Autonomous System Resilience and Guaranteed Performance in the Face of Unexpected Adversity

(hopefully different from all of my group's ULI-related SciTech papers)

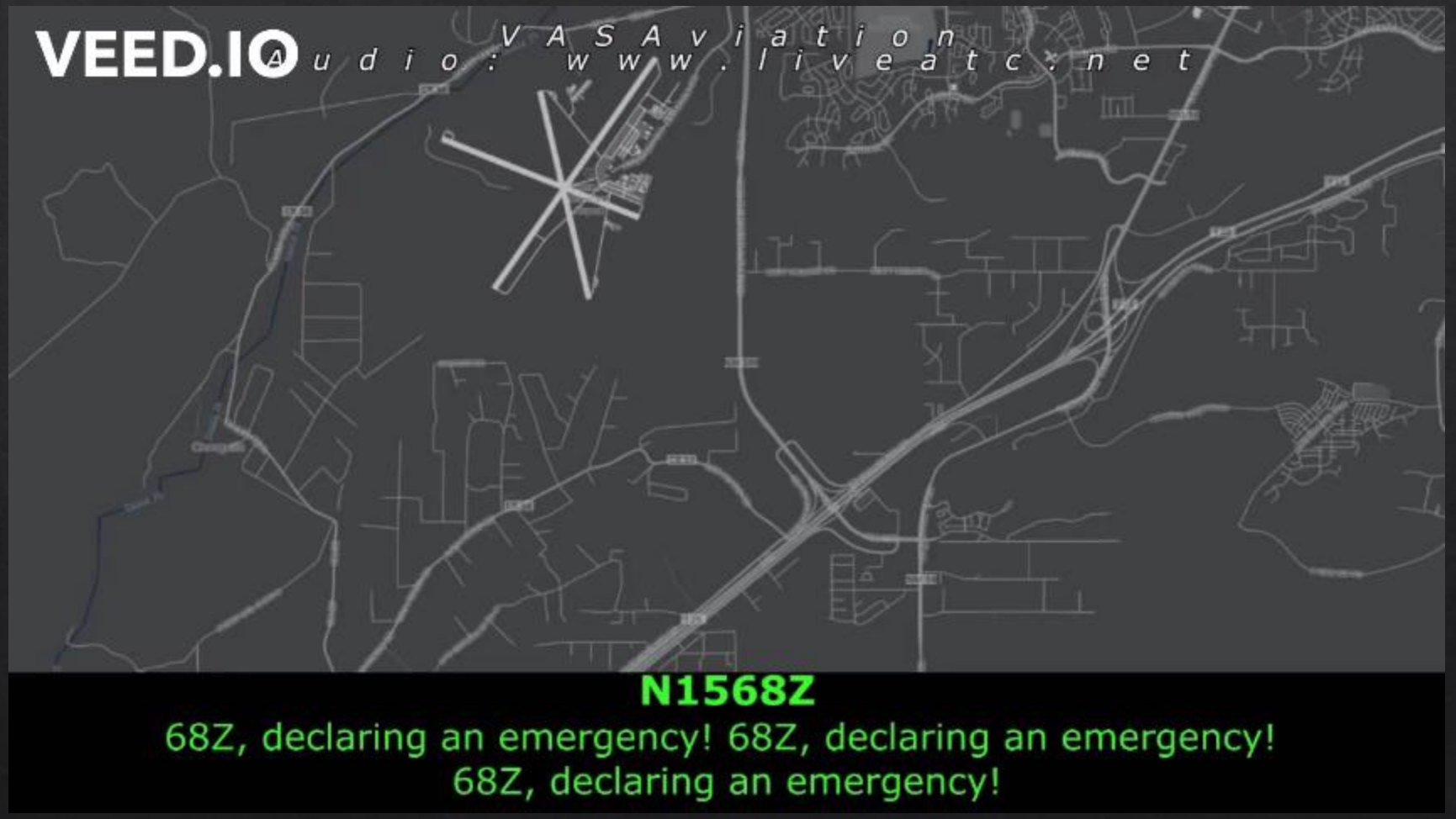
Melkior Ornik

ILLINOIS

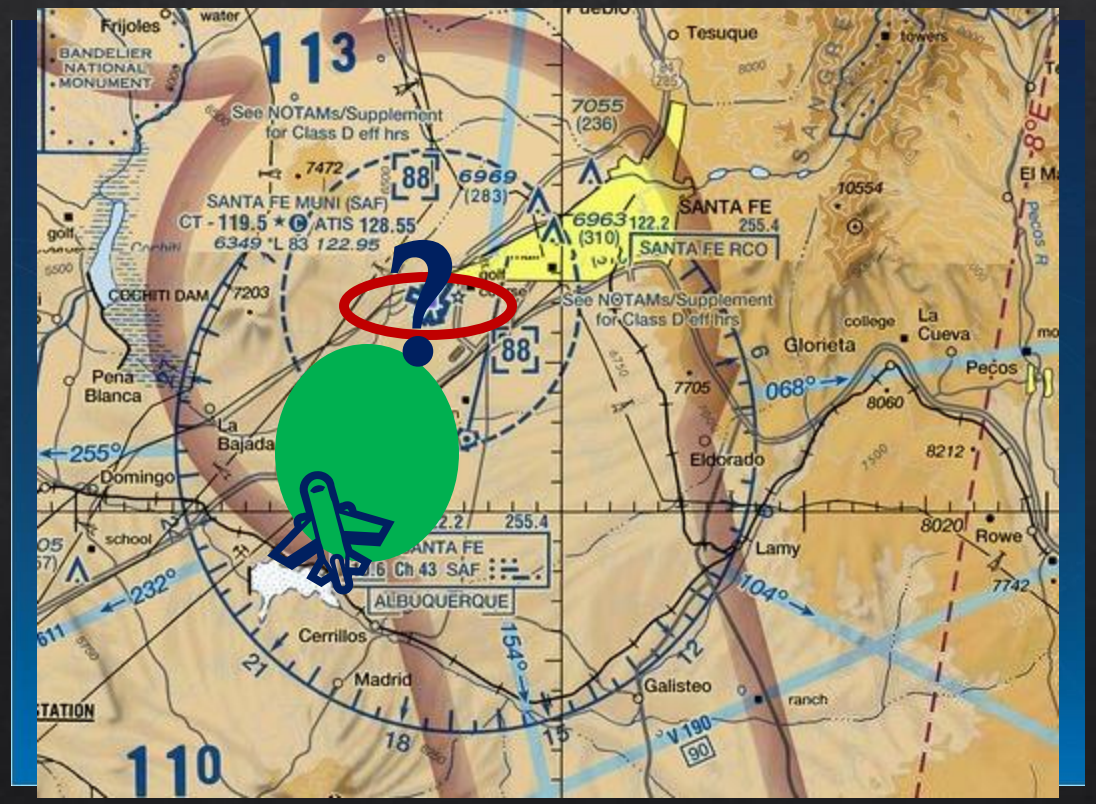
January 26, 2024

Melkior Ornik – mornik@illinois.edu

Motivating Example



Maximal Understanding in the Face of Minimal Knowledge



Adaptive / Robust Control

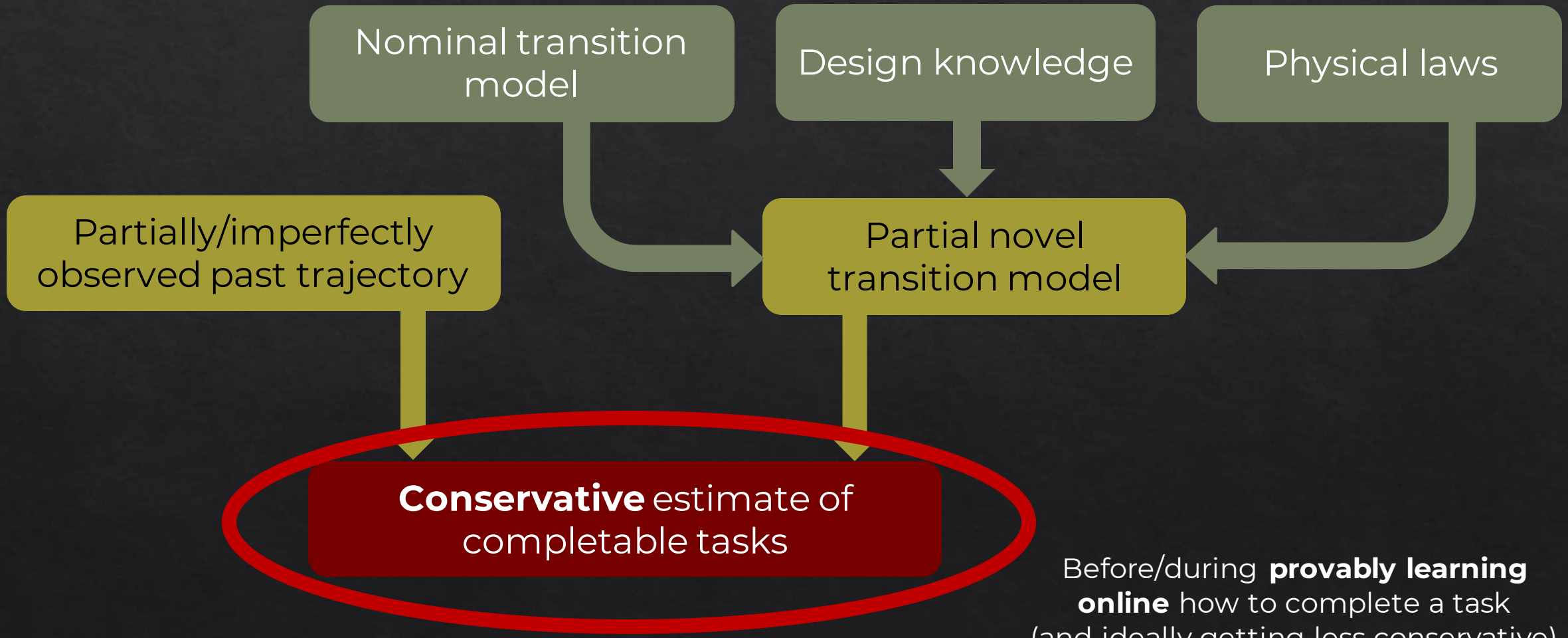
Reach the objective under (some) lack of knowledge of dynamics

Objective might not be reachable

First decide on a reachable target, *then* plan how to get there

Autopilot-in-command

Certifiable Capabilities



Before/during **provably learning online** how to complete a task (and ideally getting less conservative)

Disasters

$$\dot{x} = f(x, u), \quad u \in U$$

Change in dynamics
(e.g., physical damage)

Partial loss of control
(e.g., adversarial takeover)

Actuator
degradation

$$\dot{x} = \hat{f}(x, u, v), \quad (u, v) \in \hat{U}$$

For today: Nonlinear, continuous-time, *deterministic* (sort of – see later)

Actuator Degradation

