Inference via Interpolation: Contrastive Representations Provably Enable Planning and Inference

Benjamin Eysenbach*
Princeton University
eysenbach@princeton.edu

Vivek Myers*
UC Berkeley
vmyers@berkeley.edu

Ruslan Salakhutdinov
Carnegie Mellon University
rsalakhu@cs.cmu.edu

Sergey Levine
UC Berkeley
svlevine@eecs.berkeley.edu

Abstract

Given time series data, how can we answer questions like "what will happen in the future?" and "how did we get here?" These sorts of probabilistic inference questions are challenging when observations are high-dimensional. In this paper, we show how these questions can have compact, closed form solutions in terms of learned representations. The key idea is to apply a variant of contrastive learning to time series data. Prior work already shows that the representations learned by contrastive learning encode a probability ratio. By extending prior work to show that the marginal distribution over representations is Gaussian, we can then prove that joint distribution of representations is also Gaussian. Taken together, these results show that representations learned via temporal contrastive learning follow a Gauss-Markov chain, a graphical model where inference (e.g., prediction, planning) over representations corresponds to inverting a low-dimensional matrix. In one special case, inferring intermediate representations will be equivalent to interpolating between the learned representations. We validate our theory using numerical simulations on tasks up to 46-dimensions.[1]

1 Introduction

Probabilistic modeling of time-series data has applications ranging from robotic control [1] to material science [2], from cell biology [3] to astrophysics [4]. These applications are often concerned with two questions: *predicting* future states (e.g., what will this cell look like in an hour), and *inferring* trajectories between two given states. However, answering these questions often requires reasoning over high-dimensional data, which can be challenging as most tools in the standard probabilistic toolkit require generation. Might it be possible to use discriminative methods (e.g., contrastive learning) to perform such inferences?

Many prior works aim to learn representations that are easy to predict while retaining salient bits of information. For time-series data, we want the representation to remain a sufficient statistic for distributions related to time — for example, they should retain bits required to predict future states (or representations thereof). While generative methods [5–8] have this property, they tend to be computationally expensive (see, e.g., [9]) and can be challenging to scale to high-dimensional observations.

*Equal contribution.
[1]Code: https://github.com/vivekmyers/contrastive_planning

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

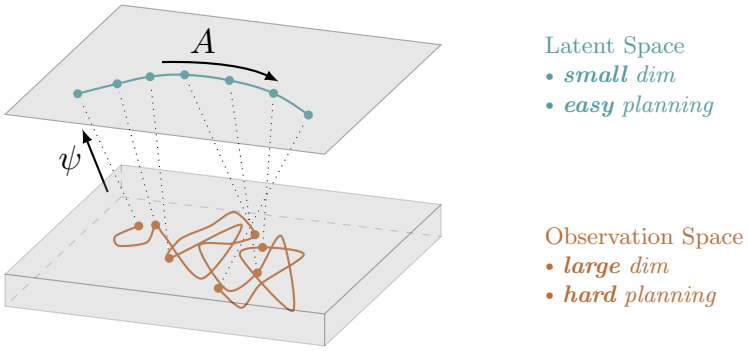


Figure 1: We apply temporal contrastive learning to observation pairs to obtain representations $(\psi(x_0), \psi(x_{t+k}))$ such that $A\psi(x_0)$ is close to $\psi(x_{t+k})$. While inferring waypoints in the high-dimensional observation space is challenging, we show that the distribution over intermediate latent representations has a closed form solution corresponding to linear interpolation between the initial and final representations.

In this paper, we will study how contrastive methods (which are discriminative, rather than generative) can perform inference over times series. Ideally, we want representations of observations x to be a sufficient statistic for temporal relationships (e.g., does x' occur after x?) but need not retain other information about x (e.g. the location of static objects). This intuition motivates us to study how contrastive representation learning methods [10–14] might be used to solve prediction and planning problems on time series data. While prior works in computer vision [10, 15] and natural language processing (NLP) [16] often study the geometry of learned representations, our results show how geometric operations such as interpolation are related to inference. Our analysis will focus on a regularized version of the symmetrized infoNCE objective [17], generating positive examples by sampling pairs of observations from the same time series data. We will study how representations learned in this way can facilitate two inference questions: prediction and planning.[2] As a stepping stone, we will build upon prior work [20] to show that regularized contrastive learning should produce representations whose marginal distribution is an isotropic Gaussian distribution.

