Pink Noise Is All You Need

Colored Noise Exploration in Deep Reinforcement Learning

Onno Eberhard[1] · Jakob Hollenstein[2,1] · Cristina Pinneri[1,3] · Georg Martius[1]

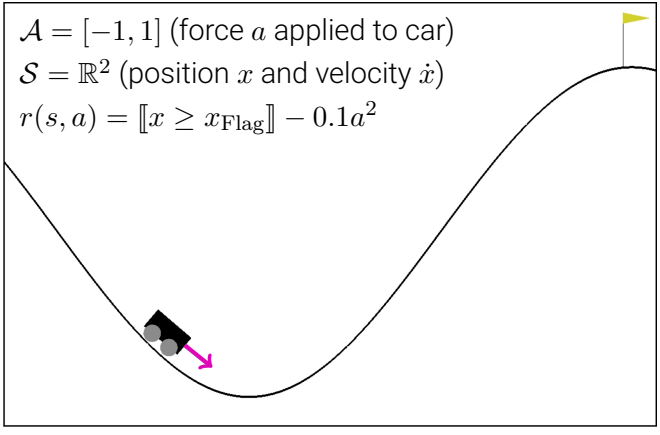
[1]Max Planck Institute for Intelligent Systems [2]Universität Innsbruck [3]ETH Zürich

AUTONOMOUS LEARNING
MAX PLANCK INSTITUTE
FOR INTELLIGENT SYSTEMS

ICLR 2023 · Kigali, Rwanda

Introduction

- Setting: Reinforcement learning for continuous control
- Mountain-car problem: Why is exploration necessary?



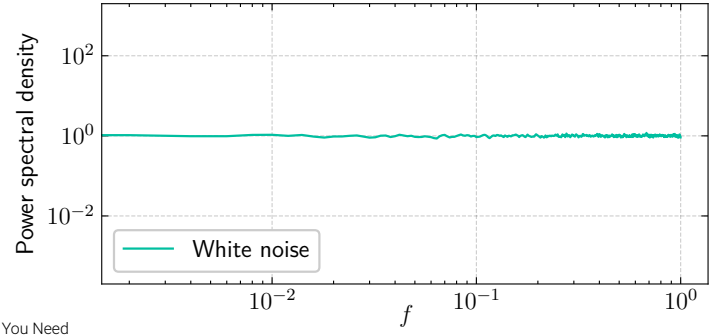
White Noise Exploration

- Usual method for exploration: add some noise ε_t to actions
- If $\varepsilon_t \sim \mathcal{N}(0, I)$ independently at every time step, then $\varepsilon_{1:T}$ is called **white noise**
  - Used as default by many algorithms: TD3, SAC, MPO, ...

White Noise Exploration

- Usual method for exploration: add some noise ε_t to actions
- If $\varepsilon_t \sim \mathcal{N}(0, I)$ independently at every time step, then $\varepsilon_{1:T}$ is called **white noise**
  - Used as default by many algorithms: TD3, SAC, MPO, ...
- The **power spectral density** (PSD) is defined for any signal $\varepsilon(t)$ as

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



Temporal Correlation

- White noise has no temporal correlation ($\mathrm{cov}[\varepsilon_t, \varepsilon_{t'}] = 0$)
- This makes exploration very slow, simple tasks like Mountain-car challenging

Temporal Correlation

- White noise has no temporal correlation ($\mathrm{cov}[\varepsilon_t, \varepsilon_{t'}] = 0$)
- This makes exploration very slow, simple tasks like Mountain-car challenging
- Simple fix: Use a temporally correlated noise process ($\mathrm{cov}[\varepsilon_t, \varepsilon_{t'}] > 0$)
- Popular choice: Ornstein-Uhlenbeck (OU) noise

