On the Identifiability of Latent Action Policies

Anonymous Author(s)
Affiliation
Address
email

Abstract

We study the identifiability of *latent action policy learning* (LAPO), a framework introduced recently to discover representations of actions from video data. We formally describe desiderata for such representations, their statistical benefits and potential sources of unidentifiability. Finally, we prove that an entropy-regularized LAPO objective identifies action representations satisfying our desiderata, under suitable conditions. Our analysis partly explains why *discrete* action representations are crucial in practice.

1 Introduction & background

In robot control, *behavior cloning* is an approach to learn a policy from action-labeled expert trajectories [12, 5]. It simply consists in training a policy $\pi(a \mid \boldsymbol{x})$ via supervised learning on data of the form $\{(\boldsymbol{x}_i, a_i)\}_{i=1}^N$ where a_i is the action taken by the expert policy in state $\boldsymbol{x}_i$. The success of the approach relies on at least two aspects: (i) the demonstrations $(\boldsymbol{x}_i, a_i)$ have to be generated from a sufficiently good expert policy, and (ii) the number of demonstrations N has to be sufficiently large to make learning possible. Although conceptually simple, the framework requires a large amount of action-labeled expert demonstrations to be successful, which can be costly to acquire.

To address this issue, Schmidt and Jiang [13] introduced *latent action policy learning* (LAPO) which can leverage large corpuses of unannotated video data in order to reduce reliance on action-labeled expert trajectories. LAPO proceeds in three stages. **First**, given a large dataset of state/next-state pairs $\{(\boldsymbol{x}_i, \boldsymbol{x}'_i)\}_{i=1}^N$, LAPO minimizes the reconstruction loss

$$\tfrac{1}{N} \textstyle\sum_{i=1}^N \mathbb{E}_{\hat{a} \sim \hat{q}(\hat{a} \mid \boldsymbol{x}_i, \boldsymbol{x}'_i)} \|\boldsymbol{x}'_i - \hat{\boldsymbol{g}}(\boldsymbol{x}_i, \hat{a})\|_2^2 \,,$$

where the *inverse dynamics model* (IDM) $\hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}')$ encodes pairs $(\boldsymbol{x}, \boldsymbol{x}')$ into action representations $\hat{a}$ which are then decoded using a *forward dynamics model* (FDM) $\hat{\boldsymbol{x}}' = \hat{\boldsymbol{g}}(\boldsymbol{x}, \hat{a})$. **Secondly**, the IDM is used to label the unlabeled video dataset, which yields $\{(\boldsymbol{x}_i, \hat{a}_i, \boldsymbol{x}'_i)\}_{i=1}^N$ where $\hat{a}_i := \arg\max_{\hat{a}} \hat{q}(\hat{a} \mid \boldsymbol{x}_i, \boldsymbol{x}'_i)$. This dataset is then used to train the *latent action policy* $\hat{\pi}(\hat{a} \mid \boldsymbol{x})$. **Thirdly**, a learnable head is applied on top of $\hat{\pi}(\hat{a} \mid \boldsymbol{x})$ and trained to map the latent actions $\hat{a}$ to actual actions a using a much smaller domain-specific action-labeled dataset $\{(\boldsymbol{x}_i, a_i)\}_{i=1}^{N_a}$ ($N_a << N$). While doing that, one can choose to either freeze the latent action policy $\hat{\pi}$ or fine-tune it. The authors showed that, thanks to this approach, fewer action-labeled samples are needed to train a good policy.[1] Since then, this idea has been applied at larger scale in Genie [2] and augmented with textual goal-conditioning in LAPA [15].

Motivated by the growing importance of this framework, we propose to study the identifiability of LAPO. Although identifiability in representation learning has been the subject of recent research efforts [6, 8, 14, 9, 3, 16], to the best of our knowledge, identifiability in the context of latent action modeling has not been investigated.

[1]The authors also show $\hat{\pi}$ can be fine-tuned via reinforcement learning more efficiently.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

Contributions. First, we postulate a data-generating process for the expert transitions $(\boldsymbol{x}, a, \boldsymbol{x}')$ (Section 2.1). Next, we provide formal desiderata for the IDM $\hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}')$ to be useful and discuss statistical consequences (Section 2.2). We further discuss two potential sources of unidentifiability (Section 2.4) and, finally, we prove that an entropy-regularized LAPO objective (Section 2.3) is guaranteed to identify an IDM satisfying the said desiderata, under suitable conditions (Section 2.5).

2 Identifiability analysis of LAPO

2.1 Data-generating process

Let $\boldsymbol{x} \in \mathcal{X} := [0,1]^d$ be the current observation, $\boldsymbol{x}' \in \mathcal{X}' := [0,1]^{d'}$ be the future observation and $a \in \mathcal{A} := \{1, \dots, k\}$ be a discrete action (in line with practical implementations [13, 2, 15]). The most natural situation is when $\boldsymbol{x}$ and $\boldsymbol{x}'$ corresponds to two consecutive frames, i.e. $\boldsymbol{x} := \boldsymbol{x}^t$ and $\boldsymbol{x}' = \boldsymbol{x}^{t+1}$ and $a = a^t$. But one could also consider different situations where the model is conditioned on a window of past observations: $\boldsymbol{x} = \boldsymbol{x}^{t-k:t}$ with $a = a^{t-k:t}$. Similarly, the $\boldsymbol{x}'$ could correspond to a window of multiple frames in the future.

We assume the current state $\boldsymbol{x} \in \mathcal{X}$ is sampled from some (Lebesgue) density function $p(\boldsymbol{x})$. Furthermore, an action $a \in \mathcal{A}$ is chosen according to a ground-truth policy π conditioned on $\boldsymbol{x}$:

$$\boldsymbol{x} \sim p(\boldsymbol{x}) \qquad \text{and} \qquad a \sim \pi(a \mid \boldsymbol{x})\,.$$

Define $p(\boldsymbol{x}, a) := p(\boldsymbol{x})\pi(a \mid \boldsymbol{x})$, $p(a) := \int p(\boldsymbol{x}, a) d\boldsymbol{x}$, and $p(\boldsymbol{x} \mid a) := p(\boldsymbol{x}, a)/p(a)$.

We assume the future observation $\boldsymbol{x}'$ is given by a deterministic transition model

$$\boldsymbol{x}' = \boldsymbol{g}(\boldsymbol{x}, a), \text{ where } \boldsymbol{g} : \mathcal{X} \times \mathcal{A} \to \mathcal{X}'\,.$$

This process induces a joint probability distribution over $(\boldsymbol{x}, \boldsymbol{x}')$, which we denote by $p(\boldsymbol{x}, \boldsymbol{x}')$.[2]

Support notation. In what follows, the support of $p(\boldsymbol{x} \mid a)$ is defined as

$$\text{supp}[p(\boldsymbol{x} \mid a)] := \{\boldsymbol{x}_0 \in \mathcal{X} \mid \forall \text{ open neighborhood } U \text{ of } \boldsymbol{x}_0, \textstyle\int_U p(\boldsymbol{x} \mid a) d\boldsymbol{x} > 0\}\,,$$

where $\int d\boldsymbol{x}$ denotes the Lebesgue integral. Note that $\text{supp}[p(\boldsymbol{x} \mid a)]$ might depend on a. Define also $\text{supp}[p(a)] := \{a \in \mathcal{A} \mid p(a) > 0\}$ and $\text{supp}[p(\boldsymbol{x}, a)] := \bigcup_{a \in \text{supp}[p(a)]} \text{supp}[p(\boldsymbol{x} \mid a)] \times \{a\}$.