The main contribution of this paper is to demonstrate how intermediate and future time steps in a time series can be inferred easily using contrastive representations. This inference problem captures a number of practical tasks: interpolation, in-filling, and even planning and control, where the intermediate steps represent states between a stand and goal. While ordinarily these problems require an iterative inference or optimization procedure, with contrastive representations this can be done simply by inverting a low-dimensional matrix. In one special case, inference will correspond to linear interpolation. Our first step is to prove that, under certain assumptions, the distribution over future representations has a Gaussian distribution, with a mean that is a linear function of the initial state representation (Lemma 1). This paves the main to our main result (Theorem 2): *given an initial and final state, we show that the posterior distribution over an intermediate state representations also follows a Gaussian distribution.* Said in other words, the representations follow a Gauss-Markov chain,[3] wherein any joint or conditional distribution can be computed by inverting a low-dimensional matrix [23, 24] (See Fig. 1). In one special case, inference will correspond to linearly interpolating between the representations of an initial state and final state. Section 5 provides numerical experiments.

2 Related Work

Representations for time-series data. In applications ranging from robotics to vision to NLP, users often want to learn representations of observations from time series data such that the *spatial* arrangement of representations reflects the *temporal* arrangement of the observations [10, 16, 25, 26].

[2]Following prior work [18, 19], we will use *planning* to refer to the problem of inferring intermediate states, not to refer to an optimal control problem.

[3]This probabilistic model is equivalent to a discretized Ornstein-Uhlenbeck process [21] and is also known as an AR(1) model [22, Eq. 3.1.16].

Ideally, these representations should retain information required to predict future observations and infer likely paths between pairs of observations. Many approaches use an autoencoder, learning representations that retain the bits necessary to reconstruct the input observation, while also regularizing the representations to compressed or predictable [6, 27–31]. A prototypical method is the sequential VAE [5], which is computationally expensive to train because of the reconstruction loss, but is easy to use for inference. Our work shares the aims of prior prior methods that attempt to linearize the dynamics of nonlinear systems [32–37], including videos [38, 39]. Our work aims to retain uncertainty estimates over predictions (like the sequential VAE) without requiring reconstruction. Avoiding reconstruction is appealing *practically* because it decreases the computational requirements and number of hyperparameters; and *theoretically* because it means that representations only need to retain bits about temporal relationships and not about the bits required to reconstruct the original observation.

Contrastive Learning. Contrastive learning methods circumvent reconstruction by learning representations that merely classify if two events were sampled from the same joint distribution [17, 40, 41]. When applied to representing states along trajectories, contrastive representations learn to classify whether two points lie on the same trajectory or not [10, 25, 42–44]. Empirically, prior work in computer vision and NLP has observed that contrastive learning acquires representations where interpolation between representations corresponds to changing the images in semantically meaningful ways [16, 45–49].

Our analysis will be structurally similar to prior theoretical analysis on explaining why word embeddings can solve analogies [50–52]. Our work will make a Gaussianity assumption similar to Arora et al. [51] and our Markov assumption is similar to the random walks analyzed in Arora et al. [51], Hashimoto et al. [53]. Our paper builds upon and extends these results to answer questions such as: "what is the distribution over future observations representations?" and "what is the distribution over state (representations) that would occur on the path between one observation and another?" While prior work is primarily aimed at explaining the good performance of contrastive word embeddings (see, e.g., [51]), we are primarily interested in showing how similar contrastive methods are an effective tool for inference over high-dimensional time series data. Our analysis will show how representations learned via temporal contrastive learning (i.e., without reconstruction) are sufficient statistics for inferring future outcomes and can be used for performing inference on a graphical model (a problem typically associated with generative methods).

Goal-oriented decision making. Much work on time series representations is done in service of learning goal-reaching behavior, an old problem [54, 55] that has received renewed attention in recent years [56–63]. Some of the excitement in goal-conditioned RL is a reflection of the recent success of self-supervised methods in computer vision [64] and NLP [65]. Our analysis will study a variant of contrastive representation learning proposed in prior work for goal-conditioned RL [42, 43]. These methods are widespread, appearing as learning objectives for learning value functions [26, 43, 66–73], as auxiliary objectives [71, 74–79], in objectives for model-based RL [32, 80–82], and in exploration methods [83, 84]. Our analysis will highlight connections between these prior methods, the classic successor representation [85, 86], and probabilistic inference.

Planning. Planning lies at the core of many RL and control methods, allowing methods to infer the sequence of states and actions that would occur if the agent navigated from one state to a goal state. While common methods such as PRM [87] and RRT [88] focus on building random graphs, there is a strong community focusing on planning methods based on probabilistic inference [19, 89, 90]. The key challenge is scaling to high-dimensional settings. While semi-parametric methods make progress on this problem this limitation through semi-parametric planning [91–93], it remains unclear how to scale any of these methods to high-dimensional settings when states do not lie on a low-dimensional manifold. Our analysis will show how contrastive representations may lift this limitation, with experiments validating this theory on 39-dimensional and 46-dimensional tasks.

