

INTRIGUING PROPERTIES OF DEEP NEURAL POLICY MANIFOLD: INTRINSIC CORRELATION AND DEEP NEURAL POLICY CURVATURE

Anonymous authors

Paper under double-blind review

A APPENDIX

A.1 DIRECTIONS OF INSTABILITY

We provided the Proposition A.1 in Section 3 in the main body of the paper. Here we provide the full proof.

Proposition A.1 (*Directions of Instability*). *Let $\pi_{\mathcal{Q}}(s, a) = \frac{\exp \mathcal{Q}(s, a)}{\sum_{a' \in \mathcal{A}} \exp \mathcal{Q}(s, a')}$ be the softmax policy defined by \mathcal{Q} . Let $a^*(\hat{s}) = \arg \max_{a \in \mathcal{A}} \mathcal{Q}(\hat{s}, a)$. Then the gradient of the cross-entropy loss $\mathcal{J}(s, \hat{s})$ is given by*

$$\nabla_s \mathcal{J}(s, \hat{s})|_{s=\hat{s}} = \mathbb{E}_{a \sim \pi_{\mathcal{Q}}(\hat{s}, a)} [\nabla_s \mathcal{Q}(s, a)|_{s=\hat{s}}] - \nabla_s \mathcal{Q}(s, a^*(\hat{s})|_{s=\hat{s}}]$$

Proof. Since the directions of instability loss is

$$\mathcal{J}(s, \hat{s}) = -\log \pi_{\mathcal{Q}}(s, \arg \max_{a \in \mathcal{A}} \mathcal{Q}(\hat{s}, a))$$

The gradient of the directions of instability with respect to states is

$$\nabla_s \mathcal{J}(s, \hat{s})|_{s=\hat{s}} = -\nabla_s [\log \pi_{\mathcal{Q}}(s, \arg \max_{a \in \mathcal{A}} \mathcal{Q}(\hat{s}, a))]|_{s=\hat{s}}$$

Since the softmax policy is $\pi_{\mathcal{Q}}(s, a) = \frac{\exp \mathcal{Q}(s, a)}{\sum_{a' \in \mathcal{A}} \exp \mathcal{Q}(s, a')}$.

$$\begin{aligned} \nabla_s \mathcal{J}(s, \hat{s})|_{s=\hat{s}} &= -\nabla_s [\log \frac{\exp \mathcal{Q}(s, \arg \max_{a \in \mathcal{A}} \mathcal{Q}(\hat{s}, a))}{\sum_{a' \in \mathcal{A}} \exp \mathcal{Q}(s, a')}]|_{s=\hat{s}} \\ \nabla_s \mathcal{J}(s, \hat{s})|_{s=\hat{s}} &= \nabla_s [\log \sum_{a' \in \mathcal{A}} e^{\mathcal{Q}(s, a')} - \mathcal{Q}(s, \arg \max_{a \in \mathcal{A}} \mathcal{Q}(\hat{s}, a))]|_{s=\hat{s}} \end{aligned}$$

By the chain rule,

$$\nabla_s \mathcal{J}(s, \hat{s})|_{s=\hat{s}} = \frac{1}{\sum_{a' \in \mathcal{A}} e^{\mathcal{Q}(s, a')}} \sum_{a' \in \mathcal{A}} \nabla_s e^{\mathcal{Q}(s, a')} - \nabla_s \mathcal{Q}(s, \arg \max_{a \in \mathcal{A}} \mathcal{Q}(\hat{s}, a))|_{s=\hat{s}}$$

Hence,

$$\nabla_s \mathcal{J}(s, \hat{s})|_{s=\hat{s}} = \mathbb{E}_{a \sim \pi_{\mathcal{Q}}(\hat{s}, a)} [\nabla_s \mathcal{Q}(s, a)|_{s=\hat{s}}] - \nabla_s \mathcal{Q}(s, \arg \max_{a \in \mathcal{A}} \mathcal{Q}(\hat{s}, a))|_{s=\hat{s}}$$

Note that $a^*(\hat{s}) = \arg \max_{a \in \mathcal{A}} \mathcal{Q}(\hat{s}, a)$. Thus,

$$\nabla_s \mathcal{J}(s, \hat{s})|_{s=\hat{s}} = \mathbb{E}_{a \sim \pi_{\mathcal{Q}}(\hat{s}, a)} [\nabla_s \mathcal{Q}(s, a)|_{s=\hat{s}}] - \nabla_s \mathcal{Q}(s, a^*(\hat{s})|_{s=\hat{s}}]$$

□

A.2 STRICT CONVEXITY OF NEGATIVE AVERAGE ADVANTAGE

Theorem A.2 (*Strict Convexity of Negative Average Advantage*). *Let S^π be a finite sequence of states by the greedy policy π with respect to state-action value function Q in an MDP M . Assume that the state space \mathcal{S} has dimension $n \geq |S^\pi|$, and that $\Omega(s, \hat{s})$ is a twice continuously differentiable function of s and strictly convex on the simplex $\text{conv}(S^\pi)$. Then Ω can be extended to a strictly convex function on the entire state space.*

