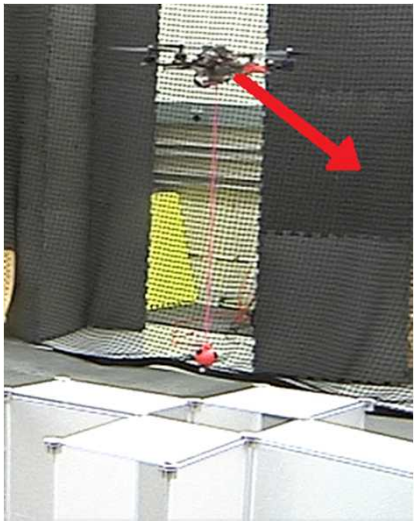
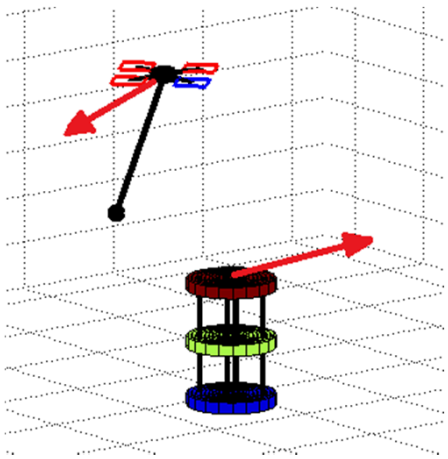
SAND2015-4141C

Sandia National Laboratories





Preference-balancing Motion Planning under Stochastic Disturbances

Aleksandra Faust[1], Nick Malone[2], and Lydia Tapia[2]

[1]Sandia National Laboratories

[2]University of New Mexico

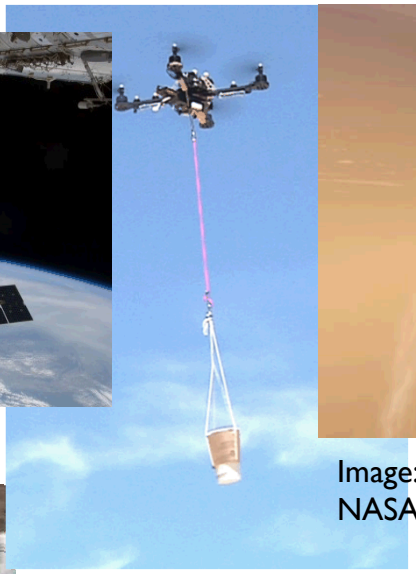
UNM

Adaptive Motion Planning Research Group

Preference Balancing Tasks (PBTs)

- Robotic motion
  - Control-affine system
  - Continuous states and actions
  - High-dimensional
- Complicated dynamics
  - Difficult demonstration
  - Lack of motion primitives
- Described with preferences

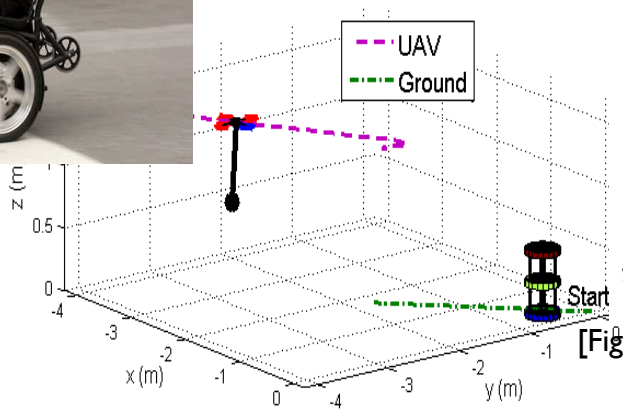
Image: SpaceX



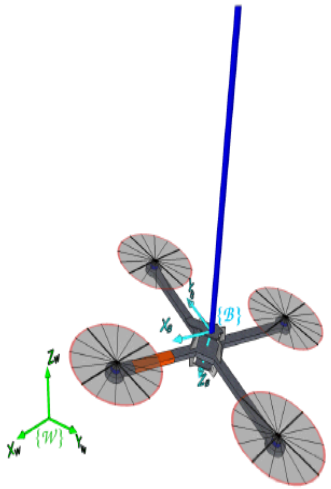
[Faust et al., 2013]