3 Preliminaries

Our aim is to learn representations of time series data such that the spatial arrangement of representations corresponds to the temporal arrangement of the underlying data: if one example occurs shortly after another, then they should be mapped to similar representations. This problem setting arises in many areas, including video understanding and reinforcement learning. To de-

fine this problem formally, we will define a Markov process with states x_t indexed by time t:[4] $p(x_{1:T} \mid x_0) = \prod_{t=0}^{T} p(x_{t+1} \mid x_t)$. The dynamics $p(x_{t+1} \mid x_t)$ tell us the immediate next state, and we can define the distribution over states t steps in the future by marginalizing over the intermediate states, $p_t(x_t \mid x_0) = \int p(x_{1:t} \mid x_0) \, \mathrm{d}x_{1:t-1}$. A key quantity of interest will be the γ-discounted state occupancy measure, which corresponds to a time-averaged distribution over future states:

$$p_{t+}(x_{t+} = x) = (1-\gamma) \sum_{t=0}^{\infty} \gamma^t p_t(x_t = x). \tag{1}$$

Contrastive learning. Our analysis will focus on applying contrastive learning to a particular data distribution. Contrastive learning [10, 40, 94] acquires representations using "positive" pairs (x, x^+) and "negative" pairs (x, x^-). While contrastive learning typically learns just one representation, we will use two different representation for the two elements of the pair; that is, our analysis will use terms like $\phi(x)$, $\psi(x^+)$ and $\psi(x^-)$. We assume all representations lie in $\mathbb{R}^k$.

The aim of contrastive learning is to learn representations such that positive pairs have similar representations ($\phi(x) \approx \psi(x^+)$) while negative pairs have dissimilar representations ($\phi(x) \neq \psi(x^-)$). Let $p(x, x^+)$ be the joint distribution over positive pairs (i.e., $(x, x^+) \sim p(x, x^+)$). We will use the product of the marginal distributions to sample negative pairs ($(x, x^-) \sim p(x)p(x)$). Let B be the batch size, and note that the positive samples x_j^+ at index j in the batch serve as *negatives* for x_i for any $i \neq j$. Our analysis is based on the infoNCE objective without resubstitution [10, 11]:

$$\max_{\phi(\cdot),\psi(\cdot)} \mathbb{E}_{\{(x_i,x_i^+)\}_{i=1}^B \sim p(x,x^+)} \left[\sum_{i=1}^{B} \log \frac{e^{-\frac{1}{2}\|\phi(x_i)-\psi(x_i^+)\|_2^2}}{\sum_{j \neq i} e^{-\frac{1}{2}\|\phi(x_i)-\psi(x_j^+)\|_2^2}} + \log \frac{e^{-\frac{1}{2}\|\phi(x_i)-\psi(x_i^+)\|_2^2}}{\sum_{j \neq i} e^{-\frac{1}{2}\|\phi(x_j)-\psi(x_i^+)\|_2^2}} \right] \tag{2}$$

We will use the symmetrized version of this objective [17], where the denominator is the sum across rows of a logits matrix and once where it is a sum across the.

While contrastive learning is typically applied to an example x and an augmentation $x^+ \sim p(x \mid x)$ of that same example (e.g., a random crop), we will follow prior work [10, 42] in using the time series *dynamics* to generate the positive pairs, so x^+ will be an observation that occurs temporally after x. While our experiments will sample positive examples from the discounted state occupancy measure ($x^+ \sim p_{t+}(x_{t+} \mid x)$) in line with prior work [43], our analysis will also apply to different distributions (e.g., always sampling a state k steps ahead).

While prior work typically constrains the representations to have a constant norm (i.e., to lie on the unit hypersphere) [10], we will instead constrain the *expected* norm of the representations is bounded, a difference that will be important for our analysis:

$$\tfrac{1}{k} \mathbb{E}_{p(x)} \left[\|\psi(x)\|_2^2 \right] \leq c. \tag{3}$$

Because the norm scales with the dimension of the representation, we have scaled down the left side by the representation dimension, k. In practice, we will impose this constraint by adding a regularization term $\lambda \, \mathbb{E}_{p(x)} \left[\|\psi(x)\|_2^2 \right]$ to the infoNCE objective (Eq. 2) and dynamically tuning the weight λ via dual gradient descent.