Temporal Correlation

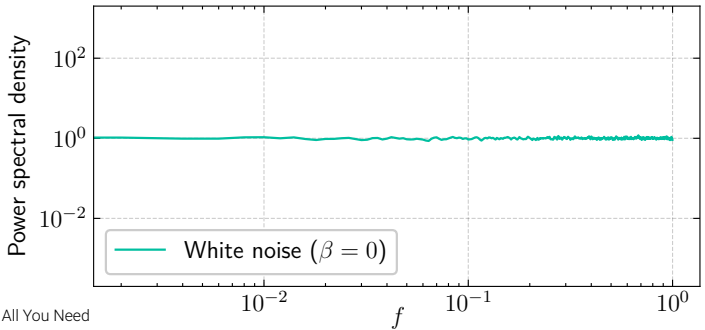
- White noise has no temporal correlation ($\mathrm{cov}[\varepsilon_t, \varepsilon_{t'}] = 0$)
- This makes exploration very slow, simple tasks like Mountain-car challenging
- Simple fix: Use a temporally correlated noise process ($\mathrm{cov}[\varepsilon_t, \varepsilon_{t'}] > 0$)
- Popular choice: Ornstein-Uhlenbeck (OU) noise
- Problem: Very strong temporal correlation $\rightarrow$ poor performance if not needed
- Idea: Use **intermediate temporal correlation** to get best of both worlds

Colored Noise

- Noise with a PSD proportional to $f^{-\beta}$ is called **colored noise** with color β
- Color parameter β controls strength of temporal correlation

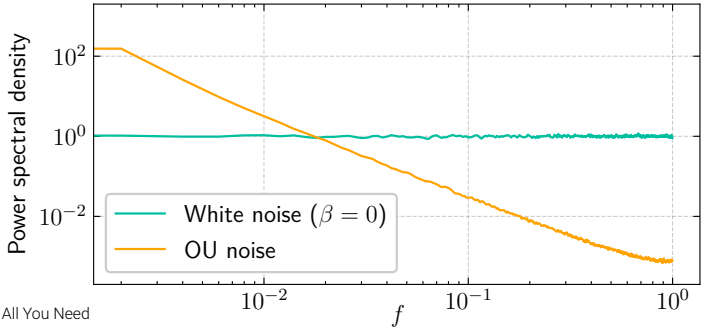
Colored Noise

- Noise with a PSD proportional to $f^{-\beta}$ is called **colored noise** with color β
- Color parameter β controls strength of temporal correlation
- White noise is colored noise with $\beta = 0$



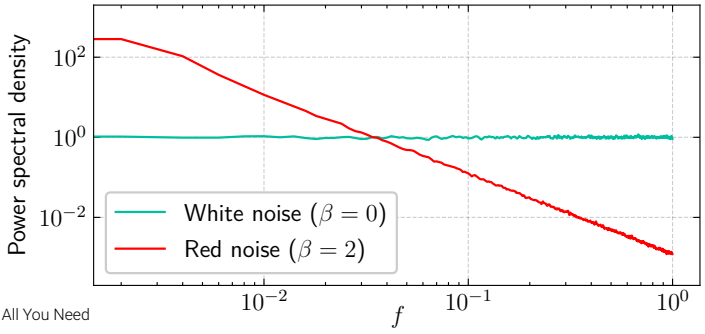
Colored Noise

- Noise with a PSD proportional to $f^{-\beta}$ is called **colored noise** with color β
- Color parameter β controls strength of temporal correlation
- White noise is colored noise with $\beta = 0$
- OU noise is related to red noise (CN with $\beta = 2$)