$\Leftrightarrow \dot{x} =$ $u \in \hat{W}$

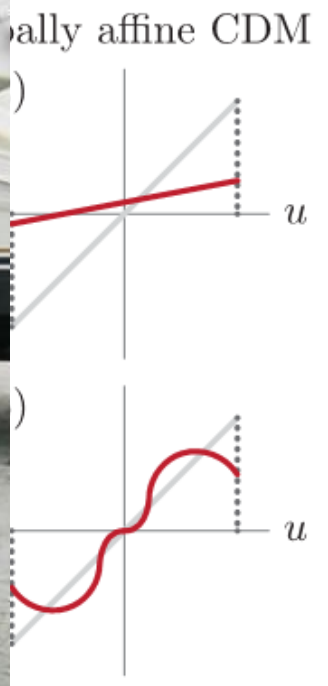
In theory **simp**

Dynamics
not exactl

reachable
set for th

its are

e reachable



H. El-Kebir, R. Berlin, J. Bentsman, M. Ornik. Viability under degraded control authority. *IEEE Control Systems Letters, 2023.*

Real-Time Reachability Analysis

In practice:

- Reachable set of an, even fully known, nonlinear system **is not computable** in real time (even meaningful **underapproximations** are difficult)
- *Operational envelope* for the nominal system is computed *beforehand*

We "know" the reachable set for the nominal system, and we know a bound on the degradation

- Our computation of the degraded reachable set **should not** start from scratch

Maximal Trajectory Difference

If an available control signal $\hat{u} : [0, T] \to \hat{U}$ differs from a desired nominal signal $u : [0, T] \to U$ by no more than ε (in some meaningful norm), then the state $\hat{x}(T)$ reached by $\hat{u}$ cannot differ from $x(T)$ reached by u by more than some $h(\varepsilon)$

How to determine $h(\varepsilon)$?

