Pink Noise is All You Need

Colored Noise Exploration in Deep Reinforcement Learning

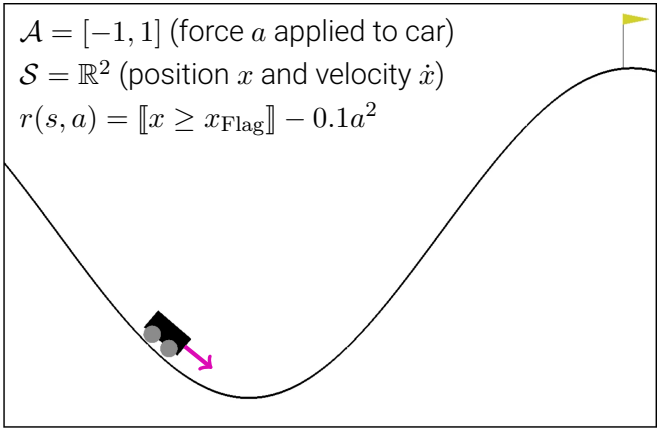
Onno Eberhard · Jakob Hollenstein · Cristina Pinneri · Georg Martius

onnoeberhard@gmail.com
University of Tübingen
Max Planck Institute for Intelligent Systems

October 12, 2022

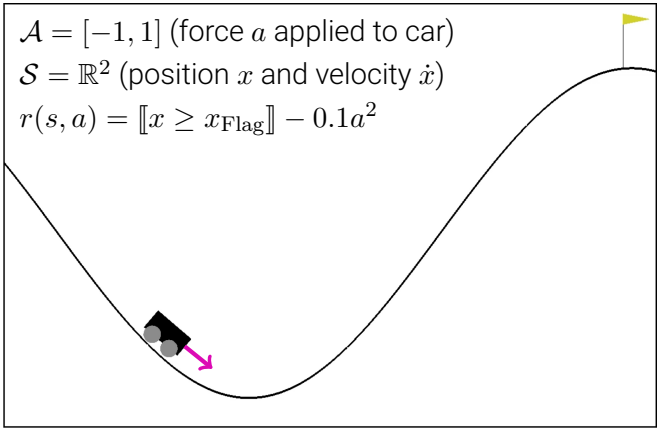
Introduction

- Setting: Reinforcement learning for continuous control
- Example: MountainCar problem



Exploration

- Why is exploration necessary for this problem?
- For an untrained policy, $r(s, a) \approx -0.1a^2$
- The agent will learn to apply no force!



Exploration

- Exploration means **not** performing the action the policy proposes: $a_t \neq \mu(s_t)$
- How is exploration usually done in continuous control?
  - Simply add some noise ε_t to the actions.
- Deterministic policies (e.g. TD3): Fixed noise scale σ

$$a_t = \mu(s_t) + \sigma\varepsilon_t$$

- Stochastic policies (e.g. SAC, MPO): Learn noise scale function

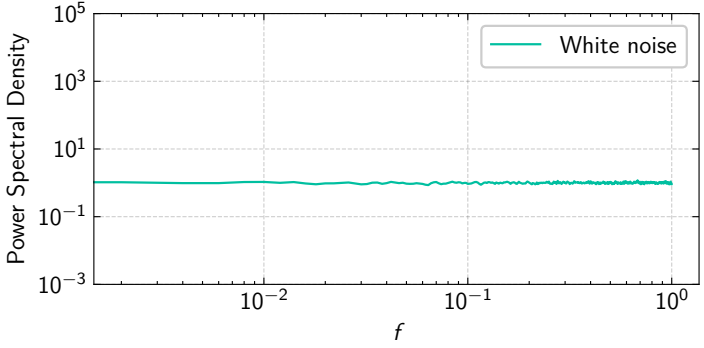
$$a_t = \mu(s_t) + \sigma(s_t) \odot \varepsilon_t$$

- Usually, $\varepsilon_t \sim \mathcal{N}(0, I)$ is sampled independently at every time step

White Noise

- If $\varepsilon_t \sim \mathcal{N}(0, I)$ independently at every time step, then $\varepsilon_{1:T}$ is called **white noise**
  - Why?
- The **power spectral density** (PSD) is defined for any signal $\varepsilon(t)$ as

$$|\hat{\varepsilon}(f)|^2 \quad \text{where} \quad \hat{\varepsilon}(f) = \mathcal{F}[\varepsilon(t)](f)$$



White Noise

- How well does white noise exploration work on MountainCar?
  - Demonstration!
  - ☹
  - Why so bad?
- Let's look at an even simpler environment: an integrator

$$s_t = \sum_{\tau=1}^{t} a_\tau$$

- How far does white noise reach ($a_\tau \sim \mathcal{N}(0, 1)$)?

$$s_t \sim \mathcal{N}(0, t) \quad \rightsquigarrow \quad \mathbb{E}[|s_t|] \propto \sqrt{t}$$

- White noise is slow, because it is temporally uncorrelated ($\mathrm{cov}[\varepsilon_t, \varepsilon_{t'}] = 0$).