2.2 Desiderata

In this section, we formalize three desiderata for an action representation and discuss statistical efficiency. Intuitively, we want to learn an encoder $\hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}')$ that captures useful information about the ground-truth action a. To formalize this, we will study

$$\boldsymbol{v}(\hat{a} \mid \boldsymbol{x}, a) := \hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a))\,,$$

defined for all $(\boldsymbol{x}, a) \in \text{supp}[p(\boldsymbol{x}, a)]$. The conditional probability mass function $\boldsymbol{v}(\hat{a} \mid \boldsymbol{x}, a)$ maps pairs $(\boldsymbol{x}, a)$ to their corresponding learned action representations $\hat{a}$, potentially in a stochastic way. This is effectively an *entanglement map*, as studied in identifiable representation learning [10]. Our first desideratum is to have a deterministic map from a to $\hat{a}$.

Desideratum 1 (Determinism)**.** *There exists a function $\boldsymbol{v}(\boldsymbol{x}, a)$ such that $\boldsymbol{v}(\hat{a} \mid \boldsymbol{x}, a) = \mathbf{1}(\hat{a} = \boldsymbol{v}(\boldsymbol{x}, a))$ for all $(\boldsymbol{x}, a) \in \text{supp}[p(\boldsymbol{x}, a)]$, where $\mathbf{1}(\cdot)$ is the indicator function.*

Notice how the learned action representation $\hat{a}$ might depend on the current state $\boldsymbol{x}$ via $\hat{a} = \boldsymbol{v}(\boldsymbol{x}, a)$. Such a dependence is undesirable since it signifies that the meaning of $\hat{a}$, i.e. how it relates to a, depends on the current state $\boldsymbol{x}$. We illustrate this unfortunate state of affairs with a simple example.

Example 1. *Consider a manipulation task where $\mathcal{A} := \{\texttt{left}, \texttt{right}\}$ and suppose*

$$\begin{aligned} 1 = \boldsymbol{v}(\boldsymbol{x}_0, a = \texttt{left}), \quad 2 = \boldsymbol{v}(\boldsymbol{x}_0, a = \texttt{right}),\\ 2 = \boldsymbol{v}(\boldsymbol{x}_1, a = \texttt{left}), \quad 1 = \boldsymbol{v}(\boldsymbol{x}_1, a = \texttt{right})\,. \end{aligned}$$

We can see that the meaning of $\hat{a}$ depends on the context $\boldsymbol{x}$: When in state $\boldsymbol{x}_0$, $\hat{a} = 1$ corresponds to $a = \texttt{left}$, whereas in state $\boldsymbol{x}_1$, $\hat{a} = 1$ corresponds to $a = \texttt{right}$.

[2] $p(\boldsymbol{x}, \boldsymbol{x}')$ is an abuse of notation since the distribution of $(\boldsymbol{x}, \boldsymbol{x}')$ has no a density (w.r.t. Lebesgue).

This undesirable phenomenon, described informally by Schmidt and Jiang [13, Section 6.2], can be thought of as a form of entanglement since $\hat{a}$ entangles both a and $\boldsymbol{x}$. This motivates:

Desideratum 2 (Disentanglement)**.** *There exists a function* $\boldsymbol{v}(a)$ *such that, for all* $(\boldsymbol{x}, a) \in \text{supp}[p(\boldsymbol{x}, a)]$, $\boldsymbol{v}(\boldsymbol{x}, a) = \boldsymbol{v}(a)$.

Furthermore, we want the latent action $\hat{a}$ to reveal all there is to know about the ground-truth action a. More formally, we want that two distinct actions a_1 and a_2 never map to the same latent action $\hat{a}$:

Desideratum 3 (Informativeness)**.** *The function* $\boldsymbol{v} : \text{supp}[p(a)] \to \hat{\mathcal{A}}$ *is injective.*

Statistical efficiency. As explained in Section 1, an encoder/IDM $\hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}')$ can be used to label the action-free video dataset, yielding $\{(\boldsymbol{x}_i, \hat{a}_i, \boldsymbol{x}'_i)\}_{i=1}^N$. If the IDM satisfies our desiderata, the newly labeled dataset is actually $\{(\boldsymbol{x}_i, \boldsymbol{v}(a_i), \boldsymbol{x}'_i)\}_{i=1}^N$ where a_i is the action taken by the expert policy $\pi(a \mid \boldsymbol{x}_i)$. Thus the latent action policy $\hat{\pi}(\hat{a} \mid \boldsymbol{x})$ trained on this data will approximate the distribution of $\boldsymbol{v}(a)$ when $a \sim \pi(a \mid \boldsymbol{x})$. This means that there exists a transformation $\sigma : \hat{\mathcal{A}} \to \mathcal{A}$ (any extension of $\boldsymbol{v}^{-1} : \boldsymbol{v}(\text{supp}[p(a)]) \to \mathcal{A}$) such that $\sigma(\hat{a}) \sim \pi(a \mid \boldsymbol{x})$ when $\hat{a} \sim \pi(\hat{a} \mid \boldsymbol{x})$. Hence, to get the expert policy π, we only need to learn $\sigma : \hat{\mathcal{A}} \to \mathcal{A}$ "on top of" the latent action policy $\hat{\pi}$ using the smaller action-labeled dataset. Had $\hat{a}$ been dependent on $\boldsymbol{x}$, such a transformation σ would not exist, forcing us to resort to either fine-tuning $\hat{\pi}$ or learning a map $\mathcal{X} \times \hat{\mathcal{A}} \to \mathcal{A}$ on top of $\hat{\pi}$, both of which are expected to be less statistically efficient than learning the simpler function $\sigma : \hat{\mathcal{A}} \to \mathcal{A}$.

2.3 A formal entropy-regularized LAPO objective

We now present a formal entropy-regularized LAPO objective. Theorem 1 will show that, under suitable assumptions, its solutions must satisfy the desiderata of Section 2.2.

In order to learn a deterministic encoder, we add an entropy regularizer $H(\hat{q}(\cdot \mid \boldsymbol{x}, \boldsymbol{x}')) := -\mathbb{E}_{\hat{q}(\hat{a}|\boldsymbol{x},\boldsymbol{x}')} \log \hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}')$. In the limit of infinite data, our entropy-regularized LAPO objective is

$$\min_{\hat{\boldsymbol{g}} \in \mathcal{G}, \hat{q} \in \mathcal{Q}} \mathbb{E}_{p(\boldsymbol{x},\boldsymbol{x}')} \left[\mathbb{E}_{\hat{q}(\hat{a}|\boldsymbol{x},\boldsymbol{x}')} \|\boldsymbol{x}' - \hat{\boldsymbol{g}}(\boldsymbol{x}, \hat{a})\|_2^2 + \beta H(\hat{q}(\cdot \mid \boldsymbol{x}, \boldsymbol{x}')) \right] , \tag{1}$$

where $\beta > 0$ controls regularization, and $\mathcal{G}$ and $\mathcal{Q}$ are respectively the hypothesis spaces for $\hat{\boldsymbol{g}}$ and $\hat{q}$.

Definition 1 (FDM hypothesis space $\mathcal{G}$)**.** *A function* $\hat{\boldsymbol{g}} : \mathcal{X} \times \hat{\mathcal{A}} \to \mathcal{X}'$ *is in* $\mathcal{G}$ *if and only if, for all* $\hat{a} \in \hat{\mathcal{A}}$, $\hat{\boldsymbol{g}}(\boldsymbol{x}, \hat{a})$ *is continuous in* $\boldsymbol{x}$.

