Robust Kinodynamic Motion Planning using Model-Free Game-Theoretic Learning

George P. Kontoudis[1] and Kyriakos G. Vamvoudakis[2]

[1]Kevin T. Crofton Department of Aerospace and Ocean Engineering,
Virginia Polytechnic Institute and State University
[2]School of Aerospace Engineering,
Georgia Institute of Technology

gpkont@vt.edu

American Control Conference, Philadelphia, USA
July 10,2019

VIRGINIA TECH

Motivation

Autonomous systems:

- Ground vehicles
- Aerial vehicles
- Boats and ships

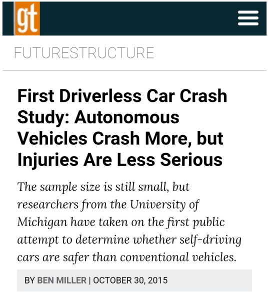
Expectations:

1. Safe navigation w/o collision
2. Robust to model uncertainties and external disturbances
3. Energy efficiency

Motivation

Autonomous systems:

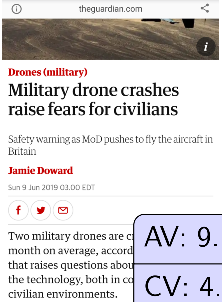
- Ground vehicles
- Aerial vehicles
- Boats and ships


FUTURESTRUCTURE

First Driverless Car Crash Study: Autonomous Vehicles Crash More, but Injuries Are Less Serious

The sample size is still small, but researchers from the University of Michigan have taken on the first public attempt to determine whether self-driving cars are safer than conventional vehicles.

BY BEN MILLER | OCTOBER 30, 2015


theguardian.com

Drones (military)

Military drone crashes raise fears for civilians

Safety warning as MoD pushes to fly the aircraft in Britain

Jamie Doward

Sun 9 Jun 2019 03.00 EDT

Two military drones are c
month on average, accord
that raises questions abou
the technology, both in co
civilian environments.

AV: 9.1 crashes/M miles
CV: 4.1 crashes/M miles
AV: 0.36 injuries/crash
CV: 0.25 injuries/crash
254 crashes/decade
70% operated by the US

Expectations:

1. Safe navigation w/o collision
2. Robust to model uncertainties and external disturbances
3. Energy efficiency

Motivation

Autonomous systems:

- Ground vehicles
- Aerial vehicles
- Boats and ships

Robust in all ocean and maritime conditions

Why We Will Never See Fully Autonomous Commercial Ships

BY COMMANDER DAVID DUBAY USCG 2019-06-25 17:24:08

The world will never see fully autonomous transoceanic commercial cargo ships. In fact, autonomous vessels are likely to operate in only very limited situations.

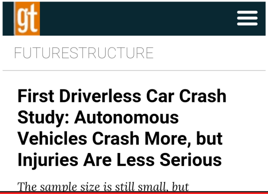
Expectations:

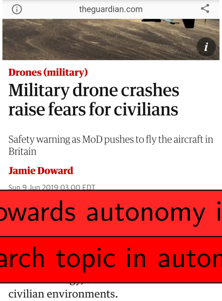
1. Safe navigation w/o collision
2. Robust to model uncertainties and external disturbances
3. Energy efficiency

Motivation

Autonomous systems:

- Ground vehicles
- Aerial vehicles
- Boats and ships


gt
FUTURESTRUCTURE
First Driverless Car Crash Study: Autonomous Vehicles Crash More, but Injuries Are Less Serious
The sample size is still small, but


theguardian.com
Drones (military)
Military drone crashes raise fears for civilians
Safety warning as MoD pushes to fly the aircraft in Britain
Jamie Doward
civilian environments.


Why We Will Never See Fully Autonomous Commercial Ships
operate in only very limited situations.

Despite hard efforts, the path towards autonomy is a **winding road**

Motion Planning is a key research topic in autonomy and robotics

Expectations:

1. Safe navigation w/o collision
2. Robust to model uncertainties and external disturbances
3. Energy efficiency

Online, Robust Kinodynamic Motion Planning

Kinodynamic motion planning: Realistic applications

- Subject simultaneously to kinematic and dynamic constraints
- More sophisticated dynamics including at least velocity and acceleration

Goal: Implementation of online, robust kinodynamic motion planning

Requirements:

1. Optimal control strategy
2. Reject external disturbances
3. Robust to model uncertainties
4. Online implementation of the algorithm
5. Safe navigation w/o collision

Online, Robust Kinodynamic Motion Planning

Kinodynamic motion planning: Realistic applications

- Subject simultaneou
- More sophisticated c

Goal: Implementation of

Requirements:

1. Optimal control strategy
2. Reject external disturbances
3. Robust to model uncertainties
4. Online implementation of the algorithm
5. Safe navigation w/o collision

Challenges
1. Dynamics are often difficult to derive
2. When obtained they are unreliable and inaccurate
3. Disturbances and parameter uncertainties affect dynamics
4. Optimality requires extensive offline computations

Problem Description

Given:

- System with undetermined linear dynamics
- Full state feedback
- Global path using RRT*

Goals:

- Model-free optimal control to drive the system
- Reject external disturbances
- Low computational effort, for online implementation

Steps:

