Guiding Evolutionary Strategies (🧬) by Differentiable Robot (🤖) Simulators

Vladislav Kurenkov*
Tinkoff
v.kurenkov@tinkoff.ru

Bulat Maksudov
Independent Researcher
b.maksudov@lua.tatar

NEURAL INFORMATION PROCESSING SYSTEMS

Motivation

Evolutionary Strategies is a class of zeroth-order black-box optimization algorithms that were successfully applied in various simulated robotics tasks. They are known as easy to parallelize, and a small number of hyperparameters makes them easy to tune. However, **these methods were observed to exhibit a high sample complexity** and a strong dependence on an initial network initialization making their adoption for training robots directly in reality troublesome.

On the other hand, there is an emerging field of **Differentiable Robot Simulators**, which already demonstrated an ability **to find successful control policies with only a handful of trajectories.** However, these methods may result in misleading gradients (e.g. when handling collisions) and it is not clear how to apply them for training robots directly in real life.

Here, **we propose a simple way to utilize gradients from Differentiable Robot Simulators to accelerate training of Evolutionary Strategies in the real world.**

Proposed Approach

Guided Evolutionary Strategies with Differentiable Simulators

We rely on a recently introduced algorithm — Guided Evolutionary Strategies. This method can make use of any surrogate gradients that are correlated with the true gradient to accelerate the convergence of evolutionary strategies. The surrogate gradients can be corrupted or biased in any way, the only requirement for them is to preserve a positive correlation with the true gradient. We presume that this is the case for the gradients computed with DRS, therefore propose to use it as a surrogate.

Algorithm 1 Evolutionary Strategies Guided by DRS

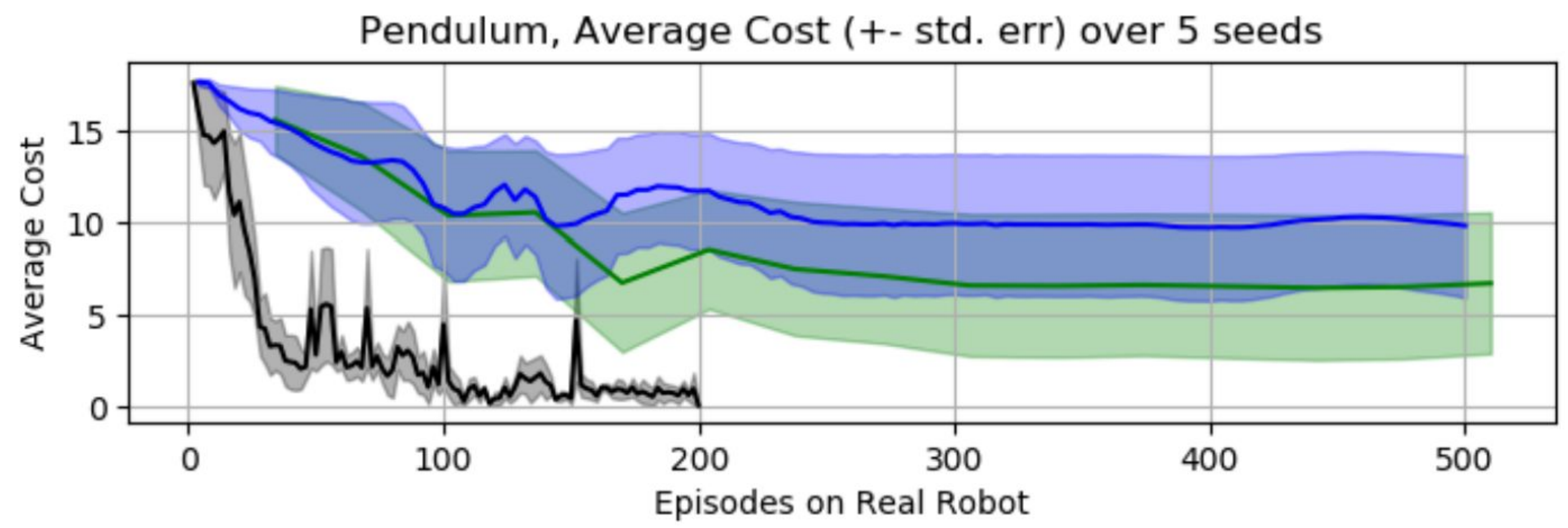
Input: Initial solution θ_0; Optimizer opt; Cost function $f(\theta)$; DRS Gradient ∇f_{drs}
Output: Final solution θ_T