Ornstein-Uhlenbeck Noise

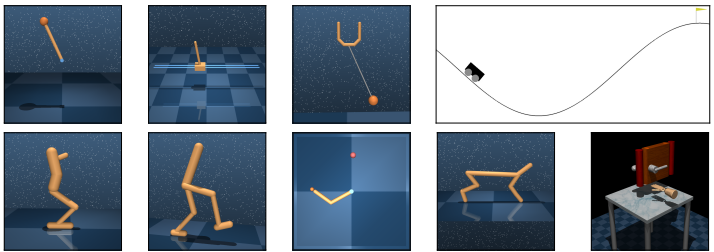
- Simple fix: Just use a temporally correlated noise process ($\mathrm{cov}[\varepsilon_t, \varepsilon_{t'}] > 0$).
- **Brownian motion** (s_t from the previous slide) is temporally correlated
  - $z_t = \int_0^t s_\tau \, \mathrm{d}\tau \quad \rightsquigarrow \quad \mathbb{E}[|z_t|] \propto t^{3/2}$
  - Not stationary (variance increases without bounds)
  - Unsuitable as action noise
- This can be fixed with **Ornstein-Uhlenbeck** (OU) noise:

$$\dot{\varepsilon}_t = -\theta \varepsilon_t + \sigma \eta_t \quad \text{where} \quad \eta_t \sim \mathcal{N}(0, 1)$$

- Equivalent to Brownian motion if $\theta = 0$, stationary if $\theta > 0$
- How does it perform on MountainCar ($a_{1:T} \sim \mathrm{OU}_T, \theta = 0.15$)?
  - Better!
  - OU noise is the default action noise on DDPG.

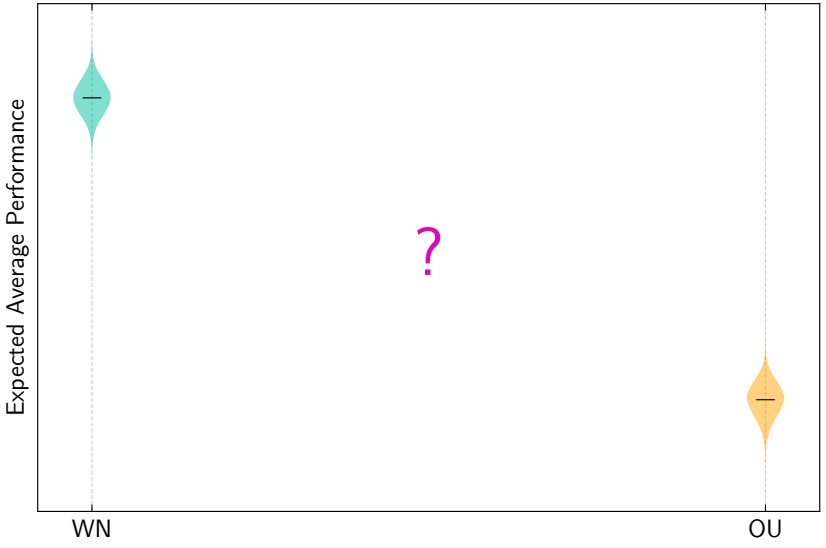
Experiments

- How do white and OU action noise perform on other environments?
- We perform experiments on a number of benchmark tasks using MPO and SAC.



