeoff}}}{m_{\text{landing}}}\right)$$

- Improve fuel efficiency by
 - Reducing structural mass
 - Reducing drag with longer, more slender wings (high aspect ratio)
 - Improving engine efficiency

Fuel Efficient Aircraft Design

Breguet Range Equation

Range =
$$V I_{sp} \frac{L}{D} \ln\left(\frac{m_{\text{takeoff}}}{m_{\text{landing}}}\right)$$

- Improve fuel efficiency by
 - Reducing structural mass
 - Reducing drag with longer, more slender wings (high aspect ratio)
 - Improving engine efficiency
- Adverse effects
 - increased coupling of structural dynamics and rigid body motion
 - increased coupling of aerodynamic loads and structural deformation
 - reduced flutter margins

Flutter



Source: NASA Dryden Flight Research



Classical Approach



Flexible Aircraft Challenges



Flexible Aircraft Challenges





Recent Flight Demonstrators

- Lockheed Martin/AFRL: Body Freedom Flutter (BFF)
 - Ref: Burnett, et al, AIAA MST Conference, 2010-7780
 - Ref: Holm-Hansen, et al, AUVSI, 2010



Recent Flight Demonstrators

- Lockheed Martin/AFRL: Body Freedom Flutter (BFF)
- NASA/Lockheed Martin: X-56A Multi-Utility Tech. Testbed (MUTT)
 - Ref: Schaefer, ACGSC, 2018



Recent Flight Demonstrators

- Lockheed Martin/AFRL: Body Freedom Flutter (BFF)
- NASA/Lockheed Martin: X-56A Multi-Utility Tech. Testbed (MUTT)
- Flutter Free FLight Envelope eXpansion for ecOnomical Performance improvement (FlexOp)
 - EU 2020 Horizon Project with B. Vanek as PI (MTA-Sztaki in Hungary)
 - Project Site: <u>https://flexop.eu/</u>



Performance Adaptive Aeroelastic Wing (PAAW)

- Goal: Suppress flutter, control wing shape and alter shape to optimize performance
 - Funding: NASA NRA with Dr. Jeffrey Ouellette as Tech. Monitor
 - Team: UMN, VT, STI, CMSoft, Aurora, Schmidt & Assoc.
 - Project Site: paaw.net





mAEWing1: BFF Replica





mAEWing1 Sensor/Actuator Configuration



Ref: Regan & Taylor, AIAA 2016-1747



[Cooper & Wright 2015; Collar 1946]

mAEWing1 Models: NASTRAN (VT), CFD/CSD (CMSoft), IO Reduced Order Model (STI/CMSoft), Flight Dynamics (Schmidt)

- 1. VT: Construct MSC NASTRAN model
 - Ref: Schmidt, Zhao, Kapania, AIAA 2016-1748
 - Finite-element model with rod / beam elements & unsteady aerodynamic model with double lattice.
 - Initial model from CAD and simple static test data from UMN

- VT: Construct MSC NASTRAN model 1
- VT/UMN: Update NASTRAN FEM with ground test data 2.
 - Ref: Gupta, Seiler, Danowsky, AIAA 2016-1753
 - Matlab Demo: "Modal Analysis of a Flexible Flying Wing Aircraft", Demonstrates frequency domain fitting in System ID Toolbox



- 1. VT: Construct MSC NASTRAN model
- 2. VT/UMN: Update NASTRAN FEM with ground test data
- 3. VT: Obtain mode shapes & frequencies from NASTRAN
 - Ref: Schmidt, Zhao, Kapania, AIAA 2016-1748



- 1. VT: Construct MSC NASTRAN model
- 2. VT/UMN: Update NASTRAN FEM with ground test data
- 3. VT: Obtain mode shapes & frequencies from NASTRAN
- 4. Schmidt: Construct low-order flight dynamics model
 - Ref: Schmidt, Zhao, Kapania, AIAA 2016-1748
 - Ref: Schmidt, Journal of Aircraft, 2016.
 - Parameter-varying model constructed using mean-axes
 - Model has longitudinal rigid body dyn. & three elastic modes

$$\dot{x} = A(V_{\infty})x + B(V_{\infty})u$$
$$y = C(V_{\infty})x + D(V_{\infty})u$$

where $x = [u \ \alpha \ \theta \ q \mid \eta_1 \ \dot{\eta}_1 \ \eta_2 \ \dot{\eta}_2 \ \eta_3 \ \dot{\eta}_3]^T$