- Looks like some version of **Grönwall's lemma**, but we can do better

Assumption 1. For all $x \in \mathbb{R}^n$, $u \in \mathcal{U}$, $t_0 \leq t < \infty$ $\|h(t, x, u)\| \leq a(t)w(\|x\|, \|u\|) + b(t)$.

Let $x(t)$ be a solution to the equation $\dot{x} = h(t, x, u)$, $0 \leq t_0 \leq t < \infty$, where $h(t, x, u) : [t_0, \infty) \times \mathbb{R}^n \times \mathcal{U} \to \mathbb{R}^n$ is continuous for $t_0 \leq t < \infty$, and $\mathcal{U} \subseteq \mathbb{R}^m$ is compact and satisfies $\max_{u \in \mathcal{U}} \|u\| = \delta$. Let Assumption 1 hold. Then,

$$\|x(t)\| \leq G^{-1}\left[G\left(\|x(t_0)\| + \int_{t_0}^{t} b(\tau)\mathrm{d}\tau\right) + \int_{t_0}^{t} a(\tau)\mathrm{d}\tau\right], \quad G(r) := \int_{r_0}^{r} \frac{\mathrm{d}s}{w(s, \delta)}, \quad r > 0, r_0 > 0.$$

Wildness of system dynamics

H. El-Kebir, M. Ornik. Online inner approximation of reachable sets of nonlinear systems with diminished control authority. *2021 SIAM Conference on Control and Its Applications*, 2021.

Bihari Inequality

Intuitively, the result should depend on the “wildness” of the system dynamics

As in similar proofs derived from Grönwall’s lemma, bound the difference in trajectories by itself:

$$r(t) = f(\hat{x}(t), \hat{u}(t)) - f(x(t), u(t)),$$

$$\|\hat{x}_i(t) - x_i(t)\| = \left\| \int_0^t r_i(\tau) d\tau \right\| \leq \int_0^t a_i(\tau)\omega_i(\hat{x}(\tau) - x(\tau), \hat{u}(\tau) - u(\tau)) + b_i(\tau) d\tau$$

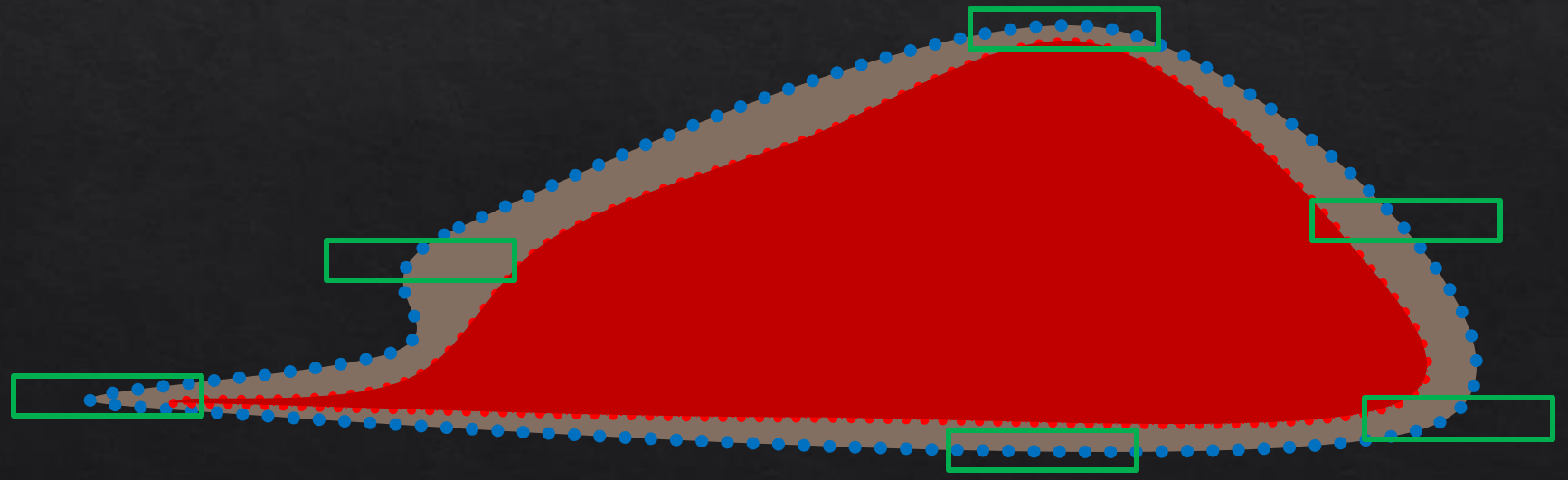
Can use Lipschitz if needed, but can do better if more information is available

H. El-Kebir, M. Ornik. Online inner approximation of reachable sets of nonlinear systems with diminished control authority. *2021 SIAM Conference on Control and Its Applications*, 2021.