Image: NASA/JPL/Caltech

Image: GM



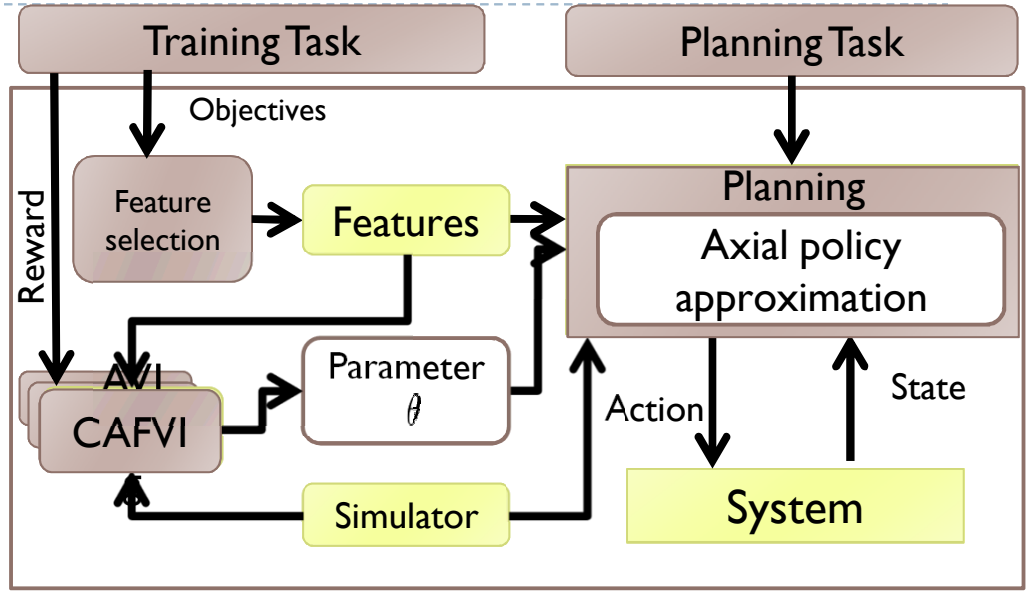
[Faust et al., in press]



[Figueroa et al., 2014]

PrEference Appraisal Reinforcement Learning (PEARL)

[Faust, 2014]



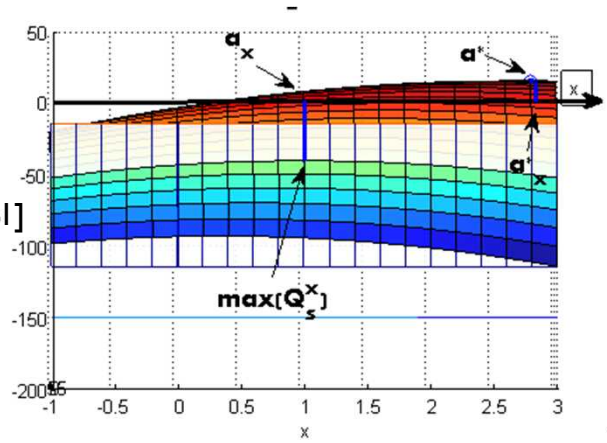
- Learns to perform PBTs*
- Batch reinforcement learning (RL)
- Learning
  - Continuous action fitted value iteration (CAFVI) [Faust, et al, in press]
  - Linear map state-value function approximation
  - Features are squared preferences $V(\mathbf{s}) = \boldsymbol{\theta}^T \mathbf{F}(\mathbf{s})$
- Planning
  - Generates trajectory
  - Real-time, one step at the time
  - Axial sum policy approximation [Faust, et al, in press]

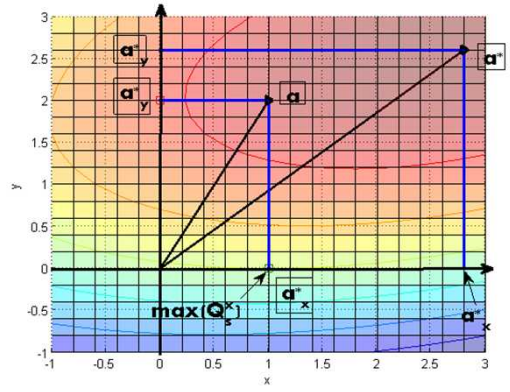
Works only with deterministic systems

*Preference Balancing Tasks

Axial sum policy approximation

- Greedy policy $h^*(x) = \text{argmax}_{u \in U} V(D(x,u))$ in continuous spaces
  - Sampling-based search space narrowing [Busoniu et al. 2013] [Mansley et al. 2011] [Walsh et al. 2010]
  - Gradient descent [Hasselt et al. 2012]
- Deterministic axial sum policy (DAS) [Faust, et al, in press]
  - Interpolate action-value function along each the axis
  - Find maximum
  - Combine with a vector or convex sum
- Sufficient conditions for convergence to goal
  - Control-affine system with bounded drift [Faust, et al, in press]
  - Squared-features
  - Negative weights

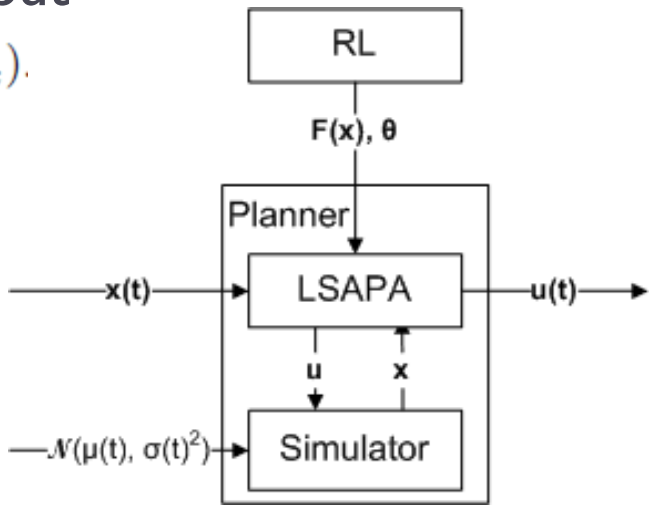




Extend DAS to work under external disturbances.