3.1 Key assumptions

This section outlines the two key assumptions behind our analysis, both of which have some theoretical justification. Our main assumption examines the distribution over representations:

Assumption 1. *Regularized, temporal contrastive learning acquires representations whose marginal distribution representations* $p(\psi) \triangleq \int p(x) \, \mathbb{1}(\psi(x) = \psi) \, \mathrm{d}x$ *is an isotropic Gaussian distribution:*

$$p(\psi) = \mathcal{N}(\psi; \mu = 0, \sigma = c \cdot I). \tag{4}$$

In Appendix A.1 we extend prior work [20] provide some theoretical intuition for why this assumption should hold: namely, that the isotropic Gaussian is the distribution that maximizes entropy subject to an expected L2 norm constraint (Eq. 3) [95–97]. Our analysis also assumes that the learned representations converge to the theoretical minimizer of the infoNCE objective:

[4] This can be extended to *controlled* Markov processes appending the previous action to the observations.

Assumption 2. *Applying contrastive learning to the symmetrized infoNCE objective results in representations that encode a probability ratio:*

$$e^{-\frac{1}{2}\|\phi(x_0)-\psi(x)\|_2^2} = \frac{p_{t+}(x_{t+}=x \mid x_0)}{p(x)C}. \tag{5}$$

This assumption holds under ideal conditions [98, 99] (see Appendix A.5),[5] but we nonetheless call this an "assumption" because it may not hold in practice due to sampling and function approximation error. This assumption means the learned representations are sufficient statistics for predicting the probability (ratio) of future states: these representations must retain all the information pertinent to reasoning about *temporal* relationships, but need not retain information about the precise contents of the observations. As such, they may be much more compressed than representations learned via reconstruction.

Combined, these assumptions will allow us to express the distribution over sequences of representations as a Gauss-Markov chain. The denominator in Assumption 2, $p(x)$, may have a complex distribution, but Assumption 1 tells us that the distribution over *representations* has a simpler form. This will allow us to rearrange Assumption 2 to express the conditional distribution over representations as the product of two Gaussian likelihoods. Note that the left hand side of Assumption 2 already looks like a Gaussian likelihood.

4 Contrastive Representations Make Inference Easy

In this section, our main result will be to show how representations learned by (regularized) contrastive learning are distributed according to a Gauss-Markov chain, making it straightforward to perform inference (e.g., planning, prediction) over these representations. Our proof technique will combine (known) results about Gaussian distributions with (known) results about contrastive learning. We start by discussing an important choice of parametrization (Section 4.1) that facilitates prediction (Section 4.2) before presenting the main result in Section 4.3.

4.1 A Parametrization for Shared Encoders

This section describes the two encoders $(\psi(\cdot), \phi(\cdot))$ to compute representations of x and x^+. While prior work in computer vision and NLP literature use the same encoder for both x and x^+, this decision does not make sense for many time-series data as it would imply that our prediction for $p(x_t \mid x_0)$ is the same as our prediction for $p(x_0 \mid x_t)$. However, the difficulty of transiting from x_0 to x_t (e.g., climbing to the peak of a mountain) might be more difficult than the reverse (e.g., sledding down a mountain). Our proposed parametrization will handle this asymmetry.

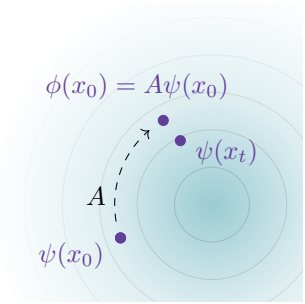


Figure 2: A parametrization for temporal contrastive learning.

We will treat the encoder $\psi(\cdot)$ as encoding the contents of the state. We will additionally learn a matrix A so that the function $\psi \mapsto A\psi$ corresponds to a (multi-step) prediction of the future representation. To map this onto contrastive learning, we will use $\phi(x) \triangleq A\psi(x)$ as the encoder for the initial state. One way of interpreting this encoder is as an additional linear projection applied on top of $\psi(\cdot)$, a design similar to those used in other areas of contrastive learning [41]. Once learned, we can use these encoders to answer questions about prediction (Section 4.2) and planning (Section 4.3).

4.2 Representations Encode a Predictive Model

Given an initial state x_0, what states are likely to occur in the future? Answering this question directly in terms of high-dimensional states is challenging, but our learned representations provide a straightforward answer. Let $\psi_0 = \psi(x_0)$ and $\psi_{t+} = \psi(x_{t+})$ be random variables representing the representations of the initial state and a future state. Our aim is to estimate the distribution over these future representations, $p(\psi_{t+} \mid \psi_0)$. We will show that the learned representations encode this distribution.

[5] While the result of Ma and Collins [98] has $C(x)$ depending on x, the symmetrized version [17] removes the dependence on x.