- 1. VT: Construct MSC NASTRAN model
- 2. VT/UMN: Update NASTRAN FEM with ground test data
- 3. VT: Obtain mode shapes & frequencies from NASTRAN
- 4. Schmidt: Construct low-order flight dynamics model
- 5. STI/UMN/Schmidt: Grey-box ID from flight tests
 - Ref: Danowsky, Schmidt, Pfifer, AIAA 2017-1394



Bode mag (dB) from symmetric L3/R3 and L4/R4 to center body accel

- 1. VT: Construct MSC NASTRAN model
- 2. VT/UMN: Update NASTRAN FEM with ground test data
- 3. VT: Obtain mode shapes & frequencies from NASTRAN
- 4. Schmidt: Construct low-order flight dynamics model
- 5. STI/UMN/Schmidt: Grey-box ID from flight tests
- 6. UMN: Component Modeling
 - Ref: Theis, Pfifer, Seiler, AIAA 2016-1751



Open-Loop Flutter at ~30m/s

Model Predictions:

- Flight Dynamics (Schmidt): 29.1 m/s
- NASTRAN (VT): 29.5 m/s
- CFD/CSD (CMSoft): 30.8 m/s
- Input/Output Reduced Order Model (STI/CMSoft): 31.7m/s



Open-Loop Flutter at ~30m/s

Mode Shape: Coupling of rigid body short period and 1st symmetric wing bending



Active Flutter Suppression

- Mixed sensitivity H_{∞} Loopshaping
 - Ref: Theis, Pfifer, Seiler, AIAA 2016-1751
 - Ref: Theis, Ph.D., 2018



Active Flutter Suppression

- Modal velocity as performance output
- Bandpass penalty on control effort





Closed-Loop Evaluation



Closed-Loop Evaluation



Open Loop

flutter beyond 30 m/s airspeed



Closed Loop

envelope expansion to 43 m/s airspeed

Closed-Loop Flight Tests

- Three controllers designed to increase damping to BFF mode at 23m/s
 - Hinf Controller (Retuned): Kotikalpudi, et al, AIAA 2018-3426)
 - MIDAAS: Danowsky (STI), 2017-4353
 - Classical Controller: Schmidt, Journal of GCD, 2016.

Closed-Loop Flight Tests

- Three controllers designed to increase damping to BFF mode at 23m/s
 - Hinf Controller (Retuned): Kotikalpudi, et al, AIAA 2018-3426)
 - MIDAAS: Danowsky (STI), 2017-4353
 - Classical Controller: Schmidt, Journal of GCD, 2016.
- Flight Tests
 - Ref: Danowsky, Kotikalpudi, Schmidt, Regan, Seiler, AIAA 2018-3427
 - Controllers tested at and above the designed airspeed.
 - All controllers added damping at the designed speed.
 - MIDAAS & classical designs flown above open-loop flutter speed.
 - Hinf controller did not increase flutter speed but this was an artifact of our design objective and flight test plan.

Pole Map for Retuned H_{∞} Controller



35

Flight Test Summary



Next Steps

- Robust Flutter Speed (RFS): Airspeeds where active flutter control has 6dB/45deg margins on all inputs & outputs
 - Metric for safe flight envelope with active flutter suppression
 - Current: Restrict envelope to 20% below (open-loop) flutter speed

Next Steps

- Robust Flutter Speed (RFS): Airspeeds where active flutter control has 6dB/45deg margins on all inputs & outputs
 - Metric for safe flight envelope with active flutter suppression
 - Current: Restrict envelope to 20% below (open-loop) flutter speed
- Redesign all 3 controllers to maximize robust flutter speed
 - Design complicated by second bending mode at higher speeds
- Preliminary Results:
 - H_{∞} control achieves robust/absolute flutter speeds of 43/41 m/s
 - Similar but slightly lower speeds for MIDAAS & classical designs
 - Tested in linear parameter varying sim with actuator limits.
 - Flight tests planned for spring 2019

Outline

- Flutter Suppression on Flexible Aircraft
- Robustness of Time-Varying Systems
 - Joint work with M. Arcak, A. Packard, M. Moore, and C. Meissen at UC, Berkeley.
 - Funded by ONR BRC with B. Holm-Hansen at Tech. Monitor
- Robustness in Reinforcement Learning

Time-Varying Systems





Wind Turbine Periodic / Parameter-Varying

Flexible Aircraft Parameter-Varying

Vega Launcher Time-Varying (Source: ESA) **Robotics** Time-Varying (Source: ReWalk)

Issue: Few numerically reliable methods to assess the robustness of time-varying systems.