for $t = 0$ to T **do**
    // DRS Part
    Get DRS gradient $\nabla f_{drs}(\theta_t)$ — **Simulation**

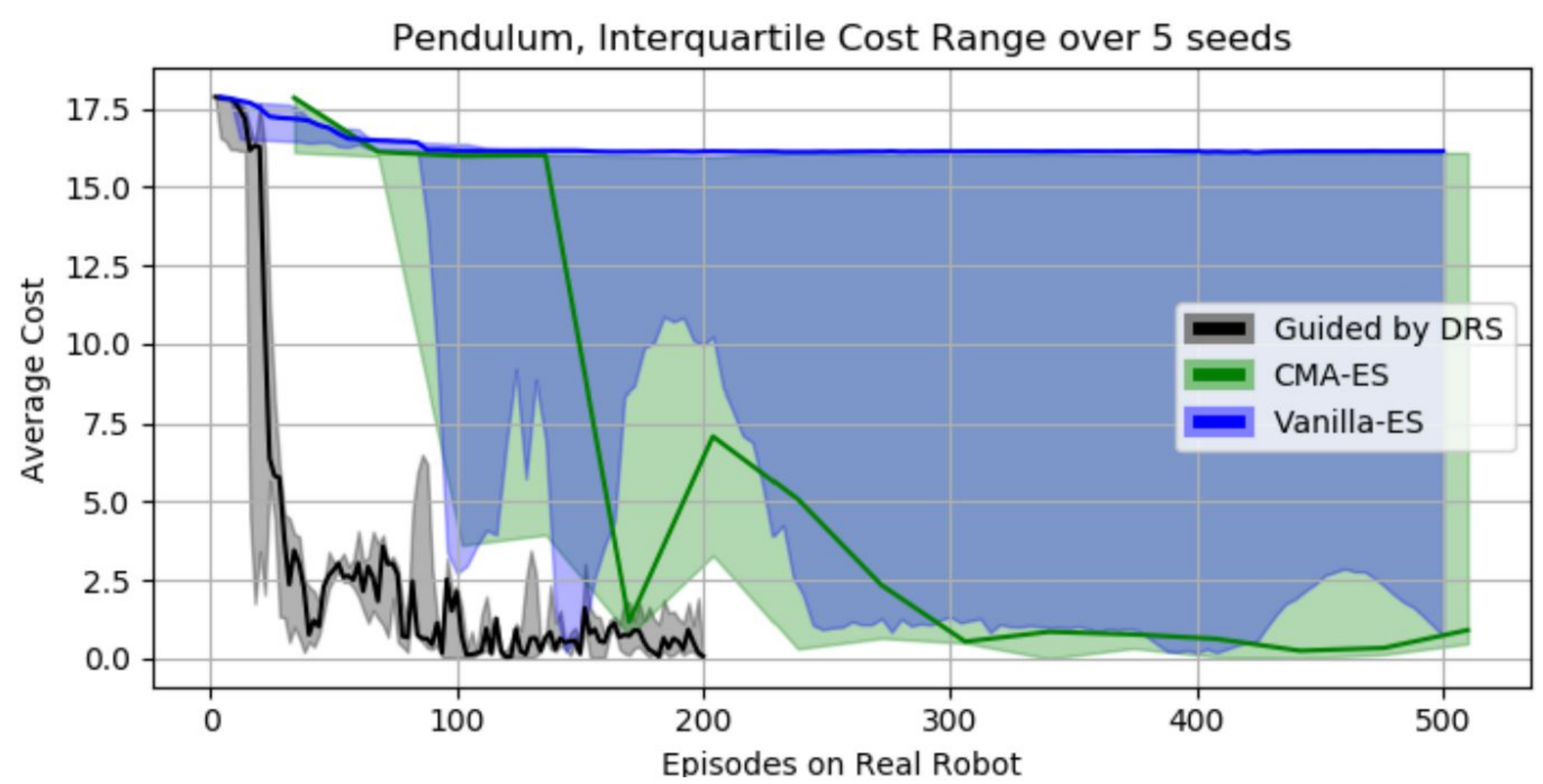
    // Guided-ES Part
    Update low-dimensional guiding subspace U with the DRS gradient
    Define search covariance $\Sigma = \frac{\alpha}{n} I + \frac{1-\alpha}{k} UU^T$
    for $i = 1$ to P **do** — **Real Robot**
        Sample perturbation $\epsilon_i \sim \mathcal{N}(0, \sigma^2 \Sigma)$
        Compute antithetic pair of losses $f(\theta_t + \epsilon_i)$ and $f(\theta_t - \epsilon_i)$
    end for
    Compute Guided ES gradient estimate $g = \frac{\beta}{2\sigma^2 P} \sum_{i=1}^{P} \epsilon_i [f(\theta_t + \epsilon_i) - f(\theta_t - \epsilon_i)]$
    Update parameters using given optimizer $\theta_{t+1} = opt.step(\theta_t, g)$
end for
return θ_T