Definition 2 (IDM hypothesis space $\mathcal{Q}$)**.** *Let* $\hat{\mathcal{A}} := \{1, \dots, \hat{k}\}$ *be the space of action representations* $\hat{a}$. *A function* $\hat{q} : \hat{\mathcal{A}} \times \mathcal{X} \times \mathcal{X}' \to [0, 1]$ *is in* $\mathcal{Q}$ *if and only if (i) for all* $(\boldsymbol{x}, \boldsymbol{x}') \in \mathcal{X} \times \mathcal{X}'$, $\sum_{\hat{a} \in \hat{\mathcal{A}}} \hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}') = 1$, *and (ii) for all* $\hat{a} \in \hat{\mathcal{A}}$, $\hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}')$ *is continuous in* $(\boldsymbol{x}, \boldsymbol{x}')$.

Note that our identifiability guarantee, Theorem 1, does not assume $\hat{k} = k$, only $\hat{k} \geq k$.

One can easily see that both terms in Problem (1) are lower bounded by zero. Additionally, Proposition 4 (in appendix) shows that there exists $(\boldsymbol{g}^*, q^*) \in \mathcal{G} \times \mathcal{Q}$ such that both terms are equal to zero (under Assumptions 1 and 2). This means that, at optimality, the entropy regularizer must be equal to zero thus forcing the learned IDM $\hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}')$ to be deterministic for all $(\boldsymbol{x}, \boldsymbol{x}') \in \text{supp}[p(\boldsymbol{x}, \boldsymbol{x}')]$. In other words, at optimality, $\hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}') = \mathbf{1}(\hat{a} = \hat{\boldsymbol{f}}(\boldsymbol{x}, \boldsymbol{x}'))$ for some function $\hat{\boldsymbol{f}} : \text{supp}[p(\boldsymbol{x}, \boldsymbol{x}')] \to \hat{\mathcal{A}}$.

Remark 1. *The above development begs the question: Why are we considering a stochastic IDM* $\hat{q}$ *to later regularize it to be deterministic? A perhaps more natural route would be to directly train a deterministic encoder* $\hat{\boldsymbol{f}} : \mathcal{X} \times \mathcal{X}' \to \hat{\mathcal{A}}$. *From an optimization perspective, a stochastic encoder is helpful as it unlocks gradient computation via the reparameterization trick [7]. From a theoretical perspective, the continuity condition on* $\hat{q}$ *is crucial for our proof, as it excludes pathological encoders* $\hat{\boldsymbol{f}}$ *that would present "jumps" on the connected components of* $\text{supp}[p(\boldsymbol{x}, \boldsymbol{x}')]$. *We conjecture that the VQ-VAE approach of Schmidt and Jiang [13], which is limited to deterministic discrete encoders, can in principle lead to such pathological behaviors. We leave this for future work.*

2.4 Potential sources of unidentifiability

We now show that, without assumptions on the data-generating process or without restrictions on the hypothesis classes $\mathcal{Q}$ and $\mathcal{G}$, Problem (1) admits degenerate solutions which do not satisfy our desiderata of Section 2.2.

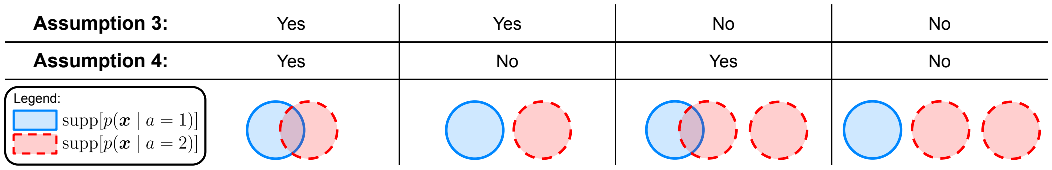

Figure 1: Illustration of Assumptions 3 and 4. Assume $\mathcal{A} := \{1, 2\}$

Example 2 (No restriction on $\hat{\mathcal{A}}$)**.** *In principle, one can choose $\hat{\mathcal{A}} := \mathcal{X}'$ and $\hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}') := \delta(\hat{a} - \boldsymbol{x}')$ where δ is the Dirac function. Hence the IDM outputs $\boldsymbol{x}'$ deterministically. By choosing $\hat{\boldsymbol{g}}(\boldsymbol{x}, \hat{a}) = \hat{a}$, we clearly solve Problem (1), but the action representation $\hat{a}$ is uninteresting. In fact, this can be understood as a violation of Desideratum 2 since $\hat{a} = \boldsymbol{x}'$ clearly depends on $\boldsymbol{x}$ via $\boldsymbol{x}' = \boldsymbol{g}(\boldsymbol{x}, a)$.*

Example 3 (Deterministic $\pi(a \mid \boldsymbol{x})$)**.** *Assume $\pi(a \mid \boldsymbol{x}) = \mathbf{1}(a = \pi(\boldsymbol{x}))$, i.e. the ground-truth policy is deterministic. In that case, one can solve the reconstruction problem by choosing $\hat{\boldsymbol{g}}(\boldsymbol{x}, \hat{a}) := \boldsymbol{g}(\boldsymbol{x}, \pi(\boldsymbol{x}))$ since $\boldsymbol{x}' = \boldsymbol{g}(\boldsymbol{x}, \pi(\boldsymbol{x}))$ with probability one. In that case, the latent action $\hat{a}$ is completely ignored by the FDM and thus the IDM $\hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}')$ could simply output the same action deterministically, which would clearly present a violation of Desideratum 3.*

2.5 Main identifiability result

In this section, we provide sufficient conditions on the data-generating process under which the solutions of Problem (1) are guaranteed to satisfy the desiderata of Section 2.2.

First of all, we require the ground-truth FDM to be continuous.

Assumption 1 (Continuous $\boldsymbol{g}$)**.** *For all $a \in \mathcal{A}$, the ground-truth FDM $\boldsymbol{g}(\boldsymbol{x}, a)$ is continuous in $\boldsymbol{x}$.*

Additionally, we require that different actions always have different effects in the data-generating process. We formalize this as a form of injectivity.

Assumption 2 (Injectivity)**.** *For all $\boldsymbol{x} \in \mathcal{X}$ and $a_1, a_2 \in \mathcal{A}$, if $a_1 \neq a_2$, then $\boldsymbol{g}(\boldsymbol{x}, a_1) \neq \boldsymbol{g}(\boldsymbol{x}, a_2)$.*

The last two assumptions put topological restrictions on the support of $p(\boldsymbol{x}, a)$. Recall that a set $S \subseteq \mathbb{R}^d$ is said to be *connected* if it "holds in one piece" [11]. See Figure 1 for an illustration.

Assumption 3. *For all $a \in \text{supp}[p(a)]$, we have that $\text{supp}[p(\boldsymbol{x} \mid a)]$ is a connected subset of $\mathcal{X}$.*

Assumption 4. *For all pairs $a_1, a_2 \in \text{supp}[p(a)]$, $\text{supp}[p(\boldsymbol{x} \mid a_1)] \cap \text{supp}[p(\boldsymbol{x} \mid a_2)] \neq \emptyset$.*

Note that Assumptions 3 and 4 are both satisfied for example if $\text{supp}[p(\boldsymbol{x}, a)] = \mathcal{X} \times \mathcal{A}$.