Lemma 1. *Under the assumptions from Section 3, the distribution over representations of future states follows a Gaussian distribution with mean parameter given by the initial state representation:*

$$p(\psi_{t+} = \psi \mid \psi_0) = \mathcal{N}\Big(\mu = \frac{c}{c+1} A\psi_0, \Sigma = \frac{c}{c+1} I\Big). \tag{6}$$

The main takeaway here is that the distribution over future representations has a convenient, closed form solution. The representation norm constraint, c, determines the shrinkage factor $\frac{c}{c+1} \in [0, 1)$; highly regularized settings (small c) move the mean closer towards the origin and decrease the variance, as visualized in Fig. 3. Regardless of the constraint c, the predicted mean is a linear function $\psi \mapsto \frac{c}{c+1} A\psi$. The proof is in Appendix A.2. The proof technique is similar to that of the law of the unconscious statistician.

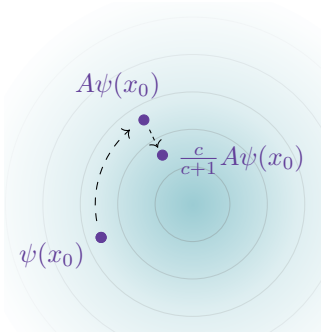


Figure 3: Predicting representations of future states.

4.3 Planning over One Intermediate State

We now show how these representations can be used for a specific type of planning: given an initial state x_0 and a future state x_{t+}, infer the representation of an intermediate "waypoint" state x_w. The next section will extend this analysis to inferring the entire sequence of intermediate states. We assume $x_0 \to x_w \to x_{t+}$ form a Markov chain where $x_w \sim p(x_{t+} \mid x_0 = x_0)$ and $x_{t+} \sim p(x_{t+} \mid x_0 = x_w)$ are both drawn from the discounted state occupancy measure (Eq. 1). Let random variable $\psi_w = \psi(x_w)$ be the representation of this intermediate state. Our main result is that the posterior distribution over waypoint *representations* has a closed form solution in terms of the initial state representation and future state representation:

Theorem 2. *Under Assumptions 1 and 2, the posterior distribution over waypoint representations is a Gaussian whose mean and covariance are linear functions of the initial and final state representations:*

$$p(\psi_w \mid \psi_0, \psi_{t+}) = \mathcal{N}\Big(\psi_w; \mu = \Sigma(A^T\psi_{t+} + A\psi_0), \Sigma^{-1} = \tfrac{c}{c+1} A^T A + \tfrac{c+1}{c} I\Big).$$

The proof (Appendix A.3) uses the Markov property together with Lemma 1. The main takeaway from this lemma is that the posterior distribution takes the form of a simple probability distribution (a Gaussian) with parameters that are linear functions of the initial and final representations.

We give three examples to build intuition:

Example 1: $A = I$ and the c is very large (little regularization). Then, the covariance is $\Sigma^{-1} \approx 2I$ and the mean is the simple average of the initial and final representations $\mu \approx \frac{1}{2}(\psi_0 + \psi_{t+})$. In other words, the waypoint representation is the midpoint of the line $\psi_0 \to \psi_{t+}$.

Example 2: A is a rotation matrix and c is very large. Rotation matrices satisfy $A^T = A^{-1}$ so the covariance is again $\Sigma^{-1} \approx 2I$. As noted in Section 4.2, we can interpret $A\psi_0$ as a *prediction* of which representations will occur after ψ_0. Similarly, $A^{-1}\psi_{t+} = A^T\psi_{t+}$ is a prediction of which representations will occur before ψ_{t+}. Theorem 2 tells us that the mean of the waypoint distribution is the simple average of these two predictions, $\mu \approx \frac{1}{2}(A^T\psi_{t+} + A\psi_0)$.

Example 3: A is a rotation matrix and $c = 0.01$ (very strong regularization). In this case $\Sigma^{-1} = \frac{0.01}{0.01+1} A^T A + \frac{0.01+1}{0.01} I \approx 100I$, so $\mu \approx \frac{1}{100}(\psi_0 + \psi_{t+}) \approx 0$. Thus, in the case of strong regularization, the posterior concentrates around the origin.

4.4 Planning over Many Intermediate States

This section extends the analysis to multiple intermediate states. Again, we will infer the posterior distribution of the representations of these intermediate states, $\psi_{w_1}, \psi_{w_2}, \cdots$. We assume that these states form a Markov chain.