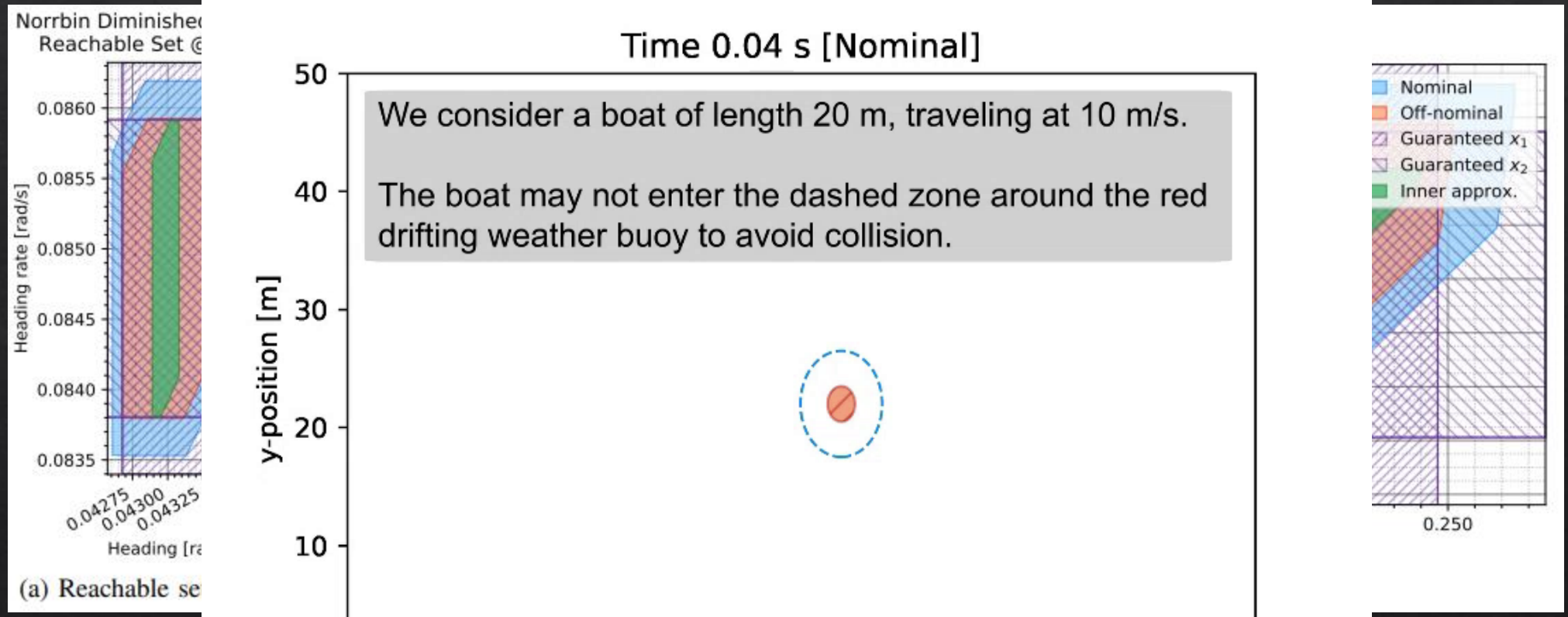
(Hyperrectangular) Tightening

Idea: if the Hausdorff distance of the nominal set of control inputs and the degraded set is bounded by M, then for each nominal control signal we can find an allowed "degraded" signal within an M-ball

Any state in the boundary of the degraded reachable set maps to a "nearby" state on the boundary of the nominal reachable set



Norrbin's Ship Steering



(Better t nates)

H. El-Kebir, A. Pirosmanishvili, M. Ornik. Online guaranteed reachable set approximation for systems with changed dynamics and control authority. *IEEE Transactions on Automatic Control*, 2023.

Disasters

$$\dot{x} = f(x, u), \quad u \in U$$

Change in dynamics
(e.g., physical damage)

Partial loss of control
(e.g., adversarial takeover)

Actuator
degradation

$$\dot{x} = \hat{f}(x, u, v), \quad (u, v) \in \hat{U}$$

Loss of Control Authority

$$\dot{x} = \in U$$

Two-player

P1 (“cont
P2 (“envi

Players p

Can P1 win? me limit?)

Relative Actuator Strength

Intuitively, P1 wins for reaching **every** state if it can:

- "cancel out" the adversarial inputs and
- successfully fight **(or exploit!)** the drift at the same time

In linear systems, if the drift pushes the state "towards infinity" or "towards zero", fighting the drift is impossible with bounded inputs: drift eventually overpowers the inputs – the only allowed drift is from $Re(\lambda(A)) = 0$

In bounded state spaces, bounded sets of states to reach, etc., fighting the drift may be possible. In all cases, the available input needs to be "stronger" than the adversary

J.-B. Bouvier, M. Ornik. Resilient reachability for linear systems. 21st IFAC World Congress, 2020.

Resilience to Loss of Authority

If the original system dynamics are $\dot{x} = Ax + \overline{B}\overline{u}$, loss of authority over several actuators means the system shifts to:

$$\dot{x} = Ax + Bu + Cv, \quad \overline{B} = \begin{bmatrix} B & C \end{bmatrix}, \quad u \in U, \quad v \in V$$

Available input stronger than the adversary: *almost the same* as

$$0 \in \mathrm{Int}\left(\bigcap_{v \in V} BU + Cv \right)$$

If system is resilient to loss of authority, **how resilient is the system?**

Quantitative Resilience

If the target state is in the guaranteed reachable set of the initial state (*guaranteed* with respect to all possible adversarial control inputs):

it also matters how long it would take to reach the state, compared to the time it would take the system with nominal dynamics.

Idea: compare the quotient of the **nominal time-optimal reach time** and the **worst-case time-optimal reach time with adversarial input**

$$r_q = \frac{\inf_{\bar{u}} T_R^{\bar{u}}(x_0, x_{goal})}{\sup_{v} \inf_{u} T_R^{(u,v)}(x_0, x_{goal})}$$

Resilience quotient = 0: no resilience; **resilience quotient = 1:** perfect resilience

The adversary chooses an input such that whatever the controller chooses, reach time will be large

J.-B. Bouvier, K. Xu, M. Ornik. Quantitative resilience of linear driftless systems. 2021 SIAM Conference on Control and Its Applications, 2021.

Driftless Systems

Computing resilience quotient: three (partly nested) optimal control problems

For **driftless systems and integrators**:

- Optimal control = optimization
- Optimal control inputs can be geometrically determined
- There is a state-invariant notion of “direction in which the adversary is the strongest”
- It is possible to determine the least resilient goal state *(“Minimax Maximax Quotient Theorem”)*

J.-B. Bouvier, M. Ornik. The maximax minimax quotient theorem. Journal of Optimization Theory and Applications, 2022.
J.-B. Bouvier, K. Xu, M. Ornik. Quantitative resilience of generalized integrators. IEEE Transactions on Automatic Control, 2023.