We are now ready to state the main identifiability result of this work. It shows that, under the assumptions introduced above, the encoder/IDM $\hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}')$ learned by optimizing Problem (1) must satisfy the desiderata of Section 2.2. Its proof can be found in the appendix. Recall $\boldsymbol{v}(\hat{a} \mid \boldsymbol{x}, a) := \hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a))$.

Theorem 1. *Suppose $\hat{k} \geq k$ and let $(\hat{\boldsymbol{g}}, \hat{q})$ be a solution[3] of Problem (1) with hypothesis classes $\mathcal{G}$ (Definition 1) and $\mathcal{Q}$ (Definition 2).*

1. *If Assumptions 1 and 2 hold, then Desideratum 1 holds, i.e. there exists a function $\boldsymbol{v} : \text{supp}[p(\boldsymbol{x}, a)] \to \hat{\mathcal{A}}$ such that, for all $(\boldsymbol{x}, a) \in \text{supp}[p(\boldsymbol{x}, a)]$,*

$$\boldsymbol{v}(\hat{a} \mid \boldsymbol{x}, a) = \mathbf{1}(\hat{a} = \boldsymbol{v}(\boldsymbol{x}, a)) \,.$$

2. *If Assumptions 1 to 3 hold, then Desideratum 2 holds, i.e. there exists a mapping $\boldsymbol{v} : \text{supp}[p(a)] \to \hat{\mathcal{A}}$ such that, for all $(\boldsymbol{x}, a) \in \text{supp}[p(\boldsymbol{x}, a)]$,*

$$\boldsymbol{v}(\hat{a} \mid \boldsymbol{x}, a) = \mathbf{1}(\hat{a} = \boldsymbol{v}(a)) \,.$$

3. *If Assumptions 1 to 4 hold, then Desideratum 3 holds, i.e. the mapping $\boldsymbol{v} : \text{supp}[p(a)] \to \hat{\mathcal{A}}$ defined above is injective.*

[3] Under Assumptions 1 and 2, a solution is guaranteed to exist by Proposition 4 in appendix.

References

[1] S. Axler. *Measure, Integration & Real Analysis*. Springer International Publishing, 2019.

[2] Jake Bruce, Michael D Dennis, Ashley Edwards, Jack Parker-Holder, Yuge Shi, Edward Hughes, Matthew Lai, Aditi Mavalankar, Richie Steigerwald, Chris Apps, Yusuf Aytar, Sarah Maria Elisabeth Bechtle, Feryal Behbahani, Stephanie C.Y. Chan, Nicolas Heess, Lucy Gonzalez, Simon Osindero, Sherjil Ozair, Scott Reed, Jingwei Zhang, Konrad Zolna, Jeff Clune, Nando de Freitas, Satinder Singh, and Tim Rocktäschel. Genie: Generative interactive environments. In *Forty-first International Conference on Machine Learning*, 2024.

[3] S. Buchholz, M. Besserve, and B. Schölkopf. Function classes for identifiable nonlinear independent component analysis. In *Advances in Neural Information Processing Systems*, 2022.

[4] Ryszard Engelking. *General Topology*. Sigma series in pure mathematics. Heldermann, 1989.

[5] Dylan J Foster, Adam Block, and Dipendra Misra. Is behavior cloning all you need? understanding horizon in imitation learning. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[6] Aapo Hyvärinen, Ilyes Khemakhem, and Hiroshi Morioka. Nonlinear independent component analysis for principled disentanglement in unsupervised deep learning. *Patterns*, 2023.

[7] E. Jang, S. Gu, and B. Poole. Categorical reparameterization with gumbel-softmax. *International Conference on Machine Learning*, 2017.

[8] I. Khemakhem, D. Kingma, R. Monti, and A. Hyvarinen. Variational autoencoders and nonlinear ica: A unifying framework. In *International Conference on Artificial Intelligence and Statistics*, 2020.

[9] S. Lachapelle, Rodriguez Lopez, P., Y. Sharma, K. E. Everett, R. Le Priol, A. Lacoste, and S. Lacoste-Julien. Disentanglement via mechanism sparsity regularization: A new principle for nonlinear ICA. In *Conference on Causal Learning and Reasoning*, 2022.

[10] Sébastien Lachapelle, Pau Rodríguez López, Yash Sharma, Katie Everett, Rémi Le Priol, Alexandre Lacoste, and Simon Lacoste-Julien. Nonparametric partial disentanglement via mechanism sparsity: Sparse actions, interventions and sparse temporal dependencies. *arXiv preprint arXiv:2401.04890*, 2024.

[11] J. R. Munkres. *Topology*. Prentice Hall, Inc., 2 edition, 2000.

[12] Dean A. Pomerleau. ALVINN: An autonomous land vehicle in a neural network. In *Advances in Neural Information Processing Systems*, 1988.

[13] D. Schmidt and M. Jiang. Learning to act without actions. In *The Twelfth International Conference on Learning Representations*, 2024.

[14] J. Von Kügelgen, Y. Sharma, L. Gresele, W. Brendel, B. Schölkopf, M. Besserve, and F. Locatello. Self-supervised learning with data augmentations provably isolates content from style. In *Advances in Neural Information Processing Systems*, 2021.

[15] S. Ye, J. Jang, B. Jeon, S. June Joo, J. Yang, B. Peng, A. Mandlekar, R. Tan, Y.-W. Chao, B. Y. Lin, L. Liden, K. Lee, J. Gao, L. Zettlemoyer, D. Fox, and M. Seo. Latent action pretraining from videos. In *The Thirteenth International Conference on Learning Representations*, 2025.

[16] J. Zhang, K. Greenewald, C. Squires, A. Srivastava, K. Shanmugam, and C. Uhler. Identifiability guarantees for causal disentanglement from soft interventions. In *Advances in Neural Information Processing Systems*, 2023.

Appendix

Multiple intermediary results are necessary in order to prove Theorem 1. The proof of the following technical lemma can be safely skipped at first read.

Lemma 2. *Let X be a metric space*[4] *and let $X_1, X_2, ..., X_K$ be a finite collection of disjoint closed sets of X. Then, there exists a function $q : [K] \times X \to [0, 1]$ such that*

- *for all $x \in X$, $\sum_{k\in[K]} q(k \mid x) = 1$,*
- *for all $k \in [K]$, $q(k \mid x)$ is continuous in x (as a function $X \to [0, 1]$), and*
- *for all $k \in [K]$ and all $x \in X_k$, $q(k \mid x) = 1$.*

Proof. We make use of the Vedenissoff theorem [4, Theorem 1.5.19]. We extract only the part of the theorem we will need: If a topological space X is perfectly normal, then for every pair of disjoint closed sets $A, B \subseteq X$, there exists a continuous function $f : X \to [0, 1]$ such that $f^{-1}(\{0\}) = A$ and $f^{-1}(\{1\}) = B$.

Since a metric space is always perfectly normal [4, Corollary 4.1.13], X is perfectly normal and thus we can apply the Vedenissoff theorem.

For each $k \in [K]$, the set $\bigcup_{k'\in[K]\setminus\{k\}} X_{k'}$ is closed since a finite union of closed sets is closed. By Vedenissoff theorem, there exists a continuous function $h_k : X \to [0, 1]$ such that $h_k^{-1}(\{1\}) = X_k$ and $h_k^{-1}(\{0\}) = \bigcup_{k'\in[K]\setminus\{k\}} X_{k'}$.