1. A two point boundary value problem (TPBVP)
2. Local replanning problem
3. Combine TPBVP with local replanning

Continuous Linear Time Invariant System

Consider a continuous linear time invariant system,

$$\dot{x}(t) = Ax(t) + Bu(t) + Fd(t), \tag{1}$$

- $x(t) \in \mathcal{X} \subseteq \mathbb{R}^n$: measurable kinodynamic state vector
- Player 1: control $u(t) \rightarrow$ aims to minimize the energy performance
- Player 2: disturbance $d(t) \rightarrow$ desires to maximize the energy performance
- Assume (A,B) is controllable and $(\sqrt{M},A)$ is detectable

To drive the system from an initial state x_0 to a final state $x(T) = x_{\mathrm{r}}$,

$$\dot{\bar{x}}(t) = A\bar{x}(t) + B\bar{u}(t) + F\bar{d}(t) \tag{2}$$

$\bar{x}(t) := x(t) - x_{\mathrm{r}}$, $\bar{d}(t) := d(t) - d_{\mathrm{r}}$, and $\bar{u}(t) := u(t) - u_{\mathrm{r}}$

Kinodynamic Motion Planning Formulation

Assume that we know,

- $\mathcal{X}_{\text{obs}} \subset \mathcal{X}$: obstacle closed space
- $\mathcal{X}_{\text{free}} := (\mathcal{X}_{\text{obs}})^{\complement} = \mathcal{X} \backslash \mathcal{X}_{\text{obs}}$: free space
- Global path $\pi(x_{0,i}, x_{\text{r},i}) \in \mathbb{R}^{2(k \times n)}$
- Essentially, k-sets of TPBVPs from the RRT*
- The k-sets include initial $\mathcal{X}_0 \in \mathbb{R}^{k \times n} \subset \mathcal{X}_{\mathsf{free}}$ and final states $\mathcal{X}_{\text{G}} \in \mathbb{R}^{k \times n} \subset \mathcal{X}_{\mathsf{free}}$

Disturbances may be,

- External wind for unmanned aerial vehicles
- Rough terrain for unmanned ground vehicles
- Sea waves for autonomous boats
- Turbulance flow in shallow water environment for unmanned underwater vehicles

Finite Horizon Minimax Cost Function

The finite horizon energy performance measure,

$$J(\bar{x}; \bar{u}, \bar{d}; t_0, T) = \phi(T) + \frac{1}{2}\int_{t_0}^{T} (\bar{x}^\intercal M \bar{x} + \bar{u}^\intercal R \bar{u} - \gamma^2 \|\bar{d}\|^2)\, \mathrm{d}\tau, \tag{3}$$

The minimax problem has the form of,

$$V^\star(\bar{x}; t_0, T) := \min_{\bar{u}} \max_{\bar{d}} \left(\phi(T) + \frac{1}{2}\int_{t_0}^{T} (\bar{x}^\intercal M \bar{x} + \bar{u}^\intercal R \bar{u} - \gamma^2 \|\bar{d}\|^2)\, \mathrm{d}\tau \right) \tag{4}$$

- $\phi(T) := \frac{1}{2}\bar{x}^\intercal(T)P(T)\bar{x}(T)$: terminal cost, w/ $x(T) \in \mathcal{X}_{\mathrm{G}} \subset \mathcal{X}_{\text{free}}$ the goal state
- $P(T) \in \mathbb{R}^{n\times n} \succ 0$: final Riccati matrix
- $M \in \mathbb{R}^{n\times n} \succeq 0$, $R \in \mathbb{R}^{m\times m} \succ 0$: user defined, penalize the state and the control
- $\gamma^\star \leq \gamma \in \mathbb{R}^+$: disturbance rejection constant, $\gamma^\star$ smallest value to stabilize system

Two-Player Zero-Sum Game Solution

- Employ the Hamilton-Jacobi-Isaacs equation $\mathscr{H}(\bar{x}; \bar{u}, \bar{d}; \frac{\partial V^\star}{\partial t}, \frac{\partial V^\star}{\partial \bar{x}})$
- Consider a value function quadratic in the state $V^\star(\bar{x}; t) = \frac{1}{2}\bar{x}^\intercal P(t)\bar{x},\ P(t) \succ 0$

The state feedback policy,

$$\bar{u}^\star(\bar{x}; t) = -R^{-1}B^\intercal P(t)\bar{x},\ \forall \bar{x}, t, \tag{5}$$

and worst-case disturbance,

$$\bar{d}^\star(\bar{x}; t) = \gamma^{-2}F^\intercal P(t)\bar{x},\ \forall \bar{x}, t, \tag{6}$$

with the differential Riccati equation,

$$-\dot{P}(t) = P(t)A + A^\intercal P(t) + M + \gamma^{-2}P(t)FF^\intercal P(t) - P(t)BR^{-1}B^\intercal P(t), \tag{7}$$

form a saddle-point equilibrium w/ value $V^\star = \bar{x}_0^\intercal P(t_0)\bar{x}_0$

Two-Player Zero-Sum Game Solution

- Employ the Hamilton-Jacobi-Isaacs equation $\mathscr{H}(\bar{x}; \bar{u}, \bar{d}; \frac{\partial V^\star}{\partial t}, \frac{\partial V^\star}{\partial \bar{x}})$
- Consider a value function quadratic in the state $V^\star(\bar{x}; t) = \frac{1}{2}\bar{x}^\intercal P(t)\bar{x}$, $P(t) \succ 0$

