

# Reinforcement-Lernen

Wintersemester 10/11

Aufgabenzettel 5

**Abgabe:** 04.02.2011 **vor** der Vorlesung.

## Aufgabe 10 (50%). Programmieraufgabe: CMA-ES

Das *Cart Pole* Problem (siehe Abbildung 1) soll mit Hilfe der CMA-ES (Covariance matrix adaptation evolution strategy) gelöst werden. Der von der CMA-ES verwendete Evolutionszyklus ist in `cartPoleCMA-ES.cpp` gegeben. Ergänze die fehlende Berechnung der Fitnessfunktion. Verwende dabei die Dateien `SinglePole.h`, `RLTask.h`, `Policy.h` und `cartPoleCMA-ES.cpp`.

Welchen Einfluss hat die initiale globale Schrittweite (`GlobalStepInit`)?

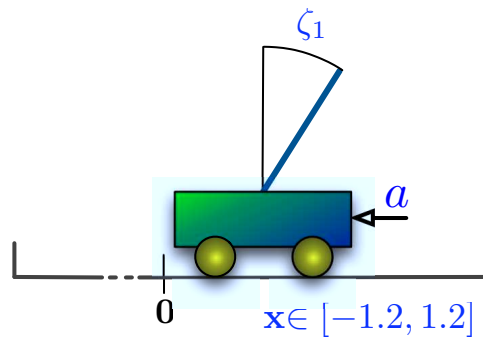


Abbildung 1: Auf einem 1-dimensionalen Wagen ist ein Stab montiert, der möglichst lange balanciert werden soll. Der Zustandsraum ist durch die Position des Wagens, die Geschwindigkeit des Wagens, der Winkel des Stabes und die Winkelgeschwindigkeit des Stabes gegeben. Um den Stab zu balancieren kann als Aktion pro Zeitschritt eine (kontinuierliche) Kraft auf den Wagen ausgeübt werden.

## Aufgabe 11 (50%). LSTD- $\lambda$

1. Welche Zeilen des LSTD - $\lambda$  Algorithmus müssen modifiziert werden, um statt eines “Least Squares TD( $\lambda$ )” eine “Least Squares SARSA( $\lambda$ )” zu implementieren?
2. Wo wird der *temporal difference*-Fehler berechnet?
3. Zeige: Für  $\lambda = \gamma = 1$  und mit einer Tabelle als Funktionsapproximator gilt:

$$\mathbf{A} = \mathbf{N} .$$

(Dabei sind die Matrizen  $\mathbf{A}$  und  $\mathbf{N}$  wie in der Vorlesung definiert.)