We now prove that $\sum_{k\in[K]} h_k(x) > 0$ for all $x \in X$. We consider two cases, $x \in \bigcup_{k\in[K]} X_k$ and $x \notin \bigcup_{k\in[K]} X_k$. If $x \in \bigcup_{k\in[K]} X_k$, then there exists a $k_0 \in [K]$ such that $x \in X_{k_0} = h_{k_0}^{-1}(\{1\})$ which means $h_{k_0}(x) = 1$. Of course, this implies that the sum is greater than zero. If $x \notin \bigcup_{k\in[K]} X_k$, in particular we have $x \notin \bigcup_{k\in[K]\setminus\{1\}} X_k = h_1^{-1}(\{0\})$, which means $h_1(x) > 0$. Of course, this implies the sum is greater than zero.

Since $\sum_{k\in[K]} h_k(x) > 0$ for all $x \in X$, we can define, for all $(k, x) \in [K] \times X$,

$$q(k \mid x) := \frac{h_k(x)}{\sum_{k'\in[K]} h_{k'}(x)} \,.$$

We now verify that q satisfies the three conditions of the theorem. First, it is clear that $\sum_{k\in[K]} q(k \mid x) = 1$. Second, $q(k \mid x)$ is continuous in x since all $h_{k'}$ are continuous functions.

Third, we check that $q(k \mid x) = 1$ when $x \in X_k$. Let $k \in [K]$ and $x \in X_k$. Since $X_k = h_k^{-1}(\{1\})$, we have that $h_k(x) = 1$. Consider $k' \in [K] \setminus \{k\}$. Clearly, $x \in X_k \subseteq \bigcup_{k''\in[K]\setminus\{k'\}} X_{k''} = h_{k'}^{-1}(\{0\})$, which means $h_{k'}(x) = 0$ for $k' \neq k$. We thus have

$$q(k \mid x) = \frac{h_k(x)}{h_k(x) + \sum_{k'\in[K]\setminus\{k\}} h_{k'}(x)} = \frac{1}{1 + 0} = 1 \,,$$

which concludes the proof. □

Define the function $G : \mathcal{X} \times \mathcal{A} \to \mathcal{X} \times \mathcal{X}'$ as $G(\boldsymbol{x}, a) := (\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a))$, which is simply the function $\boldsymbol{g}(\boldsymbol{x}, a)$ with a copy of $\boldsymbol{x}$ in its output. Note that its image $G(\mathcal{X} \times \mathcal{A})$ is effectively the set of plausible transition pairs $(\boldsymbol{x}, \boldsymbol{x}')$. In general, this is expected to be a proper subset of $\mathcal{X} \times \mathcal{X}'$.

Lemma 3. *Under Assumption 2, G is injective.*

Proof. If $G(\boldsymbol{x}_1, \boldsymbol{a}_1) = G(\boldsymbol{x}_2, \boldsymbol{a}_2)$, then $\boldsymbol{x}_1 = \boldsymbol{x}_2$ and $\boldsymbol{g}(\boldsymbol{x}_1, a_1) = \boldsymbol{g}(\boldsymbol{x}_1, a_2)$ and thus, by Assumption 2, $a_1 = a_2$. □

[4] The result generalizes to the case where X is a perfectly normal topological space.

Remark 2. *Since G is injective, it is bijective on its image $G(\mathcal{X} \times \mathcal{A}) \subseteq \mathcal{X} \times \mathcal{X}'$. Therefore, it has an inverse $F : G(\mathcal{X} \times \mathcal{A}) \to \mathcal{X} \times \mathcal{A}$ which clearly has the form $F(\boldsymbol{x}, \boldsymbol{x}') = (\boldsymbol{x}, \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{x}'))$, for some function $\boldsymbol{f} : G(\mathcal{X} \times \mathcal{A}) \to \mathcal{A}$.*

Proposition 4. *Suppose Assumptions 1 and 2 hold and $\hat{k} \geq k$. Then, there exist $\boldsymbol{g}^* \in \mathcal{G}$ and $q^* \in \mathcal{Q}$ such that the objective of Problem (1) is equal to zero.*

Proof. Take $\boldsymbol{g}^*(\boldsymbol{x}, \hat{a}) := \mathbf{1}(\hat{a} \in \mathcal{A})\boldsymbol{g}(\boldsymbol{x}, \hat{a})$. Essentially, $\boldsymbol{g}^*$ imitates the ground-truth FDM $\boldsymbol{g}$ when the action $\hat{a} \in \mathcal{A}$, otherwise it simply outputs zero. Clearly, $\boldsymbol{g}^* \in \mathcal{G}$ since, by Assumption 1, $\boldsymbol{g}(\cdot, \hat{a})$ is continuous for all $\hat{a} \in \mathcal{A}$ (and the zero function is continuous).

We now construct a $q^* \in \mathcal{Q}$ such that, for all $(\boldsymbol{x}, \boldsymbol{x}') \in G(\mathcal{X} \times \mathcal{A})$ and all $\hat{a} \in \hat{\mathcal{A}}$, we have $q^*(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}') = \delta(\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{x}') - \hat{a})$, where $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{x}')$ is defined in Remark 2. Later on, we show that the pair $(q^*, \boldsymbol{g}^*)$ sets the objective to zero.

Notice that $G(\mathcal{X} \times \mathcal{A}) = \bigcup_{a \in \mathcal{A}} G(\mathcal{X} \times \{a\})$ where $G(\mathcal{X} \times \{a\}) = \{(\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a)) \mid \boldsymbol{x} \in \mathcal{X}\}$ is the graph of $\boldsymbol{g}(\cdot, a)$. Since $\boldsymbol{g}(\cdot, a)$ is continuous by Assumption 1, the closed graph theorem implies that its graph, $G(\mathcal{X} \times \{a\})$, is closed is $\mathcal{X} \times \mathcal{X}'$. Furthermore, we know that the sets $G(\mathcal{X} \times \{a\})$ are mutually disjoint since otherwise there exists $(\boldsymbol{x}, \boldsymbol{x}') \in G(\mathcal{X} \times \{a_1\}) \cap G(\mathcal{X} \times \{a_2\})$ for distinct a_1, a_2 which implies $(\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a_1)) = (\boldsymbol{x}, \boldsymbol{x}') = (\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a_2))$, which violates Assumption 2.

To summarize, the last paragraph showed that $\{G(\mathcal{X} \times \{a\})\}_{a \in \mathcal{A}}$ is a partition of $G(\mathcal{X} \times \mathcal{A})$ where each $G(\mathcal{X} \times \{a\})$ is closed in $\mathcal{X} \times \mathcal{X}'$. By noticing that $\mathcal{X} \times \mathcal{X}'$ is a metric space, we can apply Lemma 2 to show the existence of a function $q : \mathcal{A} \times \mathcal{X} \times \mathcal{X}' \to [0, 1]$ such that the following holds:

- for all $(\boldsymbol{x}, \boldsymbol{x}') \in \mathcal{X} \times \mathcal{X}'$, $\sum_{a \in \mathcal{A}} q(a \mid \boldsymbol{x}, \boldsymbol{x}') = 1$,
- for all $a \in \mathcal{A}$, $q(a \mid \boldsymbol{x}, \boldsymbol{x}')$ is continuous in $(\boldsymbol{x}, \boldsymbol{x}')$, and
- for all $a \in \mathcal{A}$ and all $(\boldsymbol{x}, \boldsymbol{x}') \in G(\mathcal{X} \times \{a\})$, $q(a \mid \boldsymbol{x}, \boldsymbol{x}') = 1$.