The state feedback policy,

$$\bar{u}^\star(\bar{x}; t) = -R^{-1}B^\intercal P(t)\bar{x},$$

and worst-case disturbance,

$$\bar{d}^\star(\bar{x}; t) = \gamma^{-2}F^\intercal P(t)\bar{x},$$

Challenges
1. Exact knowledge of (A, B, F)
2. Solve (10) backwards in time

with the differential Riccati equation,

$$-\dot{P}(t) = P(t)A + A^\intercal P(t) + M + \gamma^{-2}P(t)FF^\intercal P(t) - P(t)BR^{-1}B^\intercal P(t), \; P(T), \; (\leftarrow)$$

form a saddle-point equilibrium w/ value $V^\star = \bar{x}_0^\intercal P(0)\bar{x}_0$

Q-advantage Function

Define the advantage function as,

$$\mathbb{Q}(\bar{x};\bar{u},\bar{d};t) := V^\star(\bar{x};t) + \mathscr{H}(\bar{x};\bar{u},\bar{d};\frac{\partial V^\star}{\partial t},\frac{\partial V^\star}{\partial \bar{x}}) \tag{8}$$

Lemma 1

The solution of the problem $\mathbb{Q}^\star(\bar{x};\bar{u}^\star,\bar{d}^\star;t) := \min_{\bar{u}} \max_{\bar{d}} \mathbb{Q}(\bar{x};\bar{u},\bar{d};t)$ yields,

$$\mathbb{Q}^\star(\bar{x};\bar{u}^\star,\bar{d}^\star;t) = V^\star(\bar{x},t)$$

Proof.

Substitute $\bar{u}^\star$ and $\bar{d}^\star$, to get $\mathscr{H}^\star = 0$. □

Augmented State Representation

Define the augmented state as $U := [\bar{x}^\intercal \ \bar{u}^\intercal \ \bar{d}^\intercal]^\intercal$ to obtain,

$$\mathbb{Q}(\bar{x};\bar{u};\bar{d};t) = \frac{1}{2}U^\intercal \bar{\mathbb{Q}}(t) U, \\ = \frac{1}{2}\mathrm{vech}(\bar{\mathbb{Q}}(t))^\intercal (U \otimes U),$$

$$\bar{\mathbb{Q}}(t) = \begin{bmatrix} Q_{\mathrm{xx}}(t) & Q_{\mathrm{xu}}(t) & Q_{\mathrm{xd}}(t) \\ Q_{\mathrm{ux}}(t) & Q_{\mathrm{uu}} & Q_{\mathrm{ud}} \\ Q_{\mathrm{dx}}(t) & Q_{\mathrm{du}} & Q_{\mathrm{dd}} \end{bmatrix} = \begin{bmatrix} \dot{P}(t) + P(t) + M + 2P(t)A + P(t)B & P(t)B & P(t)F \\ B^\intercal P(t) & R & 0 \\ F^\intercal P(t) & 0 & -\gamma^2 \end{bmatrix}$$

Model-Free Formulation

From the augmented representation,

$$\bar{\mathbb{Q}}(t) = \begin{bmatrix} Q_{\mathrm{xx}}(t) & Q_{\mathrm{xu}}(t) & Q_{\mathrm{xd}}(t) \\ Q_{\mathrm{ux}}(t) & Q_{\mathrm{uu}} & Q_{\mathrm{ud}} \\ Q_{\mathrm{dx}}(t) & Q_{\mathrm{du}} & Q_{\mathrm{dd}} \end{bmatrix} = \begin{bmatrix} \dot{P}(t) + P(t) + M + 2P(t)A + P(t)B & P(t)B & P(t)F \\ B^{\intercal}P(t) & R & 0 \\ F^{\intercal}P(t) & 0 & -\gamma^2 \end{bmatrix}$$

we obtain the model-free formulations,

$$\bar{u}^{\star}(\bar{x};t) = \arg\min_{\bar{u}} \mathbb{Q}(\bar{x};\bar{u},\bar{d};t) = -{Q_{\mathrm{uu}}}^{-1} Q_{\mathrm{ux}}(t)\bar{x}, \tag{9}$$

$$\bar{d}^{\star}(\bar{x};t) = \arg\max_{\bar{d}} \mathbb{Q}(\bar{x};\bar{u},\bar{d};t) = -{Q_{\mathrm{dd}}}^{-1} Q_{\mathrm{dx}}(t)\bar{x} \tag{10}$$

Actor-Critic Network

Critic network: $W_{\mathrm{c}}^{\mathsf{T}}\nu(t) := \frac{1}{2}\mathrm{vech}(\bar{\mathbb{Q}}(t))$,

$$\hat{\mathbb{Q}}(\bar{x};\bar{u},\bar{d};t) = \hat{W}_{\mathrm{c}}{}^{\mathsf{T}}\nu(t)(U\otimes U) \tag{11}$$

Actor network: $W_{\mathrm{a}}^{\mathsf{T}}\mu(t) := -Q_{\mathrm{uu}}^{-1}Q_{\mathrm{ux}}(t)$,

$$\hat{\bar{u}}(\bar{x};t) = \hat{W}_{\mathrm{a}}{}^{\mathsf{T}}\mu(t)\bar{x} \tag{12}$$

Disturbance network: $\hat{W}_{\mathrm{d}}^{\mathsf{T}}\xi(t) = -Q_{\mathrm{dd}}^{-1}Q_{\mathrm{dx}}(t)$,

$$\hat{\bar{d}}(\bar{x};t) = \hat{W}_{\mathrm{d}}{}^{\mathsf{T}}\xi(t)\bar{x}, \tag{13}$$