Finite Horizon Analysis

Goal: Assess the robustness of linear time-varying (LTV) systems on finite horizons.

Issue: Classical Gain/Phase Margins focus on (infinite horizon) stability and frequency domain concepts.

Instead focus on:

- Finite horizon metrics, e.g. induced gains and reachable sets.
- Effect of disturbances and model uncertainty (D-scales, IQCs, etc).
- Time-domain analysis conditions.



(Nominal) Finite Horizon Analysis

Nominal System

$$\dot{x}(t) = A(t) x(t) + B(t) d(t)$$
$$e(t) = C(t)x(t)$$
$$x(0) = 0$$



Analysis Objective

Derive bound on $||e(T)||_2$ that holds for all disturbances $||d||_{2,[0,T]} \le 1$ on the horizon [0,T].

Nominal Analysis with Dissipation Inequalities

Theorem [1]

If there exists $P(t) = P(t)^T$ such that $P(T) = C(T)^T C(T)$ and $V(x,t) \coloneqq x^T P(t) x$ satisfies $\frac{d}{dt} V(x,t) - \gamma^2 d(t)^T d(t) \le 0 \quad \forall t \in [0,T]$ then $\|e(T)\|_2 \le \gamma$

Proof

Integrate from t = 0 to t = T: $\underbrace{V(x,T)}_{e(T)^{T}e(T)} - \underbrace{V(x,0)}_{=0} \le \gamma^{2} \int_{0}^{T} d(t)^{T} d(t) dt$

[1] Moore, Finite Horizon Robustness Analysis using IQCs, MS Thesis, Berkeley, 2015.

[2] Willems, Dissipative Dynamical Systems: Parts i and ii, 1972.

[3] Green & Limebeer, Linear Robust Control, 1995.

Nominal Analysis with Dissipation Inequalities

Theorem [1]

If there exists $P(t) = P(t)^T$ such that $P(T) = C(T)^T C(T)$ and $V(x,t) \coloneqq x^T P(t) x$ satisfies $\frac{d}{dt} V(x,t) - \gamma^2 d(t)^T d(t) \le 0 \quad \forall t \in [0,T]$ then $\|e(T)\|_2 \le \gamma$

Comments

• Time-varying matrix inequality can be "solved" via convex optimization using basis functions for P(t), gridding on time.

• Extensions to parameter varying [4] and periodic [5] systems.

[1] Moore, Finite Horizon Robustness Analysis using IQCs, MS Thesis, Berkeley, 2015.

[2] Willems, Dissipative Dynamical Systems: Parts i and ii, 1972.

[3] Green & Limebeer, Linear Robust Control, 1995.

[4] Wu. Control of Linear Parameter Varying Systems. PhD thesis, Berkeley, 1995.

[5] Bittanti & Colaneri, Periodic Systems. Springer, 2009.

(Robust) Finite-Horizon Analysis

Uncertain Time-Varying Model

$$\dot{x}(t) = A(t) x(t) + B_1(t) w(t) + B_2(t) d(t)$$

$$v(t) = C_1(t) x(t)$$

$$e(t) = C_2(t) x(t)$$

$$x(0) = 0$$



Analysis Objective

Derive bound on $||e(T)||_2$ that holds for all disturbances $||d||_{2,[0,T]} \leq 1$ and uncertainties $\Delta \in \Delta$ on the horizon [0,T].

Example: Passive Uncertainty



Pointwise Quadratic Constraint

Robust Finite Horizon Analysis

Theorem [1]

Assume x(0) = 0, $||d||_{2,[0,T]} \le 1$, and Δ is pointwise passive. If there exists $P(t) = P(t)^T$ such that $P(T) = C(T)^T C(T)$ and $V(x,t) = x(t)^T P(t)x(t)$ satisfies: $\frac{d}{dt}V(x,t) - \gamma^2 d(t)^T d(t) + {v(t) \ w(t)}^T {0 \ I \ 0} {v(t) \ w(t)} \le 0 \quad \forall t \in [0,T]$ then $||e(T)||_2 \le \gamma$.