We choose, for all $(\hat{a}, \boldsymbol{x}, \boldsymbol{x}') \in \hat{\mathcal{A}} \times \mathcal{X} \times \mathcal{X}'$, $q^*(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}') := \mathbf{1}(\hat{a} \in \mathcal{A})q(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}')$. In other words, q^* imitates q when $\hat{a} \in \mathcal{A}$, and simply outputs zero when $\hat{a} \notin \mathcal{A}$.

We now check that $q^* \in \mathcal{Q}$. Take $(\boldsymbol{x}, \boldsymbol{x}') \in \mathcal{X} \times \mathcal{X}'$. We have

$$\sum_{\hat{a} \in \hat{\mathcal{A}}} q^*(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}') = \sum_{\hat{a} \in \hat{\mathcal{A}}} \mathbf{1}(\hat{a} \in \mathcal{A})q(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}') = \sum_{\hat{a} \in \mathcal{A}} q(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}') = 1\,. \tag{2}$$

where the second equality used the fact that $\mathcal{A} \subseteq \hat{\mathcal{A}}$ (since $\hat{k} \geq k$).

Now take $\hat{a} \in \hat{\mathcal{A}}$. If $\hat{a} \in \mathcal{A}$, then $q^*(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}') = q(a \mid \boldsymbol{x}, \boldsymbol{x}')$ which is continuous in $(\boldsymbol{x}, \boldsymbol{x}')$. If $\hat{a} \notin \mathcal{A}$, then $q^*(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}') = 0$ which is also continuous. Thus $q^* \in \mathcal{Q}$.

Notice that, for all $(\boldsymbol{x}, a) \in \mathcal{X} \times \mathcal{A}$, we have

$$q^*(\hat{a} \mid \boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a)) = \mathbf{1}(\hat{a} \in \mathcal{A})q(\hat{a} \mid \boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a)) = \mathbf{1}(\hat{a} \in \mathcal{A})\mathbf{1}(\hat{a} = a) = \mathbf{1}(\hat{a} = a)\,, \tag{3}$$

where the third equality holds because $(\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a)) \in G(\mathcal{X} \times \{a\})$, which implies that $q(a \mid \boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a)) = 1$.

Now, we must show that the pair $(\boldsymbol{g}^*, q^*)$ sets the loss of Problem (1) to zero. First note that

$$\begin{aligned}
\mathbb{E}_{p(\boldsymbol{x}, \boldsymbol{x}')} H(q^*(\cdot \mid \boldsymbol{x}, \boldsymbol{x}')) &= \mathbb{E}_{p(\boldsymbol{x}, a)} H(q^*(\cdot \mid \boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a))) &(4)\\
&= \mathbb{E}_{p(\boldsymbol{x}, a)} H(\mathbf{1}(\cdot = a)) &(5)\\
&= \mathbb{E}_{p(\boldsymbol{x}, a)} 0 = 0\,. &(6)
\end{aligned}$$

Also,

$$
\begin{aligned}
&\mathbb{E}_{p(\boldsymbol{x},\boldsymbol{x}')} \sum_{\hat{a}\in\hat{\mathcal{A}}} q^*(\hat{a} \mid \boldsymbol{x},\boldsymbol{x}')\|\boldsymbol{x}' - \boldsymbol{g}^*(\boldsymbol{x},\hat{a})\|_2^2 && (7)\\
&\quad = \mathbb{E}_{p(\boldsymbol{x},\boldsymbol{x}')} \sum_{\hat{a}\in\hat{\mathcal{A}}} q^*(\hat{a} \mid \boldsymbol{x},\boldsymbol{x}')\|\boldsymbol{x}' - \mathbf{1}(\hat{a}\in\mathcal{A})\boldsymbol{g}(\boldsymbol{x},\hat{a})\|_2^2 && (8)\\
&\quad = \mathbb{E}_{p(\boldsymbol{x},a)} \sum_{\hat{a}\in\hat{\mathcal{A}}} q^*(\hat{a} \mid \boldsymbol{x},\boldsymbol{g}(\boldsymbol{x},a))\|\boldsymbol{g}(\boldsymbol{x},a) - \mathbf{1}(\hat{a}\in\mathcal{A})\boldsymbol{g}(\boldsymbol{x},\hat{a})\|_2^2 && (9)\\
&\quad = \mathbb{E}_{p(\boldsymbol{x},a)} \sum_{\hat{a}\in\hat{\mathcal{A}}} \mathbf{1}(\hat{a}=a)\|\boldsymbol{g}(\boldsymbol{x},a) - \mathbf{1}(\hat{a}\in\mathcal{A})\boldsymbol{g}(\boldsymbol{x},\hat{a})\|_2^2 && (10)\\
&\quad = \mathbb{E}_{p(\boldsymbol{x},a)} \|\boldsymbol{g}(\boldsymbol{x},a) - \mathbf{1}(a\in\mathcal{A})\boldsymbol{g}(\boldsymbol{x},a)\|_2^2 && (11)\\
&\quad = \mathbb{E}_{p(\boldsymbol{x},a)} \|\boldsymbol{g}(\boldsymbol{x},a) - \boldsymbol{g}(\boldsymbol{x},a)\|_2^2 && (12)\\
&\quad = \mathbb{E}_{p(\boldsymbol{x},a)} 0 = 0\,, && (13)
\end{aligned}
$$

where the fourth equality used the fact that $\mathcal{A} \subseteq \hat{\mathcal{A}}$, which holds since $\hat{k} \geq k$. This concludes the proof. ☐

The following lemma simply states that if the integral of a non-negative continuous function f w.r.t. to some measure μ is equal to zero, then the function must be zero on the support of μ.

Lemma 5. *Let (X, τ) be a topological space and let $\mathcal{F}$ be the Borel sigma-algebra for X. Let $\mu : \mathcal{F} \to [0, \infty)$ be a measure and let $f : X \to [0, \infty)$ be a non-negative continuous function. If $\int f d\mu = 0$, then $f(x) = 0$ for all $x \in \text{supp}[\mu]$.*

Proof. We show the contrapositive statement. Suppose there exists $x_0 \in \text{supp}[\mu]$ such that $f(x_0) > 0$. Since f is continuous, we have that $f^{-1}((0,\infty))$ is an open neighborhood of x_0. Since $x_0 \in \text{supp}[\mu]$, we have that $\mu(f^{-1}((0,\infty))) > 0$. But this means $\int f d\mu > 0$ [1, Section 3A, Exercise 3]. ☐

We are finally ready to prove Theorem 1.