$\nu(t)$, $\mu(t)$, $\xi(t)$ bounded radial basis functions that depend explicitly on time

Integral Reinforcement Learning Bellman - Reward Function

Evaluation according to IRL Bellman equation,

$$\begin{aligned}\mathbb{Q}^\star(\bar{x}(t);\hat{\bar{u}}^\star(t),\hat{\bar{d}}^\star(t);t) &= \mathbb{Q}^\star(\bar{x}(t-\Delta t);\hat{\bar{u}}^\star(t-\Delta t),\hat{\bar{d}}^\star(t-\Delta t);t-\Delta t) \\ &\quad - \frac{1}{2}\int_{t-\Delta t}^{t}(\bar{x}^\intercal M\bar{x} + \hat{\bar{u}}^\intercal R\hat{\bar{u}} - \gamma^2\|\hat{\bar{d}}\|^2)\,\mathrm{d}\tau, \qquad (14)\\ \mathbb{Q}^\star(\bar{x}(T),T) &= \frac{1}{2}\bar{x}^\intercal(T)P(T)\bar{x}(T) \qquad (15)\end{aligned}$$

The imposed internal dynamics are described by,

$$\begin{aligned}\dot{p} = &\,\bar{x}^\intercal(t)M\bar{x}(t) - \bar{x}^\intercal(t-\Delta t)M\bar{x}(t-\Delta t) + \bar{u}^\intercal(t)R\bar{u}(t)\\ &-\bar{u}^\intercal(t-\Delta t)R\bar{u}(t-\Delta t) \qquad (16)\end{aligned}$$

Actor-Critic Squared-Norm Errors

Define the actor and critic errors,

$$\left.\begin{aligned} e_{\mathrm{c}_1} &:= \hat{W}_{\mathrm{c}}^{\intercal}\nu(t)\Big((\hat{U}(t)\otimes\hat{U}(t)) - (\hat{U}(t-\Delta t)\otimes\hat{U}(t-\Delta t)\Big) \\ &\quad + \frac{1}{2}\int_{t-\Delta t}^{t}(\bar{x}^{\intercal}M\bar{x} + \hat{\bar{u}}^{\intercal}R\hat{\bar{u}} - \gamma^2\|\hat{\bar{d}}\|^2)\,\mathrm{d}\tau, \\ e_{\mathrm{c}_2} &:= \frac{1}{2}\bar{x}^{\intercal}(T)P(T)\bar{x}(T) - \hat{W}_{\mathrm{c}}(T)^{\intercal}\nu(T)(U(T)\otimes U(T)) \end{aligned}\right\}\text{IRL Bellman}$$

$$e_{\mathrm{a}} := \hat{W}_{\mathrm{a}}^{\intercal}\mu(t)\bar{x} + \hat{Q}_{\mathrm{uu}}^{-1}\hat{Q}_{\mathrm{ux}}(t)\bar{x} \qquad \text{Control}$$

$$e_{\mathrm{d}} := \hat{W}_{\mathrm{d}}^{\intercal}\xi(t)\bar{x} + \hat{Q}_{\mathrm{dd}}^{-1}\hat{Q}_{\mathrm{dx}}(t)\bar{x} \qquad \text{Disturbance}$$

Define the the squared-norm errors,

$$K_1(\hat{W}_{\mathrm{c}},\hat{W}_{\mathrm{c}}(T)) := \frac{1}{2}\|e_{\mathrm{c}_1}\|^2 + \frac{1}{2}\|e_{\mathrm{c}_2}\|^2;\ K_2(\hat{W}_{\mathrm{a}}) := \frac{1}{2}\|e_{\mathrm{a}}\|^2;\ K_3(\hat{W}_{\mathrm{d}}) := \frac{1}{2}\|e_{\mathrm{d}}\|^2$$

Gradient Method for Update Laws

The update law of the critic weights yields,

$$\dot{\hat{W}}_{\mathrm{c}} = -\alpha_{\mathrm{c}} \frac{\partial K_1}{\partial \hat{W}_{\mathrm{c}}} = -\,\alpha_{\mathrm{c}} \left(\frac{1}{(1+\sigma^{\intercal}\sigma)^2} \sigma e_{\mathrm{c}_1} + \frac{1}{(1+\sigma_{\mathrm{f}}^{\intercal}\sigma_{\mathrm{f}})^2} \sigma_{\mathrm{f}} e_{\mathrm{c}_2} \right), \tag{17}$$

$\sigma := \nu(t)(\hat{U}(t) \otimes \hat{U}(t) - \hat{U}(t-\Delta t) \otimes \hat{U}(t-\Delta t))$ and $\sigma_{\mathrm{f}} = \nu(T)(U(T) \otimes U(T))$

The update law of the actor weights yields,

$$\dot{\hat{W}}_{\mathrm{a}} = -\alpha_{\mathrm{a}} \frac{\partial K_2}{\partial \hat{W}_{\mathrm{a}}} = -\alpha_{\mathrm{a}} \bar{x} {e_{\mathrm{a}}}^{\intercal}, \tag{18}$$

$$\dot{\hat{W}}_{\mathrm{d}} = -\alpha_{\mathrm{d}} \frac{\partial K_3}{\partial \hat{W}_{\mathrm{d}}} = -\alpha_{\mathrm{d}} \bar{x} {e_{\mathrm{d}}}^{\intercal}, \tag{19}$$