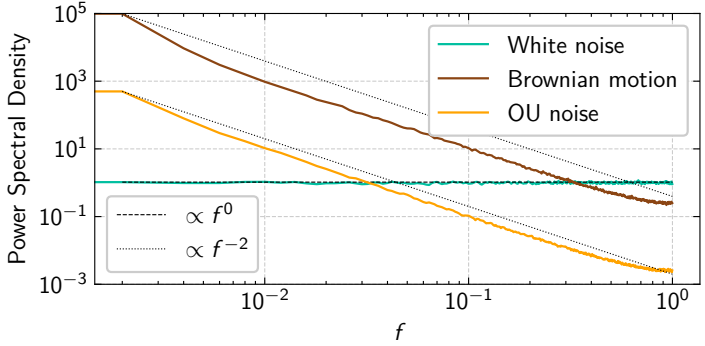
- Performance = mean evaluation return over training.
- We report the “average performance” as the normalized performance averaged over all environments

Results



Intermediate Temporal Correlation

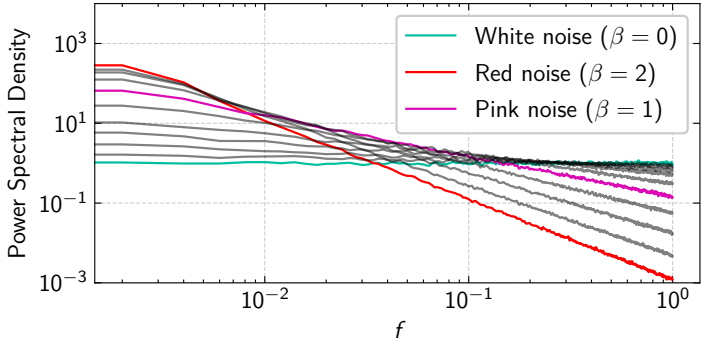
- Uncorrelated action noise (white noise) fails at hard exploration tasks
- Strongly correlated noise (OU noise) yields bad results due to off-policy data
- Action noise with **intermediate temporal correlation** might be more general



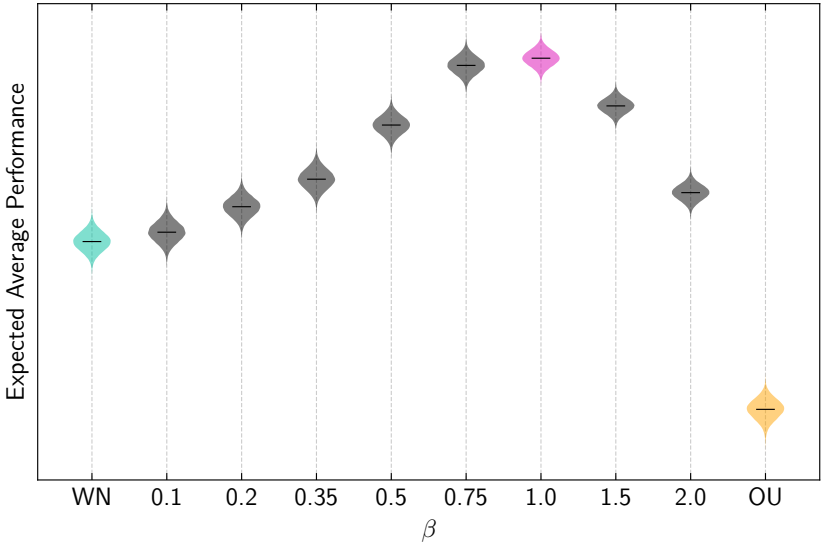
- Idea: Change the exponent of $f^{-\beta}$ to be $0 < \beta < 2$

Colored Noise

- Noise with a PSD proportional to $f^{-\beta}$ is called **colored noise** with color β
- White noise is colored noise with $\beta = 0$
- Brownian motion is colored noise with $\beta = 2$ (also called red noise)
- Stationary colored noise signals can be efficiently generated ($\varepsilon_{1:T} \sim \mathrm{CN}_T(\beta)$)



Colored Action Noise: Results



Pink Noise

- The results show that intermediate temporal correlation is a better default
- In particular, **pink noise** (CN with $\beta = 1$) performs best across tasks
  - We have also experimented with β-schedules, choosing β randomly for each rollout, and using a bandit algorithm to optimize β online.
  - Pink noise performed better than all other methods

Why does pink noise work so well as a default?

- "Faster" than white noise, but more on-policy than OU noise
  - Works both on tasks that prefer white noise, and on those that prefer OU noise
- Many environments actually prefer pink noise over both white and OU noise!
  - Not just the "best compromise"
  - Why?