Theorem 1. *Suppose $\hat{k} \geq k$ and let $(\hat{\boldsymbol{g}}, \hat{q})$ be a solution[5] of Problem (1) with hypothesis classes $\mathcal{G}$ (Definition 1) and $\mathcal{Q}$ (Definition 2).*

1. *If Assumptions 1 and 2 hold, then Desideratum 1 holds, i.e. there exists a function $\boldsymbol{v} : \text{supp}[p(\boldsymbol{x}, a)] \to \hat{\mathcal{A}}$ such that, for all $(\boldsymbol{x}, a) \in \text{supp}[p(\boldsymbol{x}, a)]$,*
$$\boldsymbol{v}(\hat{a} \mid \boldsymbol{x}, a) = \mathbf{1}(\hat{a} = \boldsymbol{v}(\boldsymbol{x}, a))\,.$$
2. *If Assumptions 1 to 3 hold, then Desideratum 2 holds, i.e. there exists a mapping $\boldsymbol{v} : \text{supp}[p(a)] \to \hat{\mathcal{A}}$ such that, for all $(\boldsymbol{x}, a) \in \text{supp}[p(\boldsymbol{x}, a)]$,*
$$\boldsymbol{v}(\hat{a} \mid \boldsymbol{x}, a) = \mathbf{1}(\hat{a} = \boldsymbol{v}(a))\,.$$
3. *If Assumptions 1 to 4 hold, then Desideratum 3 holds, i.e. the mapping $\boldsymbol{v} : \text{supp}[p(a)] \to \hat{\mathcal{A}}$ defined above is injective.*

Proof. If $(\hat{\boldsymbol{g}}, \hat{q})$ solves Problem (1), we must have $\hat{\boldsymbol{g}} \in \mathcal{G}$ and $\hat{q} \in \mathcal{Q}$. Moreover, since Assumptions 1 and 2 hold and $\hat{k} \geq k$, we can apply Proposition 4 to conclude that there exists a pair $(\boldsymbol{g}^*, q^*) \in \mathcal{G} \times \mathcal{Q}$ that reaches zero loss. From this, we conclude that $(\hat{\boldsymbol{g}}, \hat{q})$ must also reach zero loss, otherwise it is not optimal.

Thus we have

$$\mathbb{E}_{p(\boldsymbol{x},\boldsymbol{x}')} \left[\sum_{\hat{a}\in\hat{\mathcal{A}}} \hat{q}(\hat{a} \mid \boldsymbol{x},\boldsymbol{x}')\|\boldsymbol{x}' - \hat{\boldsymbol{g}}(\boldsymbol{x},\hat{a})\|_2^2 + H(\hat{q}(\cdot \mid \boldsymbol{x},\boldsymbol{x}')) \right] = 0\,. \tag{14}$$

[5] Under Assumptions 1 and 2, a solution is guaranteed to exist by Proposition 4 in appendix.

Since both terms are lower bounded by 0, both terms must equal zero. We start by using the fact that the entropy term is equal to zero:

$$\mathbb{E}_{p(\boldsymbol{x},\boldsymbol{x}')} H(\hat{q}(\cdot \mid \boldsymbol{x}, \boldsymbol{x}')) = 0 \tag{15}$$

$$\mathbb{E}_{p(\boldsymbol{x},a)} H(\hat{q}(\cdot \mid \boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a))) = 0 \tag{16}$$

$$\mathbb{E}_{p(a)}\mathbb{E}_{p(\boldsymbol{x}|a)} H(\hat{q}(\cdot \mid \boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a))) = 0 \tag{17}$$

$$\sum_{a \in \text{supp}[p(a)]} p(a)\mathbb{E}_{p(\boldsymbol{x}|a)} H(\hat{q}(\cdot \mid \boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a))) = 0\,, \tag{18}$$

where $p(\boldsymbol{x}, a) := p(\boldsymbol{x})\pi(a \mid \boldsymbol{x})$, $p(a) := \int p(\boldsymbol{x}, a)d\boldsymbol{x}$ and $p(\boldsymbol{x} \mid a) := p(\boldsymbol{x}, a)/p(a)$. The l.h.s. is a sum of positive terms. We thus have, for each $a \in \text{supp}[p(a)]$,

$$\mathbb{E}_{p(\boldsymbol{x}|a)} H(\hat{q}(\cdot \mid \boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a))) = 0\,. \tag{19}$$

Since $H(\hat{q}(\cdot \mid \boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a)))$ is greater or equal to zero and is a continuous function of $\boldsymbol{x}$ (it follows from the continuity of $\boldsymbol{g}(\boldsymbol{x}, a)$, $\hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{x}')$ and $y \mapsto y \log y$[6]), Lemma 5 implies that $H(\hat{q}(\cdot \mid \boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a))) = 0$, for all $\boldsymbol{x} \in \text{supp}[p(\boldsymbol{x} \mid a)]$.

To summarize, we showed that, for all $a \in \text{supp}[p(a)]$ and all $\boldsymbol{x} \in \text{supp}[p(\boldsymbol{x} \mid a)]$, we have that $H(\hat{q}(\cdot \mid \boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a))) = 0$. Since

$$\bigcup_{a \in \text{supp}[p(a)]} \text{supp}[p(\boldsymbol{x} \mid a)] \times \{a\} = \text{supp}[p(\boldsymbol{x}, a)]\,,$$

it is equivalent to saying that, for all $(\boldsymbol{x}, a) \in \text{supp}[p(\boldsymbol{x}, a)]$, we have $H(\hat{q}(\cdot \mid \boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a))) = 0$. This means there exists a function $\boldsymbol{v} : \text{supp}[p(\boldsymbol{x}, a)] \to \hat{\mathcal{A}}$ such that, for all $(\boldsymbol{x}, a) \in \text{supp}[p(\boldsymbol{x}, a)]$ and all $\hat{a} \in \hat{\mathcal{A}}$,

$$\hat{q}(\hat{a} \mid \boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a)) = \mathbf{1}(\hat{a} = \boldsymbol{v}(\boldsymbol{x}, a))\,, \tag{20}$$

which proves the first statement.

To prove the second statement, we rewrite the above equation as

$$\hat{q}(\hat{a} \mid G(\boldsymbol{x}, a)) = \mathbf{1}(\hat{a} = \boldsymbol{v}(\boldsymbol{x}, a))\,, \tag{21}$$

where $G : \mathcal{X} \times \mathcal{A} \to \mathcal{X} \times \mathcal{X}'$ was previously defined as $G(\boldsymbol{x}, a) := (\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}, a))$. For each pair $(a, \hat{a}) \in \mathcal{A} \times \hat{\mathcal{A}}$, define the function $\hat{q}_{a,\hat{a}} : \mathcal{X} \to [0, 1]$ as $\hat{q}_{a,\hat{a}}(\boldsymbol{x}) := \hat{q}(\hat{a} \mid G(\boldsymbol{x}, a))$. Since $\hat{q}_{a,\hat{a}}$ is the composition of two continuous functions, namely $G(\cdot, a)$ and $\hat{q}(\hat{a} \mid \cdot, \cdot)$, it must also be continuous. We rewrite (21) as follows: for all $\hat{a} \in \hat{\mathcal{A}}$, all $a \in \text{supp}[p(a)]$ and all $\boldsymbol{x} \in \text{supp}[p(\boldsymbol{x} \mid a)]$, we have

$$\hat{q}_{a,\hat{a}}(\boldsymbol{x}) = \mathbf{1}(\hat{a} = \boldsymbol{v}(\boldsymbol{x}, a))\,, \tag{22}$$

Now, fix $\hat{a} \in \hat{\mathcal{A}}$ and $a \in \text{supp}[p(a)]$. It is clear from the above equation that $\hat{q}_{a,\hat{a}}(\text{supp}[p(\boldsymbol{x} \mid a)]) \subseteq \{0, 1\}$. But since $\text{supp}[p(\boldsymbol{x} \mid a)]$ is connected (Assumption 3) and $\hat{q}_{a,\hat{a}}$ is continuous, we know that the image $\hat{q}_{a,\hat{a}}(\text{supp}[p(\boldsymbol{x} \mid a)])$ must also be connected. This implies that $\hat{q}_{a,\hat{a}}(\text{supp}[p(\boldsymbol{x} \mid a)])$ is either $\{0\}$ or $\{1\}$. Thus, there is a function $\phi : \text{supp}[p(a)] \times \hat{\mathcal{A}} \to \{0, 1\}$ that outputs the value that $\hat{q}_{a,\hat{a}}$ uniformly takes on $\text{supp}[p(\boldsymbol{x} \mid a)]$. In other words, for all $\hat{a} \in \hat{\mathcal{A}}$, all $a \in \text{supp}[p(a)]$ and all $\boldsymbol{x} \in \text{supp}[p(\boldsymbol{x} \mid a)]$, we have

$$\hat{q}_{a,\hat{a}}(\boldsymbol{x}) = \phi(a, \hat{a}) \tag{23}$$

$$\mathbf{1}(\hat{a} = \boldsymbol{v}(\boldsymbol{x}, a)) = \phi(a, \hat{a})\,. \tag{24}$$