Theorem 3. *Given observations from a Markov chain $x_0 \to x_1 \cdots x_{t+}$, the joint distribution over representations is a Gaussian distribution. Using $\psi_{1:n} = (\psi_{w_1}, \cdots, \psi_{w_n})$ to denote the concatenated representations of each observation, we can write this distribution as*

$$p(\psi_{1:n}) \propto \exp\big(-\tfrac{1}{2}\psi_{1:n}^T \Sigma^{-1} \psi_{1:n} + \eta^T \psi_{1:n}\big),$$

where Σ^{-1} is a tridiagonal matrix

$$\Sigma^{-1} = \begin{pmatrix} \frac{c}{c+1}A^TA + \frac{c+1}{c}I & -A^T & & \\ -A & \frac{c}{c+1}A^TA + \frac{c+1}{c}I & -A^T & \ddots \\ & & & \end{pmatrix} \quad \textit{and } \eta = \begin{pmatrix} A\psi_0 \\ 0 \\ \vdots \\ A^T\psi_{t+} \end{pmatrix}.$$

This distribution can be written in the canonical parametrization as $\Sigma = \Lambda^{-1}$ and $\mu = \Sigma\eta$. Recall that Gaussian distributions are closed under marginalization. Thus, once in this canonical parametrization, the marginal distributions can be obtained by reading off individual entries of these parameters:

$$p(\psi_i \mid \psi_0, \psi_{t+}) = \mathcal{N}\left(\psi_i; \mu_i = (\Sigma\eta)^{(i)}, \Sigma_i = (\Lambda^{-1})^{(i,i)}\right).$$

The key takeaway here is that this posterior distribution over waypoints is Gaussian, and it has a closed form expression in terms of the initial and final representations (as well as regularization parameter c and the learned matrix A).

In the general case of n intermediate states, the posterior distribution is

$$p(\psi_{w_1} \cdots \psi_{w_n} \mid \psi_0, \psi_{t+}) \propto e^{-\frac{1+\frac{1}{c}}{2}\sum_{i=1}^{n} \| \frac{c}{c+1}A\psi_{w_i} - \psi_{w_{i+1}} \|_2^2},$$

where $\psi_{w_0} = \psi_0$ and $\psi_{w_{n+1}} = \psi_{t+}$. This corresponds to a chain graphical model with edge potentials $f(\psi, \psi') = e^{-\frac{1+\frac{1}{c}}{2}\| \frac{c}{c+1}A\psi - \psi' \|_2^2}$.

Special case. To build intuition, consider the special case where A is a rotation matrix and c is very large, so $\frac{c}{c+1}A^TA + \frac{c+1}{c} \approx 2I$. In this case, Σ^{-1} is a (block) second difference matrix [100]:

$$\Sigma^{-1} = \begin{pmatrix} 2I & -I & & \\ -I & 2I & -I & \\ & & \ddots & \end{pmatrix}.$$

The inverse of this matrix has a closed form solution [101, Pg. 471], allowing us to obtain the mean of each waypoint in closed form:

$$\mu_i = (1 - \lambda(i))A\psi_0 + \lambda(i)A^T\psi_{t+}, \tag{7}$$

where $\lambda(i) = \frac{i}{n+1}$. Thus, each posterior mean is a convex combination of the (forward prediction from the) initial representation and the (backwards prediction from the) final representation. When A is the identity matrix, the posterior mean is simple linear interpolation between the initial and final representations!

5 Numerical Simulation

We include several didactic experiments to illustrate our results. All results and figures can be reproduced by running `make` in the source code: `https://github.com/vivekmyers/contrastive_planning`. The expected compute time is a few hours on a A6000 GPU. Figures in this section show error across different training and dataset split seeds.

5.1 Synthetic Dataset

To validate our analysis, we design a time series task with 2D points where inference over intermediate points (i.e., in-filling) requires nonlinear interpolation. Fig. 4 *(Top Left)* shows the dataset of time series data, starting at the origin and spiraling outwards, with each trajectory using a randomly-chosen initial angle. We applied contrastive learning with the parametrization in Section 4.1 to these data and used the learned representations to solve prediction and planning problems (see Fig. 4 for details). Note that these predictions correctly handle the nonlinear structure of these data — states nearby the initial state in Euclidean space that are not temporally adjacent are assigned low likelihood.

5.2 Solving Mazes with Inferred Representations

Our next experiment studies whether the inferred representations are useful for solving a control task. We took a 2d maze environment and dataset from prior work (Fig. 5, *Left*) [102] and learned encoders