Proof. Observe that since S^π is a finite set of cardinality at most n , we have that $\text{conv}(S^\pi)$ is compact. Since Ω is twice continuously differentiable, strict convexity of Ω on $\text{conv}(S^\pi)$ implies that the Hessian is positive definite. Therefore, by Theorem 3.6 in the main body of the paper Ω can be extended to a twice continuously differentiable function with positive definite Hessian on the entire state space. Positive definiteness of the Hessian then implies that the extension of Ω to the entire state-space is strictly convex. \square

A.3 STATISTICAL MEASURE

Pearson. Pearson correlation is the ratio between the covariance of two variables, i.e. shared variance, and the product of their standard deviations, i.e. their individual spreads. Pearson correlation provides a normalized version of the covariance that is always between 1 and -1. Pearson correlation intends to capture a linear correlation between two variables (Pearson, 1895).

$$\rho_{X,Y} = \frac{\mathbb{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

Spearman. Spearman correlation measures the monotonic relationship between the two variables. The Spearman correlation coefficient is essentially the Pearson correlation coefficient computed not on the original variables, but on the ranked version of the of the variables. Let $R[X_i]$ and $R[Y_i]$ be the ranked conversion of X_i and Y_i consisting of n pairs (Spearman, 1904).

$$\rho_{R[X],R[Y]} = \frac{\mathbb{E}[(R[X] - \mu_{R[X]})(R[Y] - \mu_{R[Y]})]}{\sigma_{R[X]} \sigma_{R[Y]}}$$

Kendall's Tau. Kendall's τ correlation coefficient provides a measurement of rank correlation. In particular, (X_i, Y_i) and (X_j, Y_j) are called concordant if they preserve the order with respect to each other, i.e. if $X_j > X_i$ then $Y_j > Y_i$ or $X_j < X_i$ then $Y_j < Y_i$, and called discordant if they are reverse in order. Let C be the number of concordant pairs, let D be the number of discordant pairs, X_{tied} be the number of pairs tied only on the X variable, Y_{tied} be the number of variables tied on the Y variable, and XY_{tied} is the number of pairs tied on both of the variables (Kendall, 1938).

$$\tau_b = \frac{C - D}{\sqrt{(C + D + X_{\text{tied}})(C + D + Y_{\text{tied}})}}$$

The Kendall tau for measuring order association between variables X and Y is given by the following formula:

$$\tau = \frac{2}{n(n-1)} \sum_{i < j} \text{sgn}(X_i - X_j) \text{sgn}(Y_i - Y_j)$$

A.4 ALGORITHMIC DESCRIPTIONS

Dueling. The dueling architecture proposes a new neural network architecture for deep reinforcement learning by particularly having two separate estimators for the state value function and for the state-dependent action advantage function (Wang et al., 2016).

$$\mathcal{Q}(s, a; \theta, \alpha, \beta) = \mathcal{V}(s; \theta, \beta) + (\mathcal{A}(s, a; \theta, \alpha) - \max_{a' \in A} \mathcal{A}(s, a'; \theta, \alpha))$$

where the parameters of the convolutional layers are represented with θ and the parameters of the two streams of fully-connected layers represented with α and β .

Perceptual Similarity. Perceptual similarity is a metric introduced to quantify similarity between visual inputs beyond the ℓ_p -norm metrics (Zhang et al., 2018) by leveraging the internal activations of the neural networks Krizhevsky et al. (2012); Simonyan and Zisserman (2015); Iandola et al. (2016). In particular, $\Phi_{\text{similarity}}(s, \hat{s})$ is computed via

$$\Phi_{\text{similarity}}(s, \hat{s}) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (\hat{y}_{shw}^l - \hat{y}_{\xi(s)hw}^l)\|_2^2$$

where W_l , H_l , and C_l is the width, height and number of channels in the convolutional layers respectively and $\hat{y}_s^l, \hat{y}_{\Psi(s)}^l \in \mathbb{R}^{W_l \times H_l \times C_l}$ is the vector of the unit normalized activations. The perceptual similarity metric has recently been used in deep reinforcement learning to demonstrate the vulnerabilities and issues of adversarial training methods in reinforcement learning Korkmaz (2023; 2024).

DCT. Discrete Cosine Transform is the most canonical widely used signal processing applied to almost any digital visual input (Chen et al., 1977). Hence, any AI system that operates with visual inputs will be subjected to DCT artifacts introduced to their visual processing that can naturally happen in a given environment. In particular, the DCT transform is,

$$S_k = \sum_{n=0}^{N-1} s_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right)k\right] \text{ for } k = 1, \dots, N-1$$

where DCT transforms s_1, s_2, \dots, s_N real numbers to S_1, S_2, \dots, S_N real numbers. These artifacts are shown to be naturally occurring while perceptually invisible as much as an adversarial perturbation would be and cause significant damage in deep reinforcement learning policy performance Korkmaz (2023; 2024).