Lyapunov Bounds

Even for general linear systems,

- Optimal control signal is bang-bang, but impossible to analytically determine
- The set of velocities which the system can take depends on the state

Idea:

- A lower bound on the nominal optimal reach time: a lower bound on $\dot{V}(x(t))$
- An upper bound on the nomi$\ldots$me: an upper bound on $\dot{V}(x(t))$ for a "good candidate for opti$\ldots$

$$T_N^*(x_0) \geq 2\frac{\lambda_{min}^P}{\lambda_{max}^Q}\ln\left(1+\frac{\lambda_{max}^Q\|x_0\|_P}{2\lambda_{min}^P b_{max}^P}\right)$$

- A lower bound on the worst-c$\ldots$me: a lower bound on $\dot{V}(x(t))$ for any response to the advers$\ldots$

$$T_N^*(x_0) \leq 2\frac{\lambda_{max}^P}{\lambda_{min}^Q}\ln\left(1+\frac{\lambda_{min}^Q\|x_0\|_P}{2\lambda_{max}^P b_{min}^P}\right)$$

- An upper bound on the worst$\ldots$time: an upper bound on $\dot{V}(x(t))$ for a "good candidate for optir$\ldots$ given the adversarial input

$$T_M^*(x_0) \geq 2\frac{\lambda_{min}^P}{\lambda_{max}^Q}\ln\left(1+\frac{\lambda_{max}^Q\|x_0\|_P}{2\lambda_{min}^P z_{max}^P}\right)$$

$$T_M^*(x_0) \leq 2\frac{\lambda_{max}^P}{\lambda_{min}^Q}\ln\left(1+\frac{\lambda_{min}^Q\|x_0\|_P}{2\lambda_{max}^P z_{min}^P}\right)$$

J.-B. Bouvier, M. Ornik. Quantitative resilience of linear systems. 20th European Control Conference, 2022.

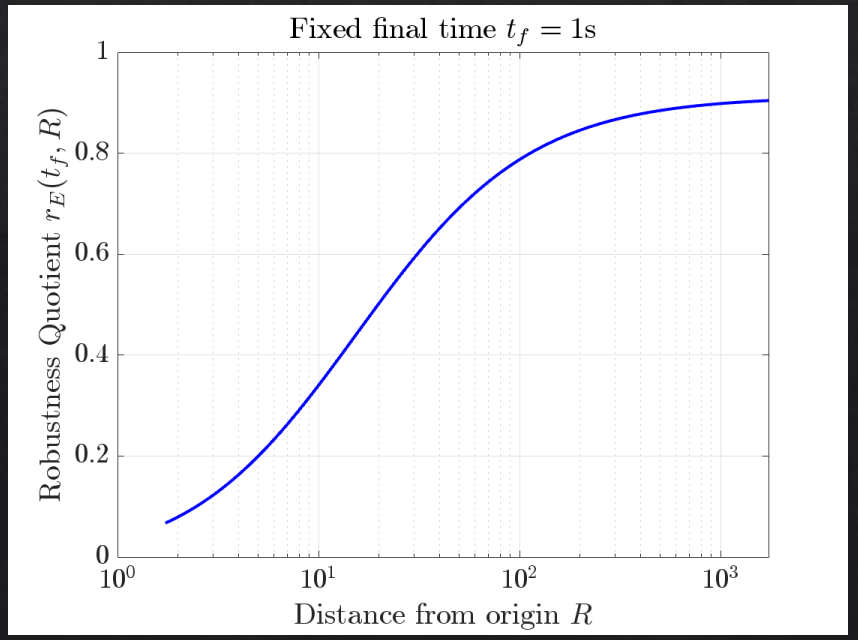
Energy Resilience

Time resilience is **hard**

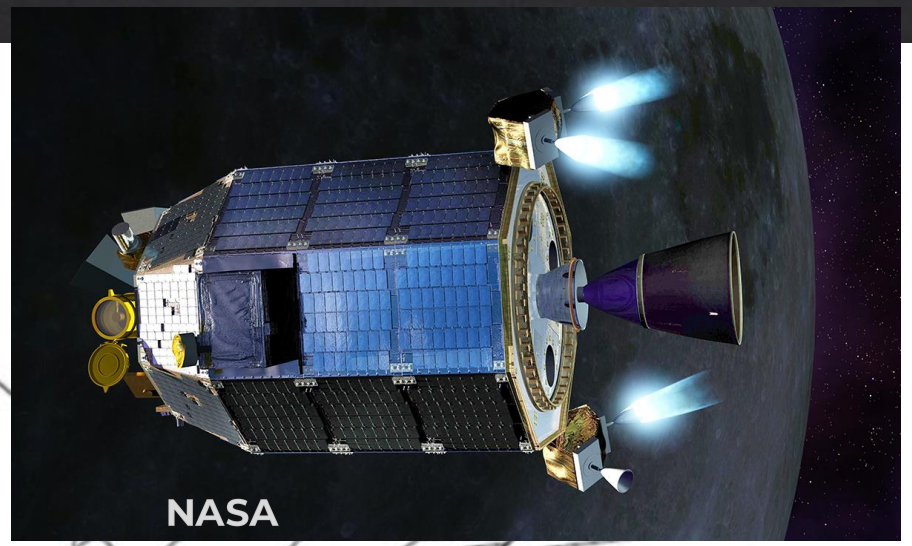
- Determining minimal time to reach a target for nominal dynamics with input constraints is not simple
- Determining minimal energy to reach a target (soft input constraints) for nominal dynamics **is simple**: controllability Gramian

for linear systems

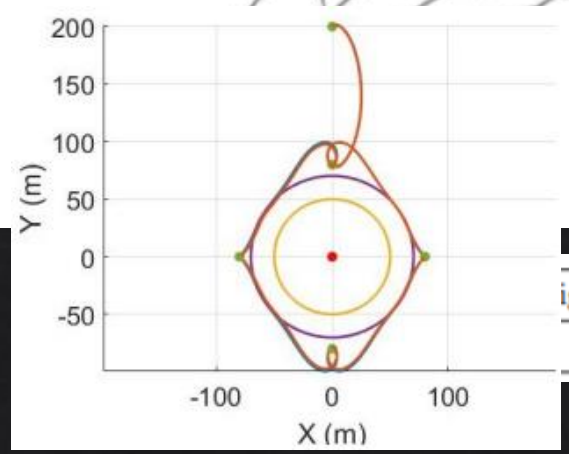
- Pathway towards nonlinear systems: energy resilience – price of nonlinearity?
- Idea: Describe nonlinear system as a higher-dimensional linear system