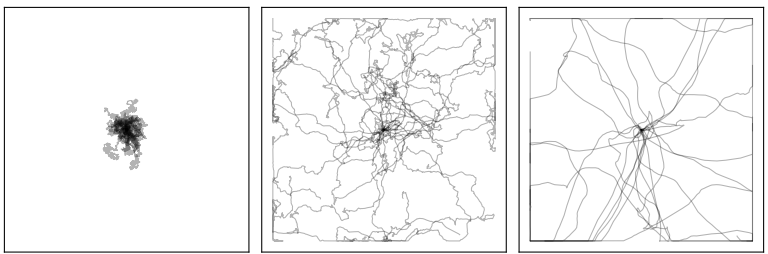
A Bounded Integrator

- Let us look at a simple 2-dimensional bounded integrator environment:

$$s_{t+1} = \text{clip}(s_t + a_t, -c, c)$$

- Parameterized by its size (area = $4c^2$)

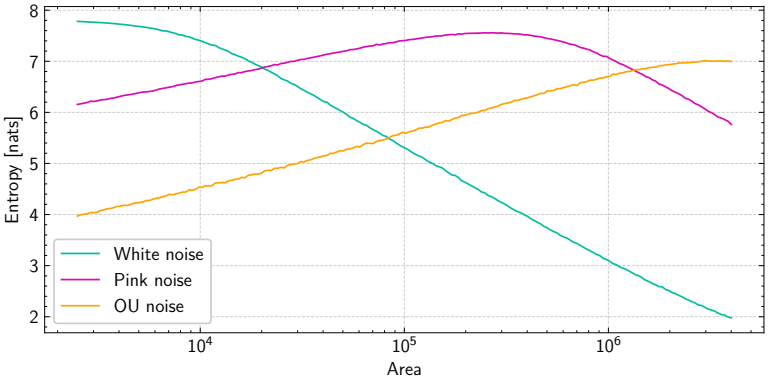
White noise | Pink noise | OU noise



$c = 250$

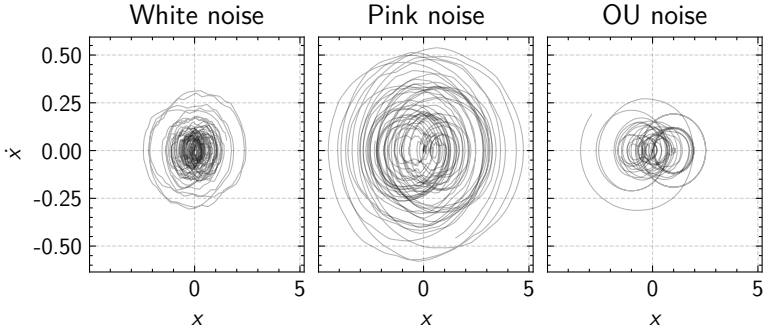
A Bounded Integrator

- Exploration and state space coverage can be measured by the entropy
- We divide the space into B bins, run N trajectories and estimate the state-visitation distribution by a histogram ($B = 2500, N = 10^4$)
- This can be repeated for a large range of environment sizes



A Harmonic Oscillator

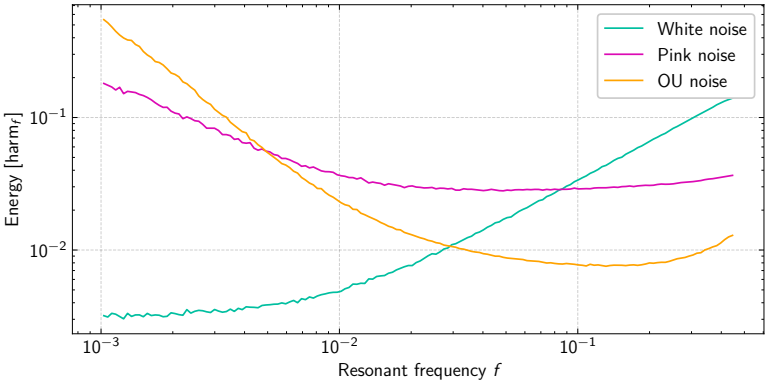
- Let us look at a second simple environment: a harmonic oscillator.
- A mass m is attached to a spring of stiffness k (no friction or gravity)
- Parameterized by its resonant frequency $f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$
- State: position x and velocity $\dot{x}$, Action: Force applied to mass



$f = 2 \times 10^{-2}$

A Harmonic Oscillator

- This oscillator is related to environments like MountainCar
- In these tasks, the exploration noise has to swing up the oscillator
- This can be measured by the energy in the system: $E = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}kx^2$
- We can vary f over the complete sensible range and measure the average energy



The Power of Pink

- In two very different settings we see very similar results
- If we don't know the exact parameterization (e.g. high or low c or f) of a given environment, pink action noise is the safest bet!
- This explains the average performance results we saw before
  - Many different tasks with different preferences $\rightarrow$ pink noise is most general

Takeaway

- Use **pink noise** as the default action noise.

Q & A

Thanks for listening!