Proof

Integrate from t = 0 to t = T to show: $\underbrace{V(x,T)}_{e(T)^T e(T)} - \underbrace{V(x,0)}_{=0} + \underbrace{\int_0^T \begin{bmatrix} v(t) \\ w(t) \end{bmatrix}^T \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \begin{bmatrix} v(t) \\ w(t) \end{bmatrix} dt}_{\ge 0} \le \gamma^2 \int_0^T d(t)^T d(t) dt$

[1] Moore, Finite Horizon Robustness Analysis using IQCs, MS Thesis, Berkeley, 2015.

More General Results

- Approach
 - Combine dissipation inequalities (Willems) and more general integral quadratic constraint framework (Megretski & Rantzer)
 - Requires technical factorization result to connect incorporate frequency domain IQCs into this time-domain analysis [1,2].
 - Developed numerical algorithms that combine differential linear matrix inequalities and Riccati differential equations [3].
- Extensions
 - Similar methods can be used to assess robustness of linear parameter varying (LPV) [4,5] and periodic systems [6].

References

[1] Seiler, Stability Analysis with Dissipation Inequalities and Integral Quadratic Constraints, TAC, 2015.

[2] Hu, Lacerda, Seiler, Rob. Analysis of Uncertain Discrete-Time Sys. with Dissipation Ineq. and IQCs, IJRNC, 2016.
 [3] Seiler, Moore, Meissen, Arcak, Packard, Finite Horizon Robustness Analysis of LTV Systems Using Integral Quadratic Constraints, arXiv / Automatica, 2018.

[4] Pfifer & Seiler, Less Conservative Robustness Analysis of LPV Systems Using IQCs, IJRNC, 2016.
[5] Hjartarson, Packard, & Seiler, LPVTools: A Toolbox for Modeling, Analysis, and Synthesis of LPV Systems, '15.
[6] Fry, Farhood, & Seiler, IQC-based robustness analysis of discrete-time LTV systems, IJRNC 2017.

Two-Link Robot Arm



Nonlinear dynamics [MZS]: $\dot{\eta} = f(\eta, \tau, d)$ where $\eta = [\theta_1, \dot{\theta}_1, \theta_2, \dot{\theta}_2]^T$ $\tau = [\tau_1, \tau_2]^T$ $d = [d_1, d_2]^T$

Two-Link Diagram [MZS]

au and d are control torques and disturbances at the link joints.

[MZS] R. Murray, Z. Li, and S. Sastry. A Mathematical Introduction to Robot Manipulation, 1994.

Nominal Trajectory (Cartesian Coords.)





Analysis

Nonlinear dynamics:

$$\dot{\eta} = f(\eta, \tau, d)$$

Linearize along a (finite –horizon) trajectory $(\bar{\eta}, \bar{\tau}, d = 0)$ $\dot{x} = A(t)x + B(t)u + B(t)d$

Design finite-horizon state-feedback LQR gain.

Goal: Compute bound on the final position accounting for disturbances and LTI uncertainty with $\|\Delta\| \le 0.8$ injected at 2nd joint.



Effect of Disturbance / Uncertainty



- Bound on final position computed in 102sec.
- Numerically robust algorithm to construct the worst-case disturbance
- Ref: Iannelli, Seiler & Marcos, submitted to AIAA JGCD.

Outline

- Flutter Suppression on Flexible Aircraft
- Robustness of Finite Horizon Systems
- Robustness in Reinforcement Learning

Benchmark: Robotic Walking on MuJoCo

- Training can exploit flaws in the simulator.
 - B. Recht, arXiv, 2018
- There are many model-based methods to ensure robustness.
- **Goal:** Develop modelfree methods to ensure robustness.
 - Venkataraman & Seiler, Recovering Robustness in Model-Free Reinforcement Learning, submitted to 2019 ACC.