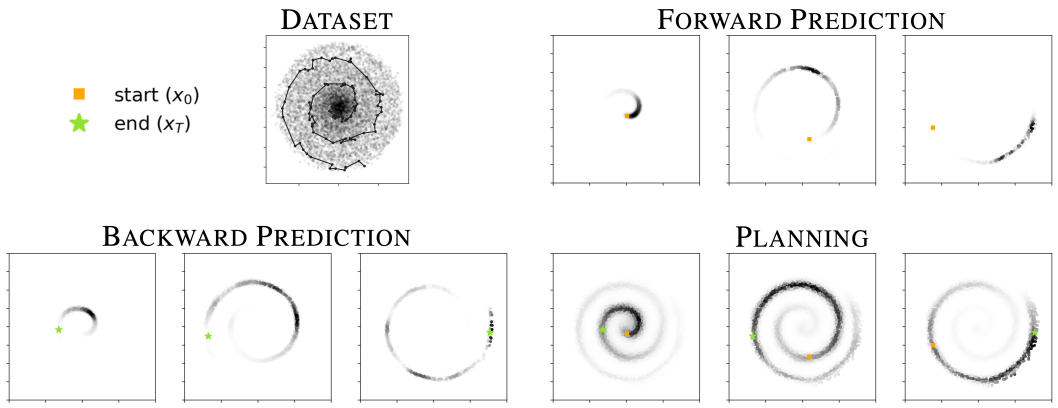


Figure 4: **Numerical simulation of our analysis.** *(Top Left)* Toy dataset of time-series data consisting of many outwardly-spiraling trajectories. We apply temporal contrastive learning to these data. *(Top Right)* For three initial observations (■), we use the learned representations to predict the distribution over future observations. Note that these distributions correctly capture the spiral structure. *(Bottom Left)* For three observations (⋆), we use the learned representations to predict the distribution over preceding observations. *(Bottom Right)* Given an initial and final observation, we plot the inferred posterior distribution over the waypoint (Section 4.3). The representations capture the shape of the distribution.

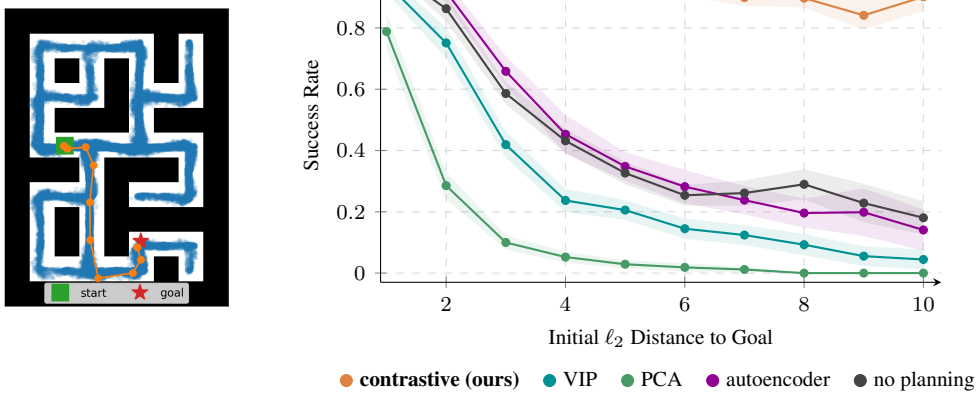


Figure 5: Using inferred paths over our contrastive representations for control boosts success rates by $4.5\times$ on the most difficult goals ($18\% \rightarrow 84\%$). Alternative representation learning techniques fail to improve performance when used for planning.

from this dataset. To solve the maze, we take the observation of the starting state and goal state, compute the representations of these states, and use the analysis in Section 4.3 to infer the sequence of intermediate representations. We visualize the results using a nearest neighbor retrieval (Fig. 5, *Left*). Figure 7 contains additional examples.

Finally, we studied whether these representations are useful for control. We implemented a simple proportional controller for this maze. As expected, this proportional controller can successfully navigate to close goals, but fails to reach distant goals (Fig. 5, *Right*). However, if we use the proportional controller to track a series of waypoints planned using our representations (i.e., the orange dots shown in Fig. 5 *(Left)*), the success rate increases by up to $4.5\times$. To test the importance of *nonlinear* representations, we compare with a "PCA" baseline that predicts waypoints by interpolating between the principal components of the initial state and goal state. The better performance of our method indicates the importance of doing the interpolation using representations that are *nonlinear* functions of the input observations. While prior methods learn representations to encode temporal distances, it