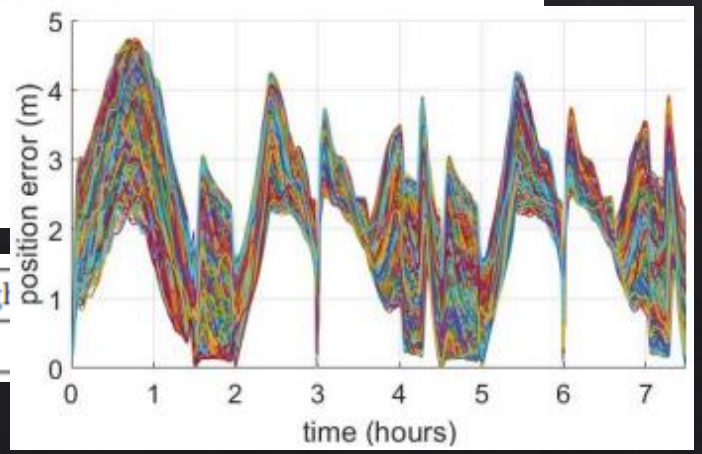


Applications



Canards





J.-B. Bouvier, M. Ornik. Designing resilient linear systems. IEEE Transactions on Automatic Control, 2022.
J.-B. Bouvier, H. Panag, R. Woollands, M. Ornik. Resilience of orbital inspections to partial loss of control authority of the chaser satellite. 2022 AAS/AIAA Astrodynamics Specialist Conference, 2022.

Disasters

$$\dot{x} = f(x, u), \quad u \in U$$

Change in dynamics
(e.g., physical damage)

Partial loss of control
(e.g., adversarial takeover)

Actuator
degradation

$$\dot{x} = \hat{f}(x, u, v), \quad (u, v) \in \hat{U}$$

Unknown Dynamics

$\dot{x}$ [redacted] J

Robust contro

Classical adapt own (finitely many) paramet

Both assume th

Guaranteed Reachability

After a change in dynamics, the system can almost certainly no longer use the same control law to reach its target

- It may not even be able to reach its target using **any** control law

What is it **certifiably capable** of doing (even if we don't yet know how)?

- If **nothing** is known, nothing is certain
- Otherwise,

$$R^{\mathcal{G}}(x_0) = \bigcap_{\tilde{f} \in D_{con}} R^{\tilde{f}}(x_0)$$

All dynamics consistent with the current knowledge

M. Ornik. Guaranteed reachability for systems with unknown dynamics. 59th IEEE Conference on Decision and Control, 2020.

Guaranteed Velocities

- In theory, guaranteed reachability set is well-defined
- In practice, how do compute it?

Idea: By finding all trajectories obtained by integrating **guaranteed velocities**

$V^{\mathcal{G}}(t) = \bigcap_{\tilde{f} \in D_{con}} V^{\tilde{f}}(t)$, we will obtain at least some (if not all) guaranteed reachable states

Velocities guaranteed at time t = velocities guaranteed at time 0 "modulo" maximal system wildness

M. Ornik. Guaranteed reachability for systems with unknown dynamics. 59th IEEE Conference on Decision and Control, 2020.

Initial Velocities + Lipschitz

Set of initially available velocities **can** be computed in arbitrarily short time: e.g., the learning part of **myopic control**

Bound on the change in set of available velocities: known system parameters + Lipschitz bound on the change in system dynamics (from physical laws and/or design knowledge)

For control-affine systems $\dot{x} = f(x) + G(x)u, \quad u \in U,$

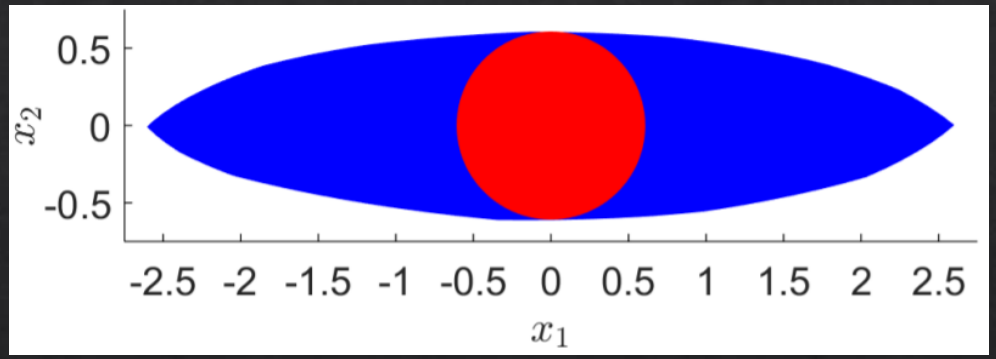
$$V^{\mathcal{G}}(x) = \bigcap_{(\hat{f},\hat{G}) \in M_x^0(f(0)) \times M_x^1(G(0))} \hat{f} + \hat{G}U$$

Infinitely many (particularly parametrized) translated/dilated/rotated copies of the same object

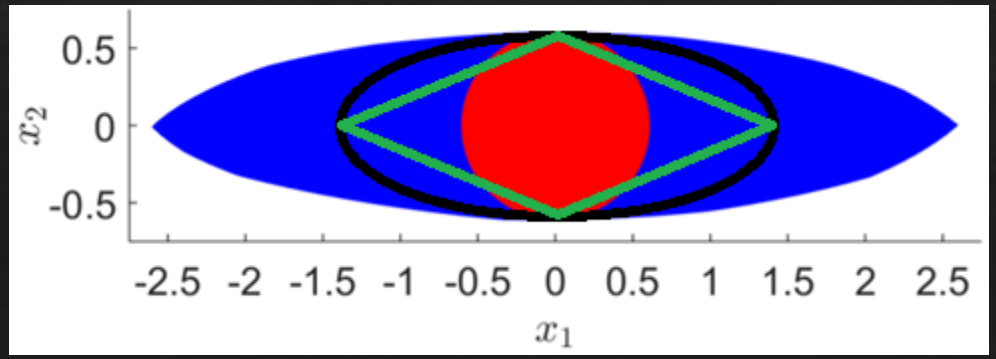
M. Ornik. Guaranteed reachability for systems with unknown dynamics. 59th IEEE Conference on Decision and Control, 2020.
T. Shafa, M. Ornik. Maximal ellipsoid method for guaranteed reachability in unknown nonlinear systems. 61st IEEE Conference on Decision and Control, 2022.