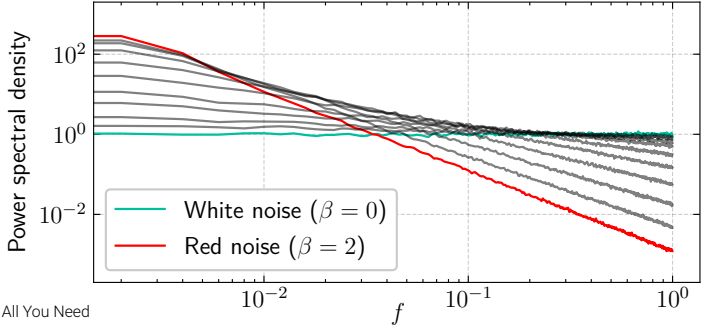
Colored Noise

- Noise with a PSD proportional to $f^{-\beta}$ is called **colored noise** with color β
- Color parameter β controls strength of temporal correlation
- White noise is colored noise with $\beta = 0$
- OU noise is related to red noise (CN with $\beta = 2$)



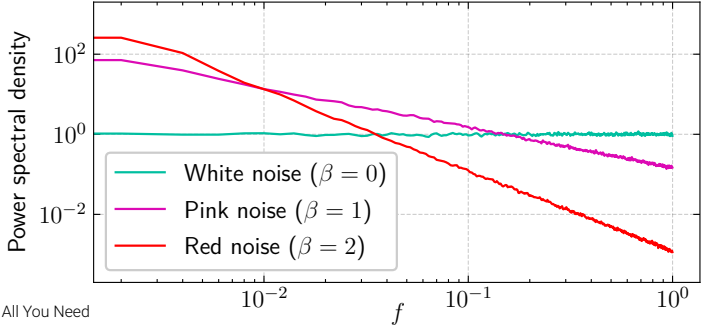
Colored Noise

- Noise with a PSD proportional to $f^{-\beta}$ is called **colored noise** with color β
- Color parameter β controls strength of temporal correlation
- White noise is colored noise with $\beta = 0$
- OU noise is related to red noise (CN with $\beta = 2$)
- Colored noise with intermediate correlation ($\beta \in [0, 2]$) is cheap to generate



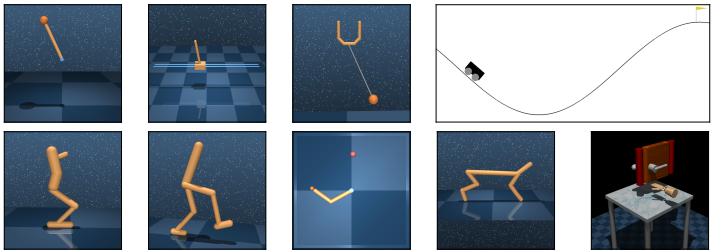
Colored Noise

- Noise with a PSD proportional to $f^{-\beta}$ is called **colored noise** with color β
- Color parameter β controls strength of temporal correlation
- White noise is colored noise with $\beta = 0$
- OU noise is related to red noise (CN with $\beta = 2$)
- Colored noise with intermediate correlation ($\beta \in [0, 2]$) is cheap to generate



