Learning to Steer on Winding Tracks
Using Semi-Parametric Control Policies

Ken Alton
Michiel van de Panne

University of British Columbia
Overview

- Introduction
- Policy Search Method
- Application to Vehicle Steering
- Summary and Future Work
Introduction

- Motion control
  - Constraints
  - Instabilities
Local Track Sensing

- Observation \((\alpha, \beta, d_1, d_2, d_3, d_4)\)
- Reactive policy
- No localization
- Track generalization
Policy Search

- A control policy $\pi$ is a mapping:

$$\pi : O \rightarrow A$$

- A specific policy is given by a vector of policy parameters $\Theta$

- Policy search optimizes $E(\Theta)$ which evaluates policies:

$$\Theta^* = \arg \max_{\Theta} E(\Theta)$$
A policy $\pi$ is defined using $n$ nodes
Each node $i$ has a location $\zeta_i$ in $O$ and an action $\mu_i$ in $A$

\[
\pi(o) = \mu_{c(o)} \\
c(o) = \arg\min_{1 \leq i \leq n} \|o - \zeta_i\|
\]
Useful Properties

- Conceptually simple
- Compact representation
- Local influence of nodes
- Can be adaptively refined
  - Allows shaping of the policy search
Related Work

- Trajectory tracking
  - Divelbiss and Wen, 1997
  - Altafini, Speranzon, and Wahlberg, 2001
  - World model required

- Truck backer-upper problem
  - Nguyen and Widrow, 1989
  - Fixed goal point
  - Global state knowledge
Related Work (2)

- Robot Automobile Racing Simulation (RARS)
  - Pyeatt and Howe, 1998
  - Coulom, 2002
  - Value function approximation
  - Rich sensory information
  - Unstable systems are not considered
  - Do not evaluate generalization
Steering Challenges

- Nonholonomic control problem
- Limited sensing
- Trailer instability: jack-knifing
Control Policy

- Takes an observation \((\alpha, \beta, d_1, d_2, d_3, d_4)\)
- Outputs a desired steering angle \(\alpha_d\)
- \(\alpha\) is driven towards \(\alpha_d\) by the dynamical system:

\[
\ddot{\alpha} = k_p (\alpha - \alpha_d) - k_d \dot{\alpha}
\]
Policy Evaluation

\[ d(s, s') = d_1 + d_2 + d_3 \]

- \( E(\Theta) \) measures the distance covered from a number of starting positions
- Distribution defines a tradeoff
Policy Search

- Stochastic greedy ascent algorithm
- Series of optimization episodes
  - with each episode a new policy node is added
- Coarse steering policy is adaptively refined, shaping the policy search
Track Generation

- Policy is trained on a single track.
- Policy is tested on tracks of the same class.
- In some cases, the policy is tested in very different scenarios.
- Randomized track generator.
Backward-Driving One-Trailer Truck

- 13476 policy evaluations
- 14 policy nodes
Backward-Driving Two-Trailer Truck

- 15482 policy evaluations
- 30 policy nodes
Backward-Driving Car

- 3665 policy evaluations
- 22 policy nodes
Forward-Driving Car on Wide Track

- 1342 policy evaluations
- 21 policy nodes
Discussion

Advantages

- Compact policies for difficult control problems
- Local policy representation
- Adaptive refinement
- Generalization

Disadvantages

- No theoretical safety guarantees
- Policy search is slow
Summary

- Nearest-neighbour policy representation and policy search to solve several non-holonomic vehicle steering problems
- Shaping the policy search through the adaptive addition of nodes
Future Work

- Smoother basis functions
- Preprocess observable state
- Theoretical statements
- Other applications
- Noisy observations