Least Squares Axial Policy Approximation (LSAPA)

- Problem
  - PBTs $V(x) = \sum_{i=1}^{d_g} \theta_i F_i(x).$
  - Control-affine system with external input disturbance $x_{k+1} = f(x_k) + g(x_k)(u_k + \eta_k).$
- Learning
  - Deterministic CAFVI
- Planning
  - Estimate disturbance in real-time
  - Least Squares Axial Policy Approximation selects an action at every time step
  - Adapts to observed disturbance



Least Squares Axial Policy Approximation Continued

- On each input axis
  - Action-value function $Q_{\boldsymbol{x},i}(u) = \boldsymbol{p}_i^T [u^2 \; u \; 1]^T$
  - Collect d_n dynamic samples $U_i = [u_{1,i} \; \ldots \; u_{d_n,i}]^T$
  - Calculate $X_i = [\boldsymbol{x}'_{1,i} \; \ldots \; \boldsymbol{x}'_{d_n,i}]^T$

$$\boldsymbol{Q_i} = [Q_{\boldsymbol{x},1}(u_{1,i}) \; \ldots \; Q_{\boldsymbol{x},d_n}(u_{d_n,i})]^T$$

$$Q_{\boldsymbol{x},j}(u_{j,i}) = \theta^T F(\boldsymbol{x}'_{j,i}) \quad j = 1, \ldots, d_n$$

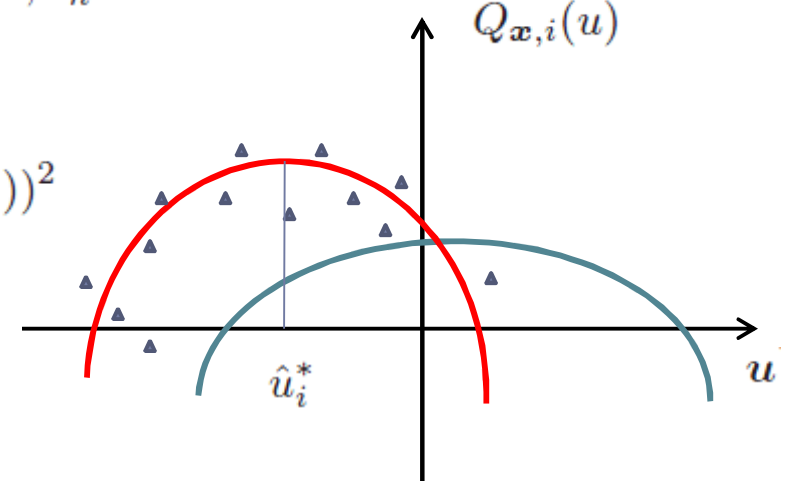
  - Solve supervised ML problem

$$\hat{\boldsymbol{p}}_i = \underset{\boldsymbol{p}_i}{\operatorname{argmin}} \sum_{j=1}^{d_n} (C_{j,i}\boldsymbol{p}_i - Q_{\boldsymbol{x},j}(u_{j,i}))^2$$

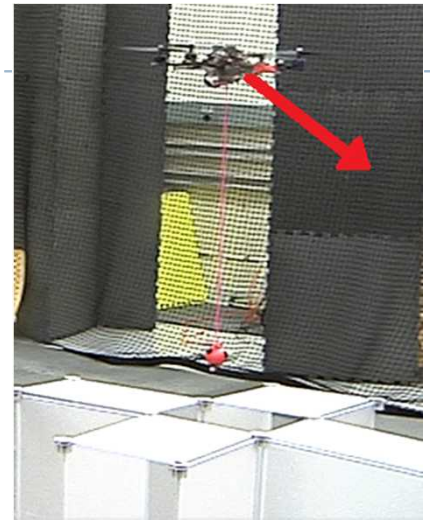
$$C_i = \begin{bmatrix} (u_{1,i})^2 & u_{1,i} & 1 \\ (u_{2,i})^2 & u_{2,i} & 1 \\ & \ldots & \\ (u_{d_n,i})^2 & u_{d_n,i} & 1 \end{bmatrix}$$

  - Find maximum

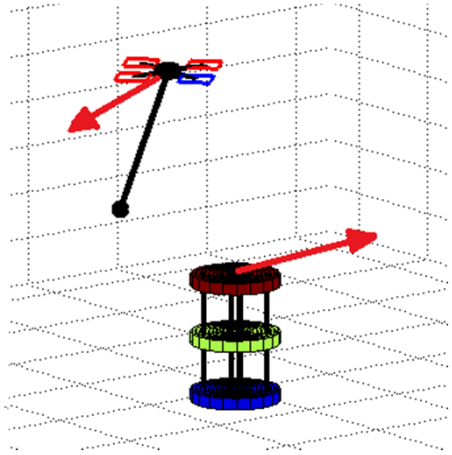
$$\hat{u}_i^* = -\frac{\hat{p}_{1,i}}{2\hat{p}_{2,i}}$$