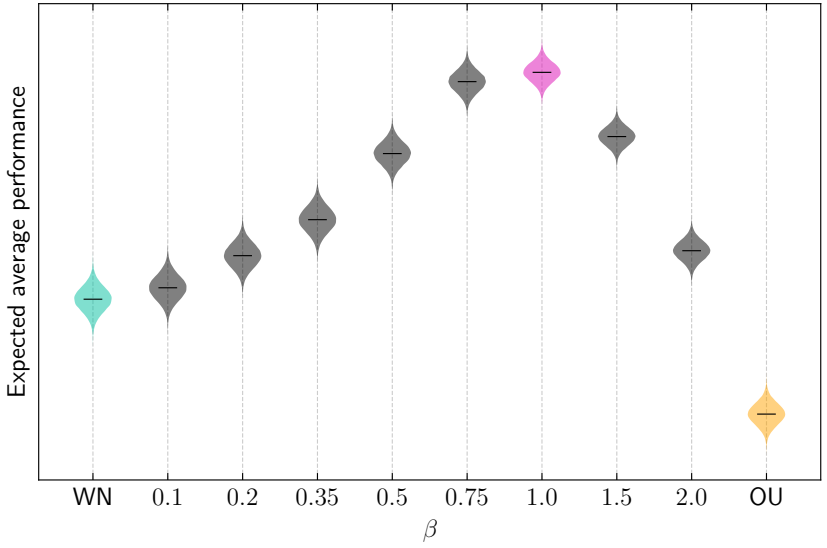
Experiments

- We perform experiments on a number of benchmark tasks using MPO and SAC

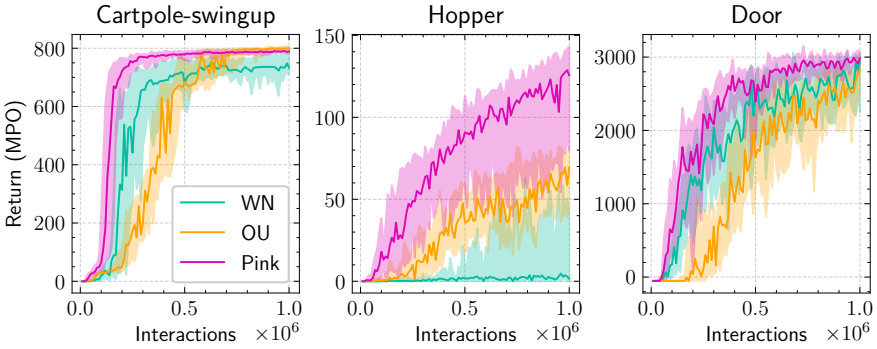


- Measure **average performance** (mean normalized performance across all tasks)
  - Default action noise should work well everywhere

Results


Expected average performance
WN
0.1
0.2
0.35
0.5
0.75
1.0
1.5
2.0
OU
β

Results



- Pink noise works well on **all** environments we tested
- Not true for white noise or OU noise!

Pink Noise

- Other experiments: β-schedules, random β selection, bandit β selection
- Pink noise performed better than all these methods

Pink Noise

- Other experiments: β-schedules, random β selection, bandit β selection
- Pink noise performed better than all these methods

Why does pink noise work so well as a default?

- Works very well on some environments
- Works well on all environments

A Bounded Integrator

- Simple 2-dimensional “bounded integrator” environment:

$$\boldsymbol{s}_{t+1} = \text{clip}(\boldsymbol{s}_t + \boldsymbol{a}_t, -c\mathbf{1}, +c\mathbf{1})$$

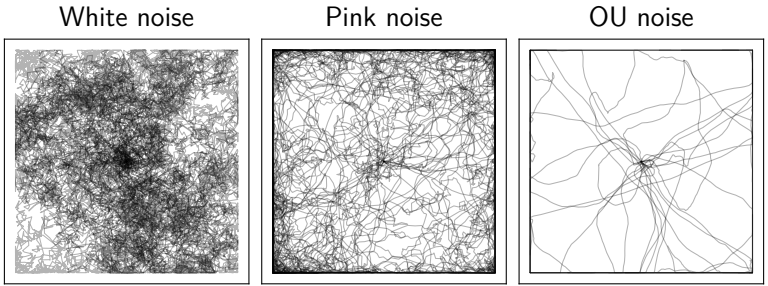
- Parameterized by its size (area = $4c^2$)

A Bounded Integrator

- Simple 2-dimensional “bounded integrator” environment:

$$\boldsymbol{s}_{t+1} = \mathrm{clip}(\boldsymbol{s}_t + \boldsymbol{a}_t, -c\mathbf{1}, +c\mathbf{1})$$