Linear Quadratic Gaussian (LQG)

Minimize
$$J_{LQG}(u) := \lim_{N \to \infty} \frac{1}{N} E\left[\sum_{t=0}^{N} x_t^T Q x_t + u_t^T R u_t\right]$$

Subject To:
$$x_{t+1} = Ax_t + Bu_t + B_w w_t$$

 $y_t = Cx_t + v_t$

The optimal controller has an observer/state-feedback form

$$\hat{x}_{t+1} = A\hat{x}_t + Bu_t + L\left(y_t - C\hat{x}_t\right)$$
$$u_t = -K\hat{x}_t$$

Gains (*K*,*L*) computed by solving two Riccati equations. This solution is model-based, i.e. it uses data *A*,*B*,*C*, etc

Reinforcement Learning

- Partially Observable Markov Decision Processes (POMDPs)
 - Set of states, S
 - Set of actions, A
 - Reward function, $r: S \times A \to \mathbb{R}$
 - State transition probability, T
 - Set of observations and observation probability, O
- Many methods to synthesize a control policy from input/output data to maximize the cumulative reward

$$J_{RL}(a) := E\left[\sum_{t=0}^{N} r(s_t, a_t)\right]$$

• The LQG problem is a special case of this RL formulation

Doyle's Example ('78 TAC)

- LQR state-feedback regulators have provably good margins.
- Doyle's example shows that LQG regulators can have arbitrarily small input margins.

Doyle's Example ('78 TAC)

- LQR state-feedback regulators have provably good margins.
- Doyle's example shows that LQG regulators can have arbitrarily small input margins.
- Doyle's example can also be solved within RL framework using direct policy search:

$$z_{t+1} = A_K(\theta) z_t + B_K(\theta) y_t$$
$$u_t = C_K(\theta) z_t$$

where

$$A_K(\theta) := \begin{bmatrix} 0 & \theta_1 \\ 1 & \theta_2 \end{bmatrix}, B_K(\theta) := \begin{bmatrix} 1 \\ 0 \end{bmatrix}, C_K^T(\theta) := \begin{bmatrix} \theta_3 \\ \theta_4 \end{bmatrix}$$

• RL will converge to the optimal LQG control with infinite data collection. Thus RL can also have poor margins.

Implications?

Recovering Robustness

- Increase process noise during training?
 - This causes margins to decrease on Doyle's example
 - Process noise is not model uncertainty

Recovering Robustness

- Increase process noise during training?
- Modify reward to increase state penalty or decrease control penalty?
 - Again, this causes margins to decrease on Doyle's example
 - Trading performance vs. robustness via the reward function can be difficult or counter-intuitive

Recovering Robustness

- Increase process noise during training?
- Modify reward to increase state penalty or decrease control penalty?
- Inject synthetic gain/phase variations at the plant input (and output?) during the training phase.



Results On Doyle's Example



Results on Simplified Flex System

- Model has 4-states (Rigid body and lightly damped modes)
- RL applied to 3-state controller parameterization
 - LQG controller is not in the control policy parameterization
 - Still converges to policy with small margins
 - Robustness recovered with synthetic pertubations during training



Next Steps

- How should synthetic perturbations be introduced during training?
- Can we make any rigorous claims about the proposed method?
- Attempt experimental tests on a simple system

Past: What is the impact of model uncertainty and nonlinearities on feedback system?

Key contributions

- 1. Theoretical connections between frequency domain and timedomain (dissipation inequality) analysis methods
- 2. Tools for uncertain time-varying and gain-scheduled systems
- 3. Applications to wind energy, UAVs, flex aircraft, hard disk drives
- Numerically reliable algorithms with transition to Matlab's Robust Control Toolbox

Future: What is the impact of model uncertainty on control systems designed via data-driven methods?

https://www.aem.umn.edu/~SeilerControl/

Acknowledgements

- US National Science Foundation
 - Grant No. NSF-CMMI-1254129: "CAREER: Probabilistic Tools for High Reliability Monitoring and Control of Wind Farms." Prog. Manager: J. Berg.
 - Grant No. NSF/CNS-1329390: "CPS: Breakthrough: Collaborative Research: Managing Uncertainty in the Design of Safety-Critical Aviation Systems".
 Prog. Manager: D. Corman.
- ONR
 - ONR BRC: "Finite-Horizon Robustness: Moving Beyond Traditional Stability Analysis." Tech Monitor: B. Holm-Hansen.
- NASA
 - NRA NNX14AL36A: "Lightweight Adaptive Aeroelastic Wing for Enhanced Performance Across the Flight Envelope," Tech. Monitor: J. Ouelette.
 - NRA NNX12AM55A: "Analytical Validation Tools for Safety Critical Systems Under Loss-of-Control Conditions." Tech. Monitor: C. Belcastro.
- Eolos Consortium and Saint Anthony Falls Laboratory
 - <u>http://www.eolos.umn.edu/</u> & <u>http://www.safl.umn.edu/</u>