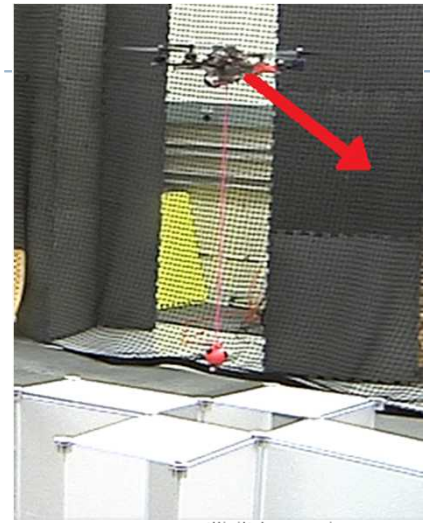
Results

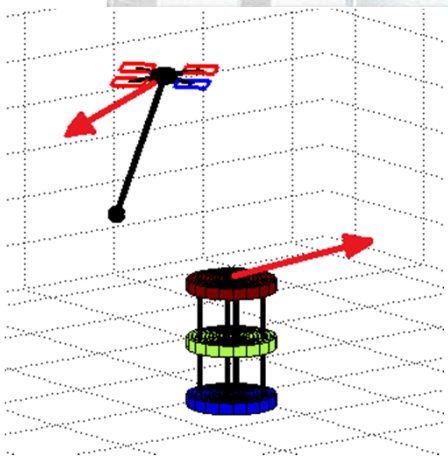


- Related work on UAV robust control
  - Trajectory tracking [Alexis et al. 2010]
  - Harmonic potential fields [Masoud 2011]
  - Low-level controllers [DeCastro and Kess-Gazit 2013]
  - Trajectory libraries [Majumdar and Tedrake 2013]
  - Blimp path planning with dynamic programming [Kawano 2011]
- Coffee-delivery Tasks
  - Swing-free aerial cargo delivery
  - Rendezvous task



Coffee Delivery: Setup

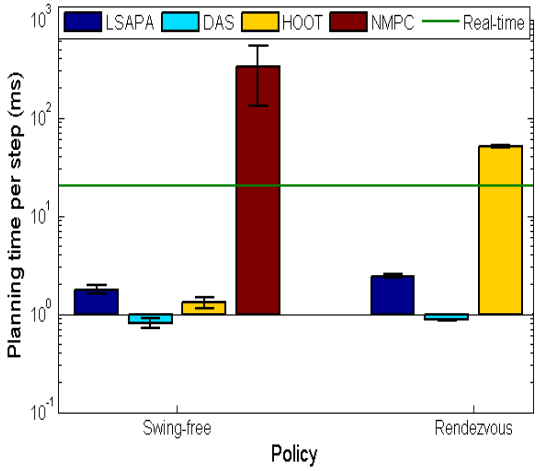


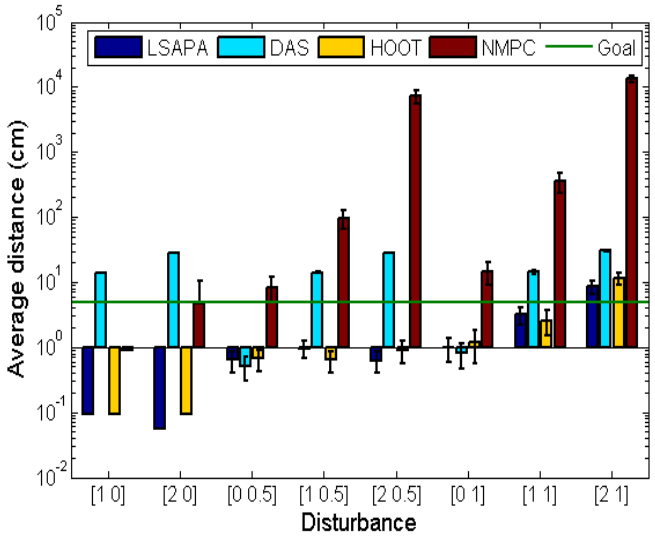


- Aerial Problem
  - Holonomic cargo-bearing UAV
  - Bring the suspended load to the destination
  - Minimal residual load oscillations at arrival
- Rendezvous Problem
  - Holonomic cargo-bearing UAV and ground robot
  - Bring the suspended load to the ground robot
  - Minimal residual load oscillations at arrival
- Preferences, reduce
  - Distance from the destination
  - Vehicle's velocity
  - Load displacement
  - Load's velocity
- MDP
  - Aerial: 10-dimensional vector states, 3-dimenstional actions
  - Rendezvous: 16-dimensional vector states, 5-dimenstional actions