Simple Guaranteed Velocities

Initial idea: fit a **maximal ball** within this infinite intersection



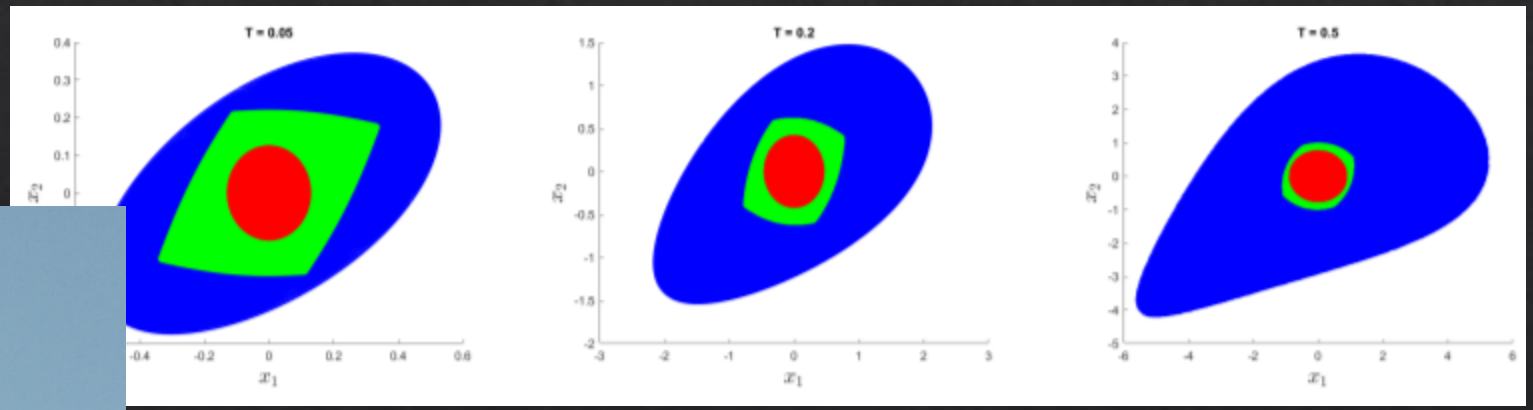
Better idea: fit a more complicated object, but still geometrically simple

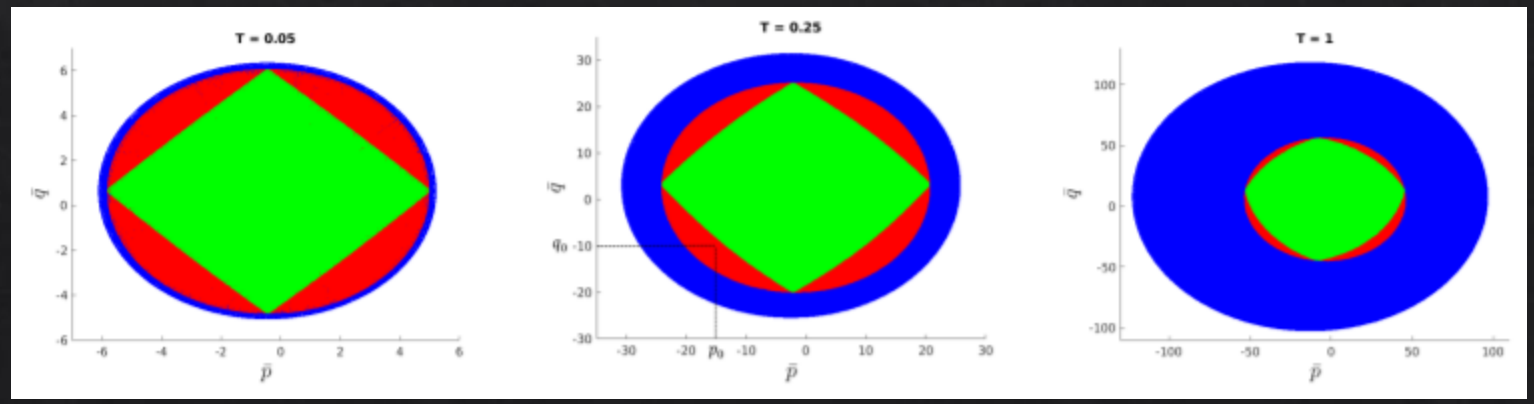


T. Shafa, M. Ornik. Reachability of nonlinear systems with unknown dynamics. *IEEE Transactions on Automatic Control*, 2023.

Reachable Set Underapproximations

A **guaranteed** reachable set: reachable set of a "simple" control system





T. Shafa, M. Ornik. Reachability of nonlinear systems with unknown dynamics. *IEEE Transactions on Automatic Control*, 2023.