Accelerated Learning on Real Robot

Training a control policy for Swinging Pendulum using Guided-ES



Guided-ES with DRS achieves considerably lower sample complexity for training on real robot in comparison to vanilla evolutionary strategies (Vanilla-ES) and Covariance Matrix Adaptation Evolution Strategy (CMA-ES).



Our method is more robust across random seeds

Evolutionary strategies and their variants were observed to be on par with reinforcement learning algorithms, but in some cases require a larger amount of training episodes. Algorithms of this class were successfully applied to train robots in simulation at the expense of higher experimental time. **We show that incorporating information from DRS into the training process accelerates the convergence of evolutionary strategies.**

We compare Guided-ES with DRS against Vanilla-ES and CMA-ES. We observe that the proposed method converges faster than both Vanilla-ES and CMA-ES, moreover, the convergence is robust to random seeds in opposition to other considered algorithms. We notice that **Vanilla-ES and CMA-ES are highly dependent on initial network initialization, which is not the case for Guided-ES with DRS**

When DRS Gradients are Misleading

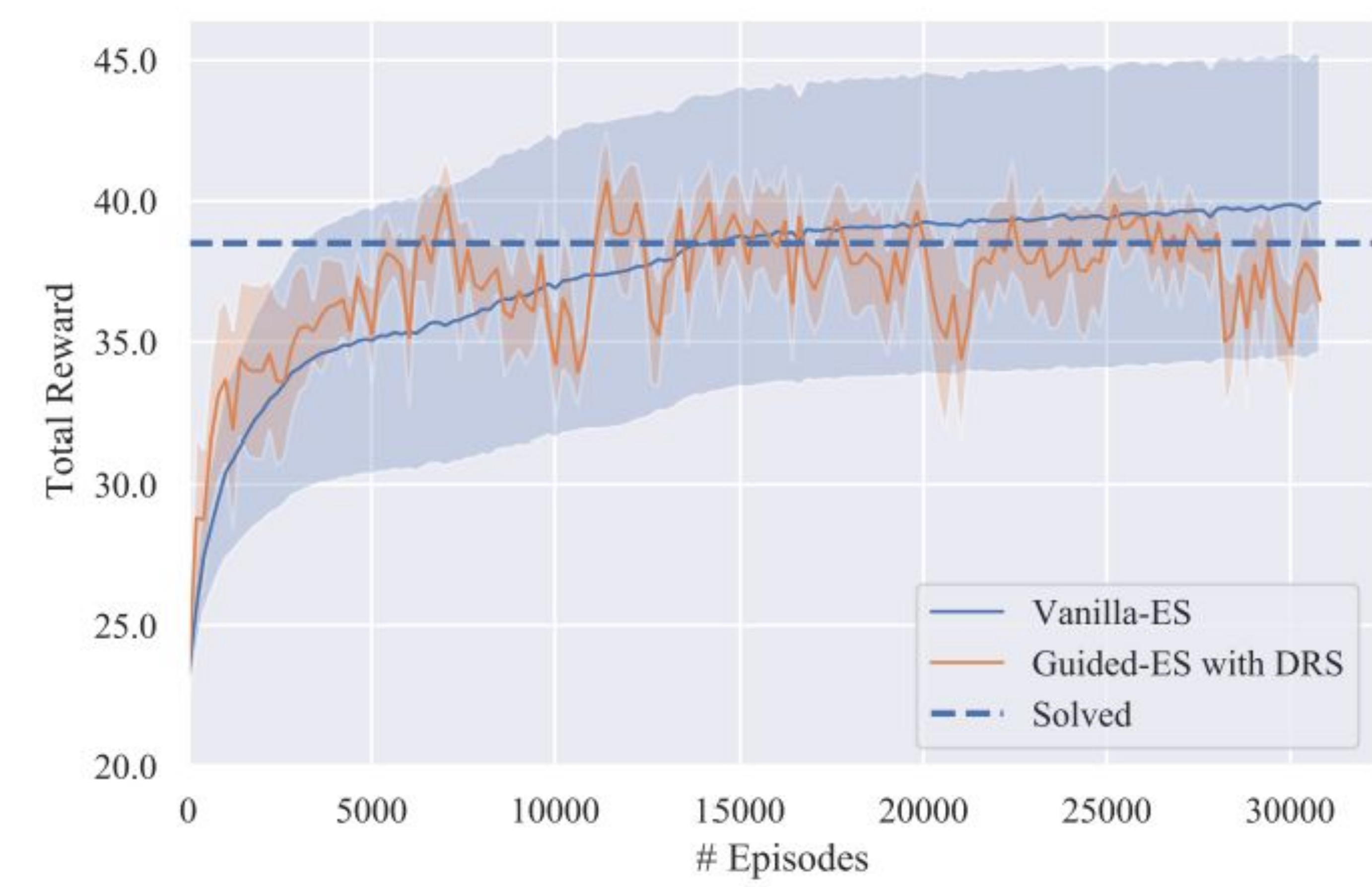
Guided-ES for Mass-Spring without Time-of-Impact fix

To demonstrate that the **convergence of Guided-ES with DRS is possible even when the gradients are not useful** for first-order optimization methods, we rely on a Mass-Spring simulator.