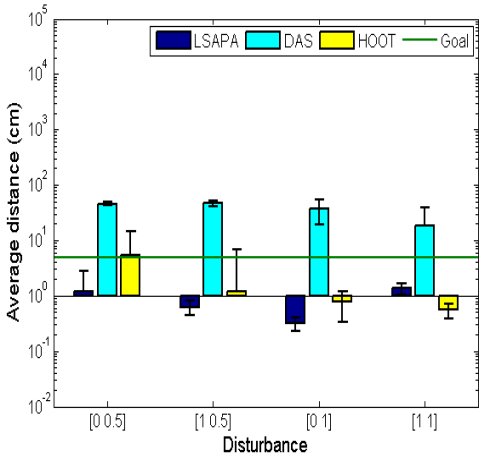
 [Faust et al., 2015]

Trajectory Characteristics under Varying Disturbance





Aerial Coffee Delivery



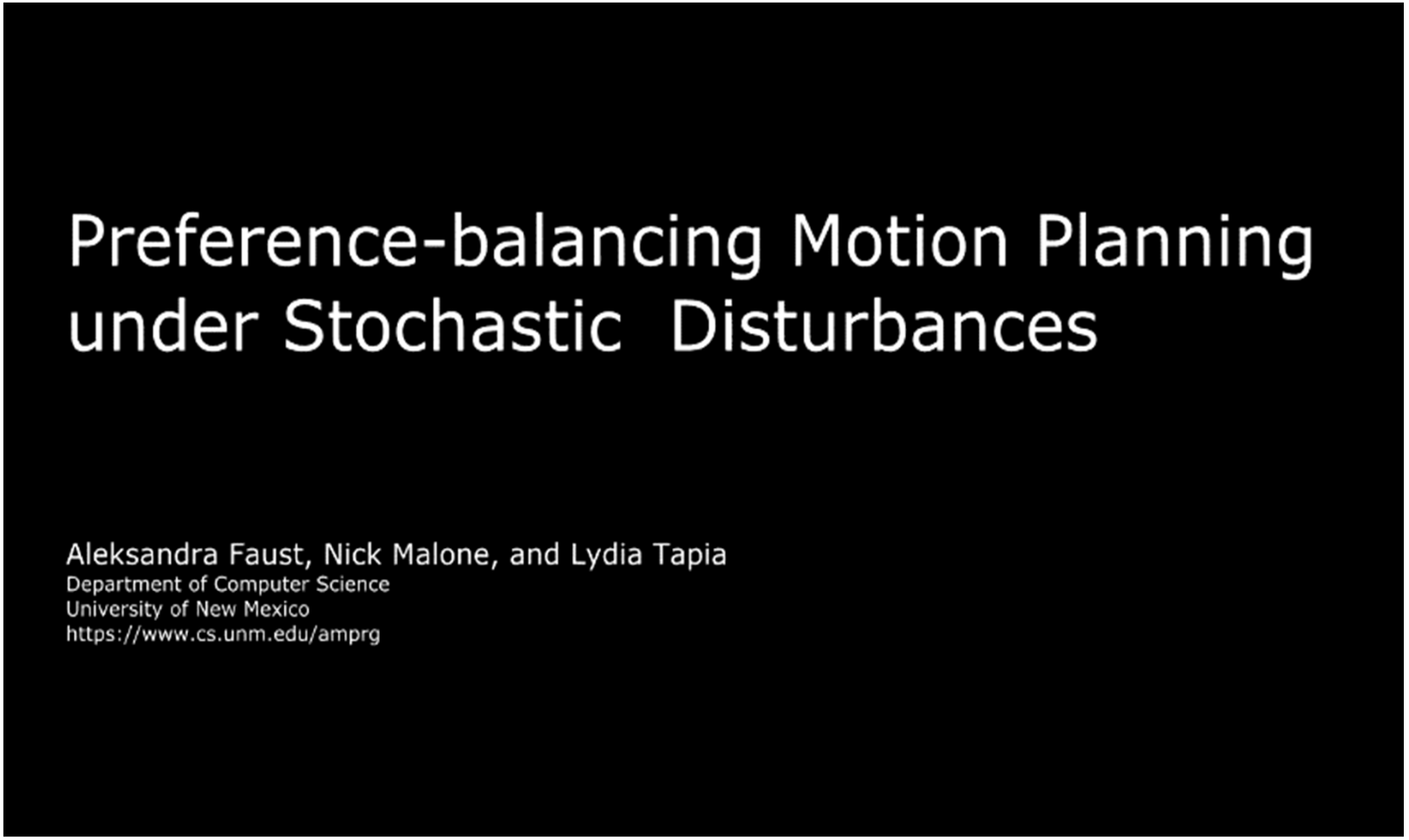
Rendezvous

- Least Squares Axial Policy Approximation (LSAPA) [Faust et al., 2015]
- Deterministic Axial Sum (DAS) [Faust et al., 2014]
- HOOT [Mansley et al. '11]
- Nonlinear Model Predictive Control (NMPC) [Grune and Pannek, 2011]

LSAPA and DAS perform decision-making in real-time.