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

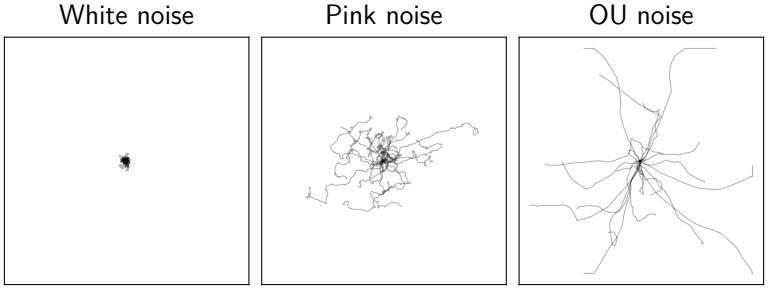


A Bounded Integrator

- Simple 2-dimensional "bounded integrator" environment:

$$\boldsymbol{s}_{t+1} = \text{clip}(\boldsymbol{s}_t + \boldsymbol{a}_t, -c\mathbf{1}, +c\mathbf{1})$$

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



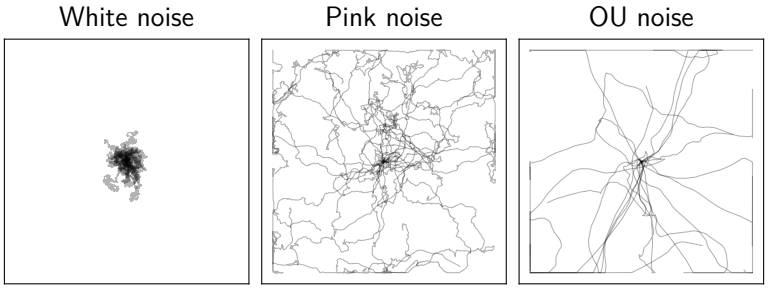
Large space ($c = 1000$)

A Bounded Integrator

- Simple 2-dimensional “bounded integrator” environment:

$$\boldsymbol{s}_{t+1} = \text{clip}(\boldsymbol{s}_t + \boldsymbol{a}_t, -c\mathbf{1}, +c\mathbf{1})$$

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



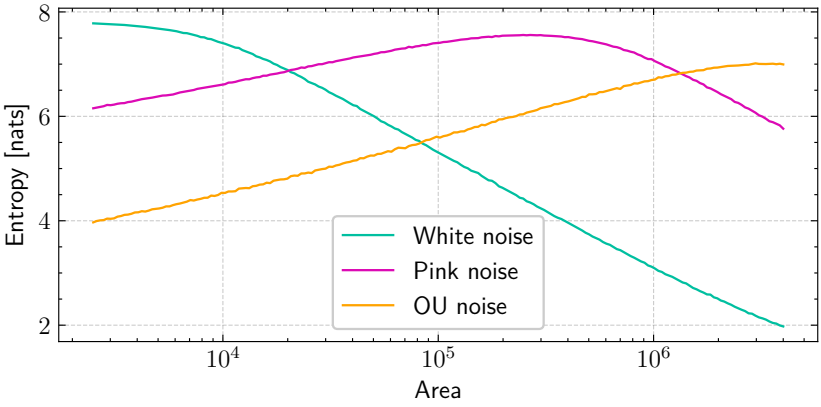
Medium space ($c = 250$)

A Bounded Integrator

- Measure exploration by estimating state-visitation entropy
- Repeat for a large range of environment sizes

A Bounded Integrator

- Measure exploration by estimating state-visitation entropy
- Repeat for a large range of environment sizes



The Power of Pink

- Very similar results on a second simplified environment
- Pink noise is **general**: less sensitive to the environment parameterization
- Explains average performance results (benchmark experiments)
  - Many different tasks with different preferences $\rightarrow$ general noise preferable

The Power of Pink

- Very similar results on a second simplified environment
- Pink noise is **general**: less sensitive to the environment parameterization
- Explains average performance results (benchmark experiments)
  - Many different tasks with different preferences → general noise preferable

Takeaway

- Try **pink noise** as the default action noise

```
pip install pink-noise-rl
```

Thank you!

More Info:

`https://bit.ly/pink-noise-rl`