Hu et al. observed that a **naive DRS implementation for first-order optimization does not lead to a satisfactory solution**, and requires a specific approach to collision handling (time-of-impact fix). In this setup, we evaluate Guided-ES without the fix proposed by the DiffTaichi framework.

Our approach is able to find a satisfactory solution even when guided by the ineffective gradient with the faster convergence on average than the vanilla evolutionary strategies

The results suggest that if one has access to a differentiable robot simulator with inaccurate backward propagation, one could still utilize it. As there is a possibility it can still be used to reduce sample complexity of evolutionary strategies, which are easier to tune than reinforcement learning algorithms.



References

Niru Maheswaranathan et al. (2019). “Guided evolutionary strategies: augmenting random search with surrogate gradients.”. In: Proceedings of the 36th International Conference on Machine Learning, PMLR 97:4264-4273.

Hu et al. (2020). “DiffTaichi: Differentiable Programming for Physical Simulation.”. In: ICLR 2020.

H. Mania et al. (2018). “Simple random search of static linear policies is competitive for reinforcement learning.”. In: Advances in Neural Information Processing Systems 31.

T. Salimans et al. (2017). “Evolution Strategies as a Scalable Alternative to Reinforcement Learning.”. In: arXiv:1703.03864 [cs, stat].

N. Hansen et al. (2003). “Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES).”. In: Evolutionary Computation, 11(1):1–18.