LSAPA reaches the goal for non-zero mean disturbances.

Swing-free aerial cargo delivery with disturbances



http://www.cs.unm.edu/~afaust/movies/afaustIcra15.mp4

Questions

- Thank you!

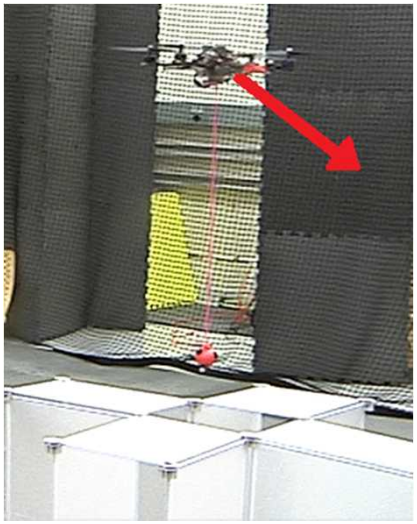
- Acknowledgements
  - Dr. Peter Ruymgaart for discussing disturbance modelling
  - Patricio Cruz for assisting with experiments
  - Reviewers for very helpful and constructive feedback

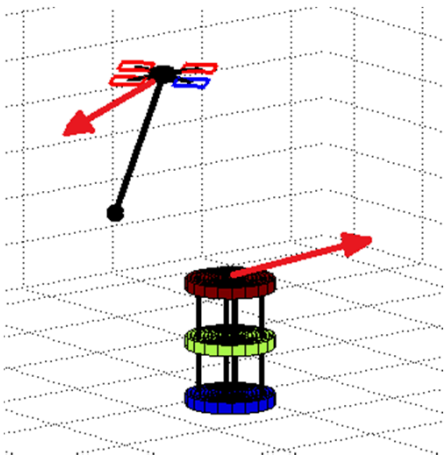
This work was in part supported by NM Space Grant and National Institutes of Health (NIH) Grant P20RR018754 to the Center for Evolutionary and Theoretical Immunology.

Sandia National Laboratories

UNM

Adaptive Motion Planning Research Group





Preference-balancing Motion Planning under Stochastic Disturbances

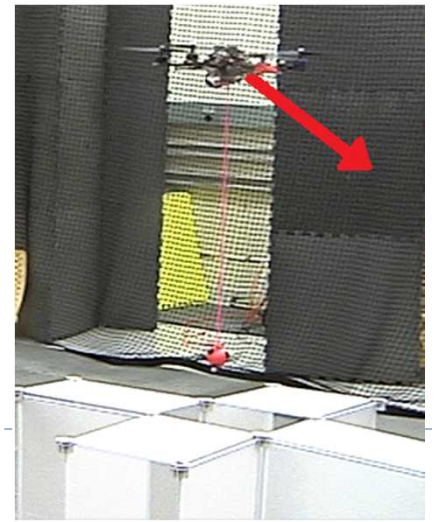
Aleksandra Faust[1], Nick Malone[2], and Lydia Tapia[2]