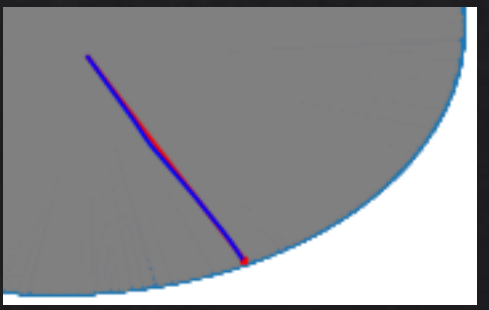
Learning-Control Pipeline

Once we have established what the system is capable of doing, we still do not know *how* to do it

- **End-to-end planning, learning and control**:
  - **Task assignment**: what task can be provably completed
  - **Real-time learning**: what do we need to know in order to be able to complete it
    - persistent excitation allows us to learn local dynamics – we "need" a little time, but not much
  - **Assured control:** complete the task

Lemma 4.5: Let $n \geq 1$ be fixed. Let $\mathbf{u}(t) = u_{n,j}$ for $t \in [\tau_n + j\delta t, \tau_n + (j+1)\delta t]$ and for $j \in \mathcal{I}$. Let $\mathfrak{x}_n$ be given and consider $z_n = \theta_n y$ with $|z_n - \mathfrak{x}_n| = r$ for some $\theta_n \in (0,1)$. Suppose that $\dot{d}_{z_n}(\phi_{\mathbf{u}}(t, \mathfrak{x}_n)) < 0$ for all $t \in [\tau_n, \tau_{n+1}]$. Then, there exists a $z_{n+1} = \theta_{n+1} y$ such that $\theta_{n+1} > \theta_n$ and $|z_{n+1} - \mathfrak{x}_{n+1}| = r$. Particularly, $r - |\mathfrak{x}_{n+1} - z_n| \leq |z_{n+1} - z_n| < 2r$.

Corollary 4.12: Let $r = r(k, \tau)$ be the maximal radius of $\hat{\mathcal{R}}(k\tau, \mathfrak{x}_n)$ for some k. Suppose that $2(M_0(m+1)^2\delta t + rL_{\max})M_0(m+1)^2\delta t + \mu(\delta t, \epsilon) \leq (r - M_0(m+1)^2\delta t)(b - c|\mathfrak{x}_n|)$. Let $\mathbf{u}(\tau_n) = u_{n,0} := \operatorname{argmin}_{u \in \mathcal{U}_\lambda} \dot{d}_{z_n}(\phi_{\mathbf{u}}(\tau_n, x_0))$ be the controller at τ_n. Then $\dot{d}_{z_n}(\phi_{\mathbf{u}}(t, \mathfrak{x}_n)) < 0$ for all $t \in [\tau_n, \tau_n + \tau]$.



Disasters

$$\dot{x} = f(x, u), \quad u \in U$$

Change in dynamics
(e.g., physical damage)

Partial loss of control
(e.g., adversarial takeover)

Actuator
degradation

$$\dot{x} = \hat{f}(x, u, v), \quad (u, v) \in \hat{U}$$

Vision

Complexity

For what systems can we estimate capabilities? What capabilities can we estimate? What tasks can we complete?

Validation

But really…
what can we do *onboard during a disaster*?

Short Term: Complexity

Much remains to be done: we only just started

- Technical assumptions (linear systems in loss of control authority, full-ish actuation for unknown dynamics, …)
- Combination of scenarios: e.g., partially unknown dynamics with partial loss of control
- **We can be smarter**
  - Structural knowledge of “undamaged” part of the unknown dynamics
  - Not just reachability, but safe reachability **(reach-avoid)** and output reachability
- **Time-delayed systems, system operating under state constraints**

Teaser: Time Delays

Sensing of state and control inputs takes time:

$$\dot{x}(t) = f\left(x(t), u\left(t, x(t-\tau), v(t-\tau)\right), v(t)\right)$$

No perfect response

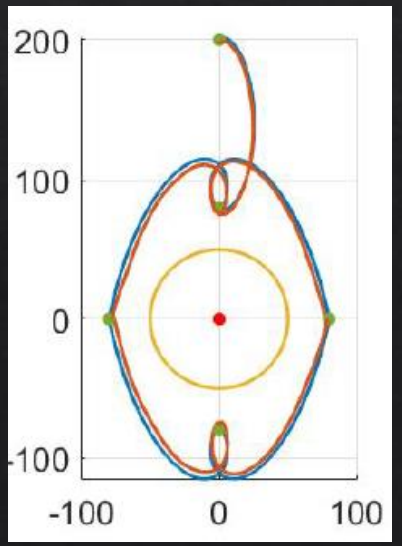
- Reachability is unlikely, but *approximate reachability* may be attainable
- How much damage can an adversary do by exploiting delay?
  - How large is the delay?
  - How wild are the system dynamics?

Adversary's strength

Delay size

$$\rho = \frac{\max\{\|Cv\| \mid v \in V\}}{\lambda_{max}\left(A + A^T\right)/2}\left(e^{T_c \lambda_{max}\left(A+A^T\right)/2} - 1\right), \quad T_c \geq \tau$$

Wildness in dynamics



J.-B. Bouvier, H. Panag, R. Woollands, M. Ornik. Delayed resilient trajectory tracking after partial loss of control authority over actuators. Preprint: arXiv:2303.12877 [eess.SY].

Teaser: Guaranteed Reachability on a Manifold

Another kind of side knowledge: no "full actuation"

$$\dot{x} = f^?(x, u), \quad u \in U$$

$\dot{x} \in T_x\mathcal{M}$, $x \in \mathcal{M}$

Finding the reachable set of a *known* system on a manifold is a challenge

- Analytical methods are hard enough on a Euclidean space
- How to approximate solving ODEs on a manifold?
- No good tools

Idea for guaranteed reachability: exploit "maximal wildness" of dynamics

- On a Euclidean space: "maximal wildness" = Lipschitz constant
- Lipschitz on surfaces ← Riemannian metric

Validation and More

- **Using sensors and perception** to recognize fault type and gather information
  - Fault detection; sensor fusion; state estimation

- **Complex missions** in high-fidelity simulation
  - Concurrent sensing and actuation faults
  - Noise+hostile action

Ongoing joint work and demonstrations with TC3 team (e.g., GUAM)

Acknowledgments

Extensive support from NASA (incl. current ULI), US Department of Defense – Air Force Office of Scientific Research, and Discovery Partners Institute

Real research (among others)**:**

- **Actuator degradation:** Hamza El-Kebir, Ani Pirosmanishvili
  collaboration with Joseph Bentsman, Changrak Choi et al. (JPL)
- **Partial loss of control:** Jean-Baptiste Bouvier, Kathleen Xu (UIUC/MIT), Ram Padmanabhan
  collaboration with Sai Pushpak Nandanoori et al. (PNNL), Robyn Woollands, Himmat Panag
- **Unknown dynamics:** Taha Shafa, Yiming Meng

mornik@illinois.edu