A.5 COMPUTATION OF ADVERSARIAL DIRECTIONS.

Carlini and Wagner. Adversarial directions find the minimum distance to a decision boundary where the optimal action cannot be taken by the reinforcement learning agent anymore. Originally proposed in the image classification domain by Carlini and Wagner (2017), and the formulation is later used to break the robust training defenses in deep reinforcement learning Ezgi (2022).

$$\min_{\hat{s} \in \mathcal{D}_{p,\epsilon}} \|\hat{s} - s\|_p \tag{1}$$

$$\text{subject to } \arg \max_a \mathcal{Q}(s, a) \neq \arg \max_a \mathcal{Q}(\hat{s}, a) \tag{2}$$

Elastic-Net. Further algorithms proposed to use the regularized Carlini and Wagner (2017) formulation which is regularized by the ℓ_1 -norm of the distance from the original state to the adversarial state observation Chen et al. (2018). Several other methods focused on optimization techniques to produce optimal adversarial directions.

Iterative Methods. In particular, Dong et al. (2018) proposed to use momentum iteration on top of the current iterative version Kurakin et al. (2016) of fast gradient sign method initially proposed by Goodfellow et al. (2015).

$$x_{\text{adv}}^{K+1} = \text{clip}_{\epsilon}(x_{\text{adv}}^K + \alpha \text{sign}(\nabla_x J(x_{\text{adv}}^K, y)))$$

Momentum. Particularly, Dong et al. (2018) took the iterative form of fast gradient sign method above and proposed to use Momentum iteration,

$$x_{\text{adv}}^{K+1} = x_{\text{adv}}^K + \alpha \frac{g_{K+1}}{g_K}$$

where the accumulated gradient g_{K+1} is,

$$g_{K+1} = \mu \cdot g_K + \frac{J(x_{\text{adv}}^K, y)}{\|\nabla_x J(x_{\text{adv}}^K, y)\|_1}$$

Nesterov Momentum. Later methods proposed Nesterov momentum in deep reinforcement learning,

$$v_{t+1} = \mu \cdot v_t + \frac{\nabla_{s_{\text{adv}}} J(s_{\text{adv}}^t + \mu \cdot v_t, a)}{\|\nabla_{s_{\text{adv}}} J(s_{\text{adv}}^t + \mu \cdot v_t, a)\|_1} \text{ and } s_{\text{adv}}^{t+1} = s_{\text{adv}}^t + \alpha \cdot \frac{v_{t+1}}{\|v_{t+1}\|_2}$$

where v_t is velocity and μ is the decay factor.

REFERENCES

- N. Carlini and D. Wagner. Towards evaluating the robustness of neural networks. *In 2017 IEEE Symposium on Security and Privacy (SP)*, pages 39–57, 2017.
- P. Chen, Y. Sharma, H. Zhang, J. Yi, and C. Hsieh. EAD: elastic-net attacks to deep neural networks via adversarial examples. In S. A. McIlraith and K. Q. Weinberger, editors, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 10–17. AAAI Press, 2018.
- W.-H. Chen, C. M. Smith, and S. Fralick. A fast computational algorithm for the discrete cosine transform. *IEEE Transactions on Communications*, 1977.
- Y. Dong, F. Liao, T. Pang, H. Su, J. Zhu, X. Hu, and J. Li. Boosting adversarial attacks with momentum. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9185–9193, 2018.
- K. Ezgi. Deep reinforcement learning policies learn shared adversarial features across mdps. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 7229–7238. AAAI Press, 2022.
- I. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples. *International Conference on Learning Representations*, 2015.
- F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.
- M. Kendall. A new measure of rank correlation. *Biometrika, JSTOR*, 30:81–89, 1938.
- E. Korkmaz. Adversarial robust deep reinforcement learning requires redefining robustness. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 8369–8377. AAAI Press, 2023.
- E. Korkmaz. Diagnosing and understanding deep reinforcement learning decision making. In *Proceedings of the 41th International Conference on Machine Learning, ICML 2024, 2024*.
- A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 2012.

- A. Kurakin, I. Goodfellow, and S. Bengio. Adversarial examples in the physical world. *arXiv preprint arXiv:1607.02533*, 2016.
- K. Pearson. Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58:240–242, 1895.
- K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations, ICLR*, 2015.
- C. Spearman. The proof and measurement of association between two things. *The American Journal of Psychology, JSTOR*, 58:72–101, 1904.
- Z. Wang, T. Schaul, M. Hessel, H. Van Hasselt, M. Lanctot, and N. De Freitas. Dueling network architectures for deep reinforcement learning. *International Conference on Machine Learning ICML*, page 1995–2003, 2016.
- R. Zhang, P. Isola, A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.