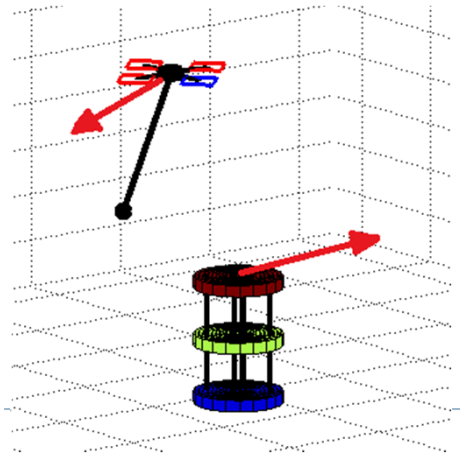
[1]Sandia National Laboratories

[2]University of New Mexico

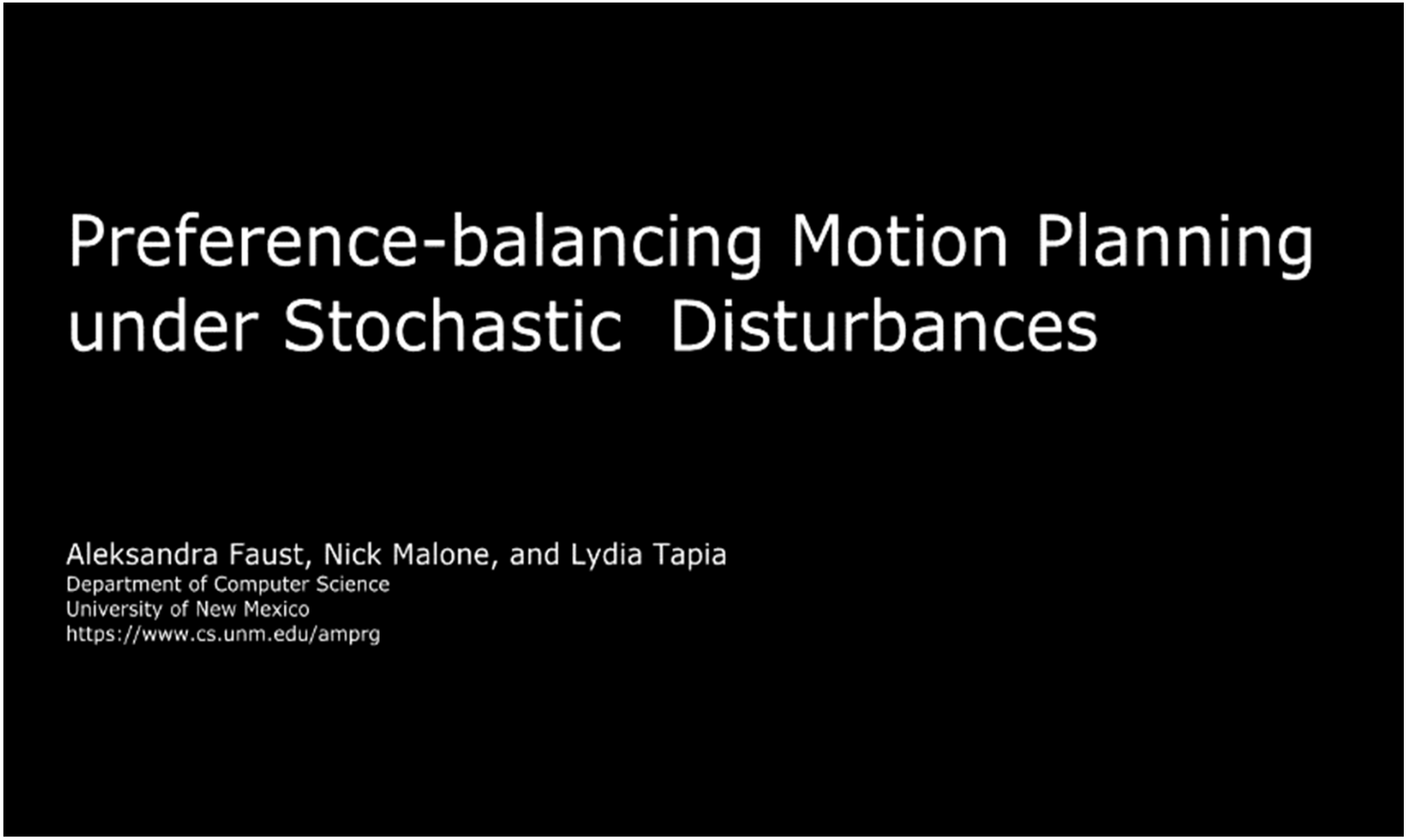
Preference-balancing Motion Planning under Stochastic Disturbances Summary

- Reinforcement learning with no disturbances
- Online planning in the presence of disturbances
- Applicable for continuous actions
- Linear in the input dimensionality
- Works through a Least Squares Axial Sum Policy Approximation