$\alpha_{\mathrm{c}}, \alpha_{\mathrm{a}}, \alpha_{\mathrm{d}} \in \mathbb{R}^{+}$ the gradient descent gains

Gradient Method for Update Laws

The update law of the critic weights yields,

$$\dot{\hat{W}}_c = -\alpha_c \frac{\partial K_1}{\partial \hat{W}_c} = -\alpha_c \left(\frac{1}{(1+\sigma^\intercal \sigma)^2} \sigma e_{c_1} + \frac{1}{(1+\sigma_f^\intercal \sigma_f)^2} \sigma_f e_{c_2} \right),$$

$$\sigma := \nu(t)(\hat{U}(t) \otimes \hat{U}(t) - \hat{U}(t-\Delta t) \otimes \hat{U}(t-\Delta t)) \text{ and } \sigma_f = \nu(T)(U(T) \otimes U(T)),$$

The update law of the actor weights yields,

$$\dot{\hat{W}}_a = -\alpha_a \frac{\partial K_2}{\partial \hat{W}_a} = -\alpha_a \bar{x} {e_a}^\intercal,$$

$$\dot{\hat{W}}_d = -\alpha_d \frac{\partial K_3}{\partial \hat{W}_d} = -\alpha_d \bar{x} {e_d}^\intercal,$$

$\alpha_c, \alpha_a, \alpha_d \in \mathbb{R}^+$ the gradient descent gains

Learning Steps
1. Measure (x)
2. Compute $e_{c_1}, e_{c_2}, e_a, e_d$
3. Run $\hat{W}_c, \hat{W}_a, \hat{W}_d$
4. Compute $\mathbb{Q}$ and u, d

Gradient Method for Update Laws

The update law of the critic weights yields,

$$\dot{\hat{W}}_{\mathrm{c}} = -\alpha_{\mathrm{c}} \frac{\partial K_1}{\partial \hat{W}_{\mathrm{c}}} = -\,\alpha_{\mathrm{c}} \left(\frac{1}{(1+\sigma^{\mathsf{T}}\sigma)^2} \sigma e_{\mathrm{c}_1} + \frac{1}{(1+\sigma_{\mathrm{f}}^{\mathsf{T}}\sigma_{\mathrm{f}})^2} \sigma_{\mathrm{f}} e_{\mathrm{c}_2} \right),$$

$\sigma := \nu(t)(\hat{U}(t) \otimes \hat{U}(t) - \hat{U}(t-\Delta t) \otimes \hat{U}(t-\Delta t))$ and $\sigma_{\mathrm{f}} = \nu(T)(U(T) \otimes U(T))$,

The update law of the actor weights yields,

Results

1. Model-free control
2. Works forward in time
3. Computationally efficient

Learning Steps

1. Measure (x)
2. Compute $e_{\mathrm{c}_1}, e_{\mathrm{c}_2}, e_{\mathrm{a}}, e_{\mathrm{d}}$
3. Run $\hat{W}_{\mathrm{c}}, \hat{W}_{\mathrm{a}}, \hat{W}_{\mathrm{d}}$
4. Compute $\mathbb{Q}$ and u, d

$\alpha_{\mathrm{c}}, \alpha_{\mathrm{a}}, \alpha_{\mathrm{d}} \in \mathbb{R}^{+}$ the gradient descent gains

Theorem

Lemma 3

Peristently exciting signal

- If the gradient descent gains follow the inequalities,

$$0 < \alpha_{\mathrm{a}} < \frac{\underline{\lambda}(M + Q_{\mathrm{xu}} R^{-1} Q_{\mathrm{xu}}^{\intercal} - \gamma^{-2} Q_{\mathrm{xd}} Q_{\mathrm{xd}}^{\intercal}) - \overline{\lambda}(Q_{\mathrm{xu}} Q_{\mathrm{xu}}^{\intercal})}{\delta_1 \overline{\lambda}\left(\frac{\mu(t) R^{-1}}{\|1+\mu(t)^{\intercal}\mu(t)\|^2}\right)},$$

$$0 < \alpha_{\mathrm{d}} < \frac{\underline{\lambda}(Q_{\mathrm{xd}} Q_{\mathrm{xd}}^{\intercal})}{\delta_2 \overline{\lambda}\left(\frac{\xi(t)\gamma^{-2}}{\|1+\xi(t)^{\intercal}\xi(t)\|^2}\right)}; \quad \alpha_{\mathrm{c}} \gg \alpha_{\mathrm{a}}.$$

- Then, origin is globally uniformly asymptotically stable equilibrium point $\forall\ \psi(0)$
- Closed-loop system with state $\psi = [\bar{x}^{\intercal} \;\; \tilde{W}_{\mathrm{c}}^{\intercal} \;\; \tilde{W}_{\mathrm{a}}^{\intercal} \;\; \tilde{W}_{\mathrm{d}}^{\intercal}]^{\intercal}$

Proof.