The last equation above implies that $\boldsymbol{v}(\boldsymbol{x}, a)$ is constant in $\boldsymbol{x}$, for all values of $a \in \text{supp}[p(a)]$. This means there is a function $\boldsymbol{v} : \text{supp}[p(a)] \to \hat{\mathcal{A}}$ such that, for all $a \in \text{supp}[p(a)]$ and all $\boldsymbol{x} \in \text{supp}[p(\boldsymbol{x} \mid a)]$,

$$\boldsymbol{v}(\boldsymbol{x}, a) = \boldsymbol{v}(a)\,, \tag{25}$$

which shows the second statement.

[6] In the definition of entropy, $0 \log 0$ is defined to be equal to zero, which makes $y \mapsto y \log y$ a continuous function on $[0, \infty)$ since $\lim_{y \to 0^+} y \log y = 0$.

To prove the third statement ($\boldsymbol{v}(a)$ injective), we leverage the fact that the reconstruction term is equal to zero:

$$\mathbb{E}_{p(\boldsymbol{x},\boldsymbol{x}')} \sum_{\hat{a}\in\hat{\mathcal{A}}} \hat{q}(\hat{a} \mid \boldsymbol{x},\boldsymbol{x}') \|\boldsymbol{x}' - \hat{\boldsymbol{g}}(\boldsymbol{x},\hat{a})\|_2^2 = 0 \tag{26}$$

$$\mathbb{E}_{p(\boldsymbol{x},a)} \sum_{\hat{a}\in\hat{\mathcal{A}}} \hat{q}(\hat{a} \mid \boldsymbol{x},\boldsymbol{g}(\boldsymbol{x},a)) \|\boldsymbol{g}(\boldsymbol{x},a) - \hat{\boldsymbol{g}}(\boldsymbol{x},\hat{a})\|_2^2 = 0 \tag{27}$$

$$\mathbb{E}_{p(a)}\mathbb{E}_{p(\boldsymbol{x}|a)} \sum_{\hat{a}\in\hat{\mathcal{A}}} \hat{q}(\hat{a} \mid \boldsymbol{x},\boldsymbol{g}(\boldsymbol{x},a)) \|\boldsymbol{g}(\boldsymbol{x},a) - \hat{\boldsymbol{g}}(\boldsymbol{x},\hat{a})\|_2^2 = 0 \tag{28}$$

$$\tag{29}$$

The term inside the expectation $\mathbb{E}_{p(a)}$ are greater or equal to zero, thus each of them must be equal to zero, i.e. for all $a \in \text{supp}[p(a)]$, we have

$$\mathbb{E}_{p(\boldsymbol{x}|a)} \sum_{\hat{a}\in\hat{\mathcal{A}}} \hat{q}(\hat{a} \mid \boldsymbol{x},\boldsymbol{g}(\boldsymbol{x},a)) \|\boldsymbol{g}(\boldsymbol{x},a) - \hat{\boldsymbol{g}}(\boldsymbol{x},\hat{a})\|_2^2 = 0\,. \tag{30}$$

Since the inside of the expectation is always greater or equal to zero and is a continuous function of $\boldsymbol{x}$, Lemma 5 implies that, for all $\boldsymbol{x} \in \text{supp}[p(\boldsymbol{x} \mid a)]$.

$$\sum_{\hat{a}\in\hat{\mathcal{A}}} \hat{q}(\hat{a} \mid \boldsymbol{x},\boldsymbol{g}(\boldsymbol{x},a)) \|\boldsymbol{g}(\boldsymbol{x},a) - \hat{\boldsymbol{g}}(\boldsymbol{x},\hat{a})\|_2^2 = 0\,. \tag{31}$$

To summarize, we showed that, for all $a \in \text{supp}[p(a)]$ and all $\boldsymbol{x} \in \text{supp}[p(\boldsymbol{x} \mid a)]$, (31) holds. We can thus derive that

$$0 = \sum_{\hat{a}\in\hat{\mathcal{A}}} \hat{q}(\hat{a} \mid \boldsymbol{x},\boldsymbol{g}(\boldsymbol{x},a)) \|\boldsymbol{g}(\boldsymbol{x},a) - \hat{\boldsymbol{g}}(\boldsymbol{x},\hat{a})\|_2^2 = \sum_{\hat{a}\in\hat{\mathcal{A}}} \mathbf{1}(\hat{a} = \boldsymbol{v}(\boldsymbol{x},a)) \|\boldsymbol{g}(\boldsymbol{x},a) - \hat{\boldsymbol{g}}(\boldsymbol{x},\hat{a})\|_2^2 \tag{32}$$

$$= \sum_{\hat{a}\in\hat{\mathcal{A}}} \mathbf{1}(\hat{a} = \boldsymbol{v}(a)) \|\boldsymbol{g}(\boldsymbol{x},a) - \hat{\boldsymbol{g}}(\boldsymbol{x},\hat{a})\|_2^2 \tag{33}$$

$$= \|\boldsymbol{g}(\boldsymbol{x},a) - \hat{\boldsymbol{g}}(\boldsymbol{x},\boldsymbol{v}(a))\|_2^2 \tag{34}$$

where the first line leverages (20) and the second line uses (25). This means that, for all $a \in \text{supp}[p(a)]$ and all $\boldsymbol{x} \in \text{supp}[p(\boldsymbol{x} \mid a)]$,

$$\boldsymbol{g}(\boldsymbol{x},a) = \hat{\boldsymbol{g}}(\boldsymbol{x},\boldsymbol{v}(a))\,. \tag{35}$$

We now show that $\boldsymbol{v} : \text{supp}[p(a)] \to \hat{\mathcal{A}}$ is injective. We proceed by contradiction. Suppose it is not injective. This means there exist two distinct $a_1, a_2 \in \text{supp}[p(a)]$ such that $\boldsymbol{v}(a_1) = \boldsymbol{v}(a_2)$. By Assumption 4, we know there exists an $\boldsymbol{x}_0$ that is in both $\text{supp}[p(\boldsymbol{x} \mid a_1)]$ and $\text{supp}[p(\boldsymbol{x} \mid a_2)]$. We note that

$$\boldsymbol{g}(\boldsymbol{x}_0,a_1) = \hat{\boldsymbol{g}}(\boldsymbol{x}_0,\boldsymbol{v}(a_1)) = \hat{\boldsymbol{g}}(\boldsymbol{x}_0,\boldsymbol{v}(a_2)) = \boldsymbol{g}(\boldsymbol{x}_0,a_2)\,, \tag{36}$$

where the first equality holds because $\boldsymbol{x}_0 \in \text{supp}[p(\boldsymbol{x} \mid a_1)]$ and the last equality holds because $\boldsymbol{x}_0 \in \text{supp}[p(\boldsymbol{x} \mid a_2)]$. But this is contradicting Assumption 2. Thus, $\boldsymbol{v}(a)$ is injective. □