Swing-free aerial cargo delivery with disturbances



http://www.cs.unm.edu/~afaust/movies/afaustIcra15.mp4

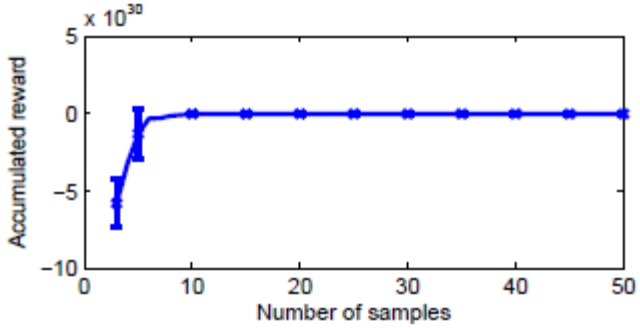
Sandia National Laboratories

UNM

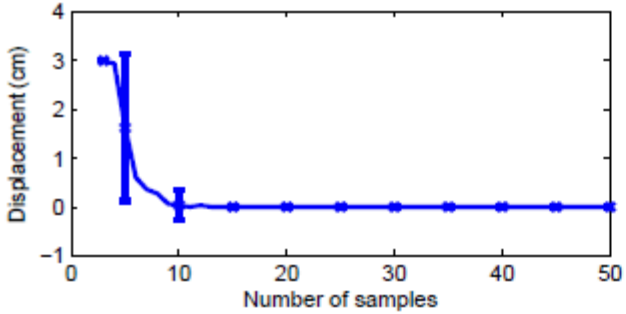
Adaptive Motion Planning Research Group

Flying Inverted Pendulum

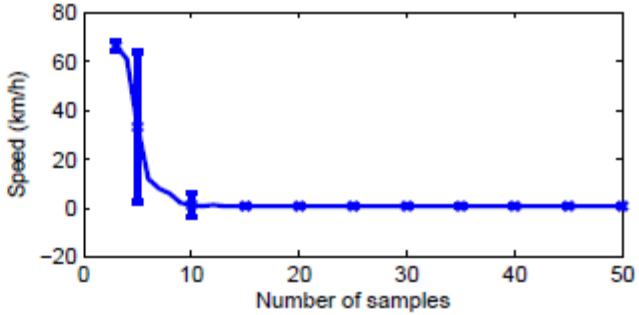
Number of samples need



(a) Accumulated reward



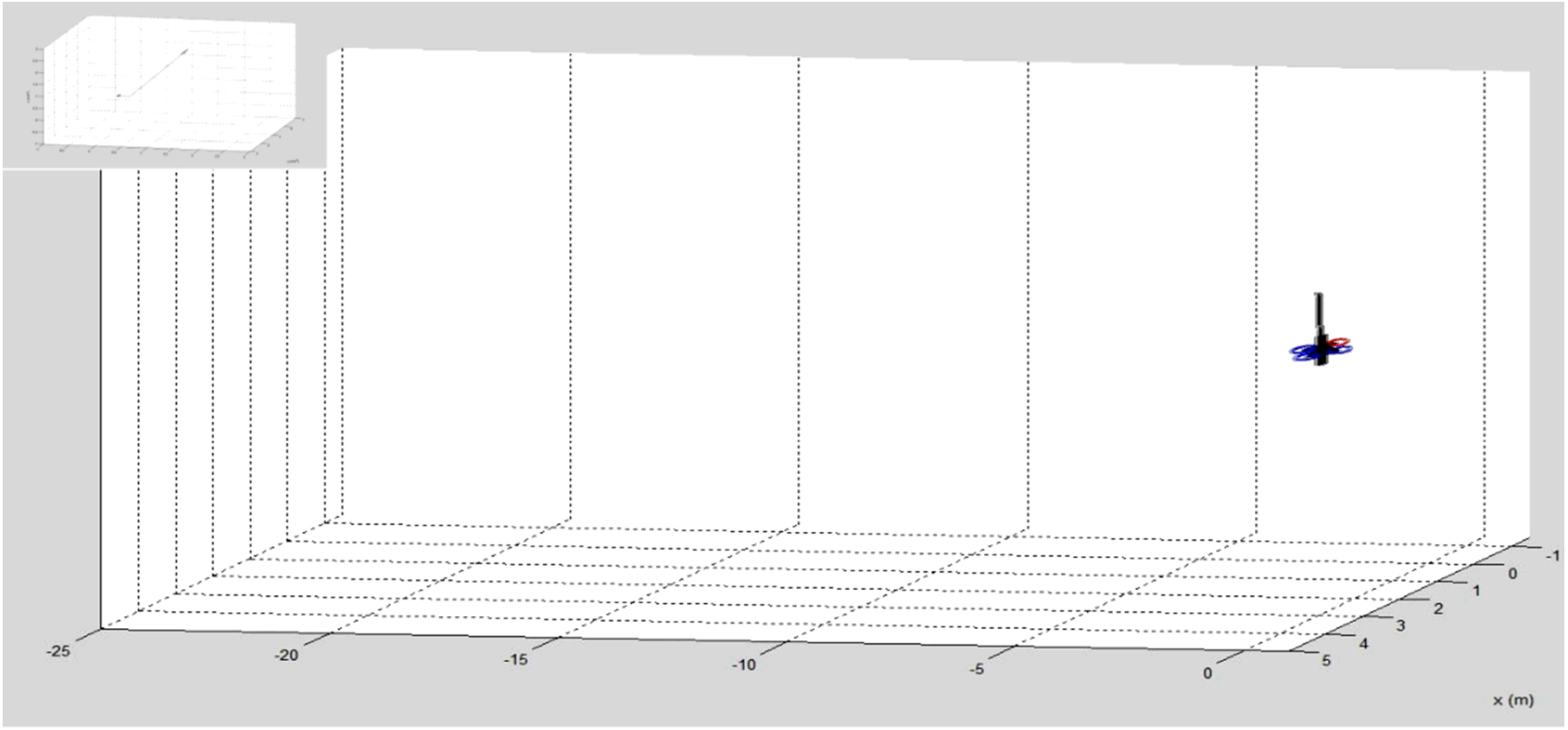
(b) Displacement



(c) Speed

Trajectory characteristics improve exponentially with number of samples.

Flying Inverted Pendulum



Task: *Flying inverted pendulum*
Planning: *Least squares axial policy approximation (LSAPA)*
Upper left: *Stochastic disturbance ~N(1,1)*

Related work

- UAV Robust control
  - Trajectory tracking [Alexis et al. 2010]
  - Harmonic potential fields [Masoud 2011]
  - Low-level controllers [DeCastro and Kess-Gazit 2013]
  - Trajectory libraries [Majumdar and Tedrake 2013]
  - Blimp path planning with dynamic programming [Kawano 2011]
- Greedy policy approximation
  - Sampling based planning search space narrowing [Mansley et al. 2011] [Busoniu et al. 2013] [Walsh et al. 2010] [Bubek et al. 2011]
  - Gradient descent [Hasselt et al. 2012]