Lyapunov-based proof,

$$\mathscr{L}(\psi; t) := V^{\star}(\bar{x}; t) + \tfrac{1}{2}\|\tilde{W}_{\mathrm{c}}\|^2 + \tfrac{1}{2}\mathrm{tr}\{\tilde{W}_{\mathrm{a}}^{\intercal}\tilde{W}_{\mathrm{a}}\} + \tfrac{1}{2}\mathrm{tr}\{\tilde{W}_{\mathrm{d}}^{\intercal}\tilde{W}_{\mathrm{d}}\} > 0 \qquad \square$$

Theorem

Lemma 3

Peristently exciting signal

- If the gradient descent gains follow the inequalities,

$$0 < \alpha_{\mathrm{a}} < \frac{\underline{\lambda}(M + Q_{\mathrm{xu}} R^{-1} Q_{\mathrm{xu}}^{\intercal} - \gamma^{-2} Q_{\mathrm{xd}} Q_{\mathrm{xd}}^{\intercal}) - \overline{\lambda}(Q_{\mathrm{xu}} Q_{\mathrm{xu}}^{\intercal})}{\delta_1 \overline{\lambda}\left(\frac{\mu(t) R^{-1}}{\|1+\mu(t)^{\intercal}\mu(t)\|^2}\right)},$$

$$0 < \alpha_{\mathrm{d}} < \frac{\underline{\lambda}(Q_{\mathrm{xd}} Q_{\mathrm{xd}}^{\intercal})}{\delta_2 \overline{\lambda}\left(\frac{\xi(t)\gamma^{-2}}{\|1+\xi(t)^{\intercal}\xi(t)\|^2}\right)}; \quad \alpha_{\mathrm{c}} \gg \alpha_{\mathrm{a}}$$

- Then, origin is globally uniformly asymptotically stable equilibrium point $\forall\ \psi(0)$
- Closed-loop system with state $\psi = [\bar{x}^{\intercal} \quad \tilde{W}_{\mathrm{c}}^{\intercal} \quad \tilde{W}_{\mathrm{a}}^{\intercal} \quad \tilde{W}_{\mathrm{d}}^{\intercal}]^{\intercal}$

Proof.

Lyapunov-based proof,

$\mathscr{L}(\psi; t) := V^{\star}(\bar{x}; t) + \frac{1}{2}\|\tilde{W}_{\mathrm{c}}\|^2 + \frac{1}{2}\mathrm{tr}\{\tilde{W}_{\mathrm{a}}^{\intercal}\tilde{W}_{\mathrm{a}}\} + \frac{1}{2}\mathrm{tr}\{\tilde{W}_{\mathrm{d}}^{\intercal}\tilde{W}_{\mathrm{d}}\} > 0$ □

Transition to Motion Planning

Question

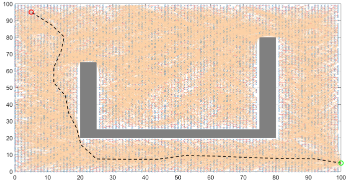
Is the solution of the game sufficient for online and safe kinodynamic motion planning?

Transition to Motion Planning

Question

Is the solution of the game sufficient for online and safe kinodynamic motion planning?

- Offline computation for the RRT*
-



Transition to Motion Planning

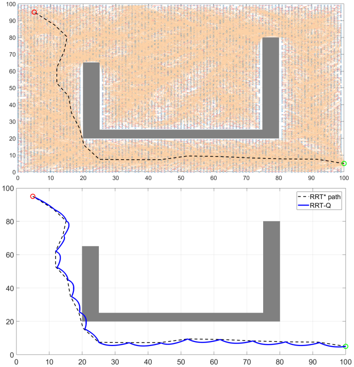
Question

Is the solution of the game sufficient for online and safe kinodynamic motion planning?

- Offline computation for the RRT*
- Apply the proposed control strategy

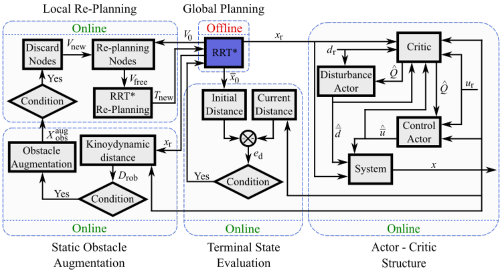
Answer: NO

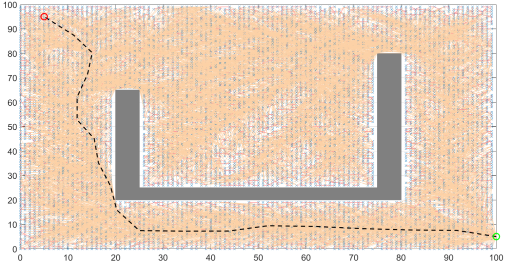
Collision due to kinodynamic constraints



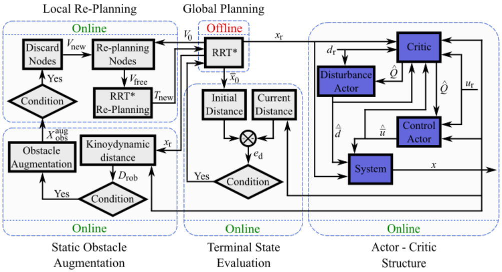
Global Planning

- Offline computation of the RRT*
- Computed path: dashed black line [- -]



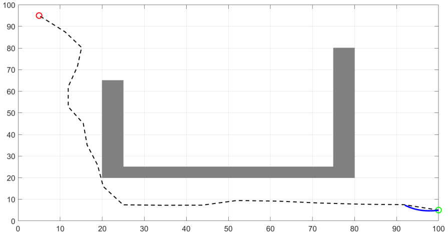


Actor-Critic Structure



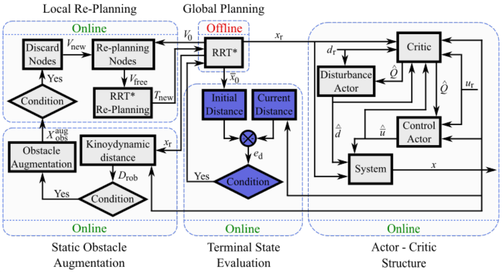
- Online approximation of the optimal policy for the worst-case disturbance

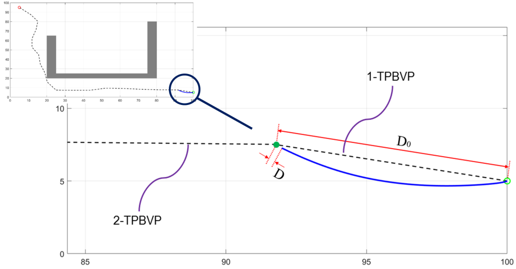
- Motion: blue solid line [—]



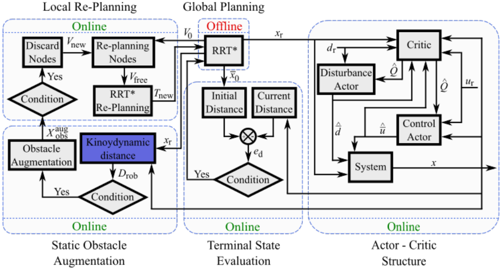
Terminal State Evaluation (Final State Free)

- Initial distance: $D_0(\bar{x}_0) := \|\bar{x}_0\|_n$
- Relative distance: $D(\bar{x}) := \|\bar{x}\|_n$
- Distance error $e_{\mathrm{d}}(\bar{x}_0, \bar{x}) := |D_0(\bar{x}_0) - D(\bar{x})|$
- If $e_d \leq \beta D_0$, we continue to next TPBVP
- β: admissible window

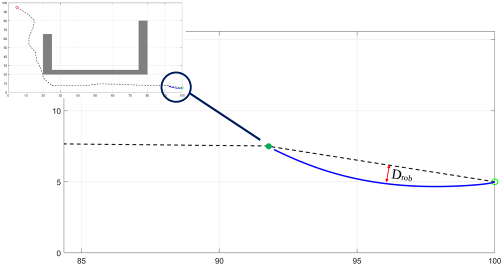




Kinodynamic Distance

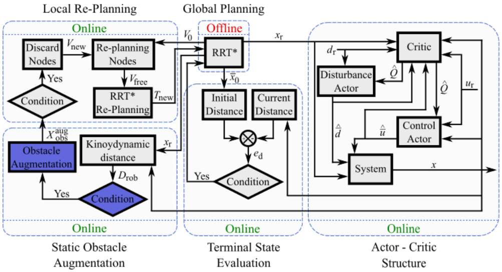


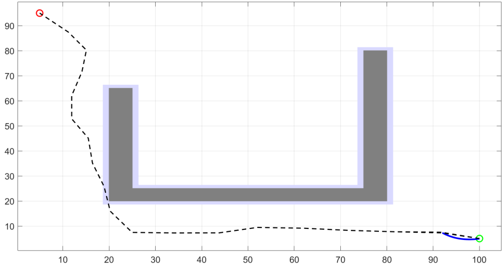
- Kinodynamic distance: $D_{\text{rob}}(\bar{x}_0, \bar{x}) = \frac{|\bar{x}_0 \times \bar{x}|}{D_0(\bar{x}_0)}$
- Maximum deviation of the robot from the straight line D_{rob} at every TPBVP



Obstacle Augmentation

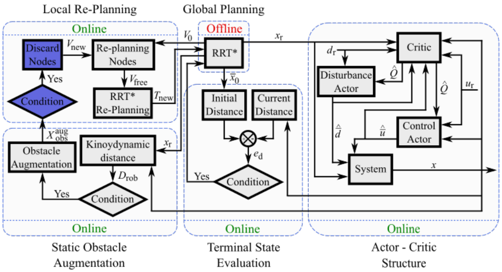
- If kinodynamic distance is greater than the previously measured deviations, $D_{\text{rob},i} > \max\{D_{\text{rob},1}, \ldots, D_{\text{rob},i-1}\}$
- Compute the augmented closed obstacle space, $\mathcal{X}_{\text{obs}}^{\text{aug}} := \mathcal{X}_{\text{obs}} \oplus \mathcal{X}_{\text{rob}}$
- Obstacle augmentation: purple [ ]

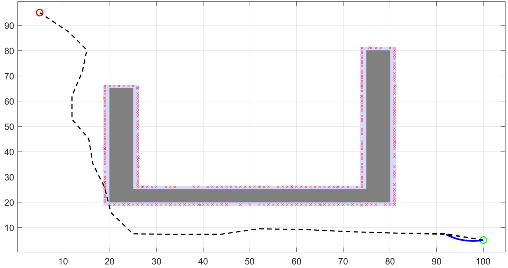




Discard Nodes

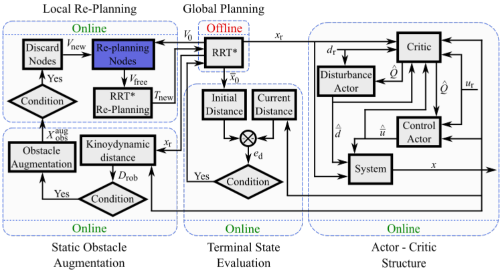
- If collision of global path with the augmented obstacle space occurs
- Discard the nodes of the graph $\mathcal{G}$ in the augmented obstacle space, $V_{\text{new}} = V \backslash V_{\text{aug}},\ V_{\text{aug}} = V \in \mathcal{X}_{\text{obs}}^{\text{aug}}$
- Discarded nodes: red x mark [$\times$]

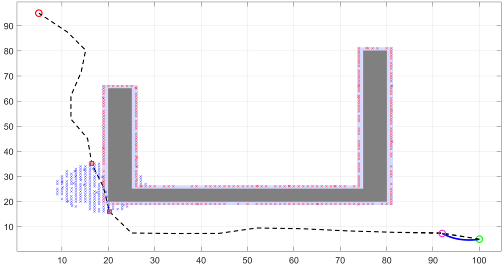




Replanning Area

- Search for the two closest states to the collision area: red rectangles [□]
- Establish a circle wrt to the closest states, $\mathfrak{X}^{\text{loc}}_{\text{circle}} := \{x \in \mathfrak{X} \mid \|x - O_{\text{loc}}\|^2 \leq r^2_{\text{loc}}\}$
- Compute the local free space, $\mathfrak{X}^{\text{loc}}_{\text{free}} := \mathfrak{X}^{\text{loc}}_{\text{circle}} \setminus \mathfrak{X}^{\text{aug}}_{\text{obs}}$
- Local free space nodes: blue x mark [×]





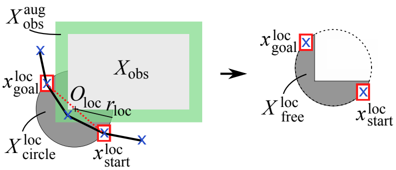
Connected Space Lemma

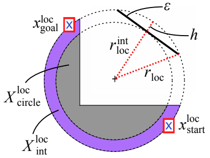
Lemma 2

- If there exists a sufficient, connected, and closed local free space $\mathcal{X}_{\text{free}}^{\text{loc}}$
- Forms a ring, based on the fixed incremental distance ϵ of the RRT*
- Then, we can obtain a collision-free path with the local re-planning framework

Proof.

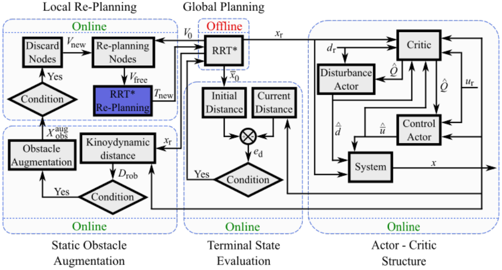
Proof based on topological connectedness tools[1] □

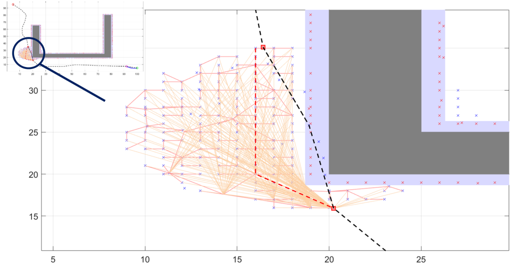




[1] G. P. Kontoudis and K. G. Vamvoudakis, "Kinodynamic Motion Planning with Continuous-Time Q-Learning: An Online, Model-Free, and Safe Navigation Framework," in IEEE Transactions on Neural Networks and Learning Systems 2019.

Local Replanning





- Compute the local path in the reduced local free space with the RRT*
- Edges of the new tree: orange lines [—]
- Local path: red dashed line [- -]

Simulation

For the linear system,

$$\begin{bmatrix} \dot{x}_1 \\ \dot{y}_1 \\ \dot{x}_2 \\ \dot{y}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{20}{40} & 0 & -\frac{45}{40} & 0 \\ 0 & -\frac{20}{40} & 0 & -\frac{45}{40} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{1}{20} & 0 \\ 0 & \frac{1}{20} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} d, \qquad (20)$$

- Set the finite horizon: $T = 10$ s
- User defined matrices, disturbance rejection constant: $M = I$, $R = 0.1I$, $\gamma = 0.5$
- Final Riccati matrix: $P(T) = 0.5I$
- Gradient descent gains: $\alpha_{\mathrm{c}} = 50$, $\alpha_{\mathrm{a}} = 4$, $\alpha_{\mathrm{a}} = 2.5$
- Incremental distance and neighborhood radius of the RRT*: $\epsilon = 2$, $r = 10$
- Start state: $x_{\mathrm{start}} = [90 \quad 10 \quad 0 \quad 0]^{\mathsf{T}}$
- Goal state: $x_{\mathrm{goal}} = [5 \quad 80 \quad 0 \quad 0]^{\mathsf{T}}$

Robust, Online Kinodynamic Motion Planning

Conclusions and Future Work

Conclusions

- Game-based Q-learning to approximate optimal policy for worst-case disturbance
- Stability proof of learning algorithm
- Safe replanning of kinodynamic motion planning with unknown dynamics
- Reduced computational effort to facilitate online and collision-free navigation

Future work:

- Extension to moving obstacle environments
- Deep intermittent Q-learning implementation

George P. Kontoudis and Kyriakos G. Vamvoudakis
Kinodynamic motion planning with continuous-time Q-learning: An online, model-free, and safe navigation framework
IEEE Transactions on Neural Networks and Learning Systems, 2019.

Thank You!