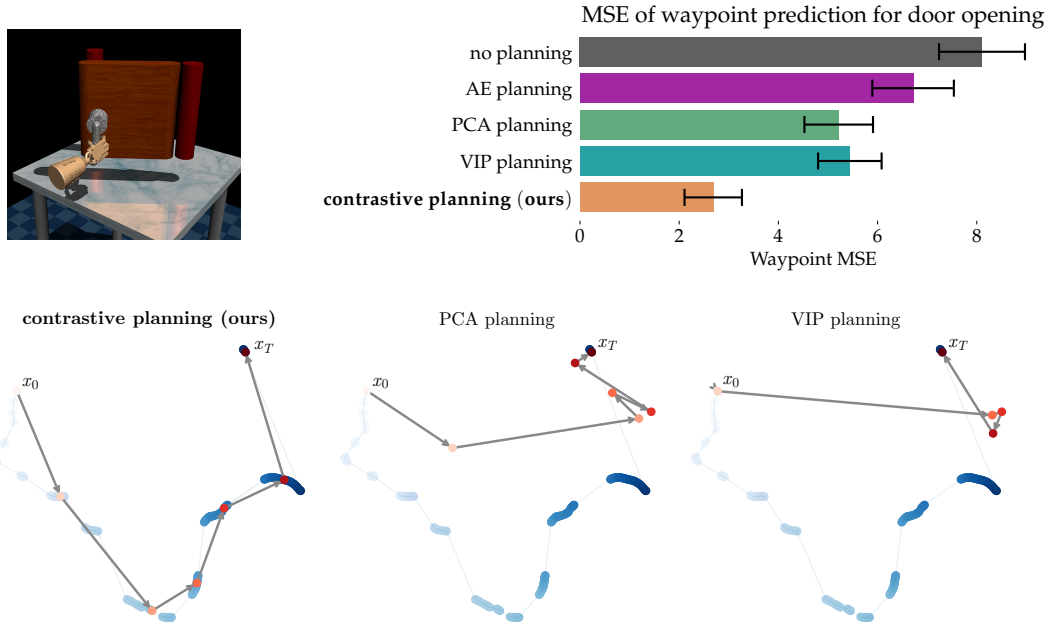


Figure 6: Planning for 39-dimensional robotic door opening. *(Top Left)* We use a dataset of trajectories demonstrating door opening from prior work [102] to learn representations. *(Top Right)* We use our method and three baselines to infer one intermediate waypoint between the first and last observation in a trajectory from a held-out validation set. Errors are measured using the mean squared error with the true waypoint observation; predicted representations are converted to observations using nearest neighbors on a validation set. *(Bottom)* We visualize a TSNE [103] of the states along the sampled trajectory as blue circles, with the transparency indicating the index along the trajectory. The inferred plan is shown as red circles connected by arrows. Our method generates better plans than alternative representation learning methods (PCA, VIP).

is unclear whether these methods support inference via interpolation. To test this hypothesis, we use one of these methods ("VIP" [69]) as a baseline. While the VIP representations likely encode similar bits as our representations, the better performance of the contrastive representations indicates that the VIP representations do not expose those bits in a way that makes planning easy.

5.3 Higher dimensional tasks

In this section we provide preliminary experiments showing the planning approach in Section 4 scales to higher dimensional tasks. We used two datasets from prior work [102]: `door-human-v0` (39-dimensional observations) and `hammer-human-v0` (46-dimensional observations). After learning encoders on these tasks, we evaluated the inference capabilities of the learned representations. Given the first and last observation from a trajectory in a validation set, we use linear interpolation (see Eq. 7) to infer the representation of five intermediate waypoint representations.

We evaluate performance in two ways. **Quantitatively**, we measure the mean squared error between each of the true waypoint observations and those inferred by our method. Since our method infers representations, rather than observations, we use a nearest-neighbor retrieval on a validation set so that we can measure errors in the space of observations. **Qualitatively**, we visualize the high-dimensional observations from the validation trajectory using a 2-dimensional TSNE [103] embedding, overlying the infer waypoints from our method; as before, we convert the representations inferred by our method to observations using nearest neighbors.

We compare with three alternative methods in Fig. 6. To test the importance of representation learning, we first naïvely interpolate between the initial and final observations ("no planning"). The poor performance of this baseline indicates that the input time series are highly nonlinear. Similarly, interpolating the principle components of the initial and final observations ("PCA") performs poorly, again highlighting that the input time series is highly nonlinear and that our representations are doing more than denoising (i.e., discarding directions of small variation). The third baseline, "VIP" [69],

learns representations to encode temporal distances using approximate dynamic programming. Like our method, VIP avoids reconstruction and learns nonlinear representations of the observations. However, the results in Fig. 6 highlight that VIP's representations do not allow users to plan by interpolation. The error bars shown in Fig. 6 (*Top Right*) show the standard deviation over 500 trajectories sampled from the validation set. For reproducibility, we repeated this entire experiment on another task, the 46-dimensional `hammer-human-v0` from D4RL. The results, shown in Appendix Fig. 8, support the conclusions above. Taken together, these results show that our procedure for interpolating contrastive representations continues to be effective on tasks where observations have dozens of dimensions.

6 Discussion

Representation learning is at the core of many high-dimensional time-series modeling questions, yet how those representations are learned is often disconnected with the inferential task. The main contribution of this paper is to show how *discriminative* techniques can be used to acquire compact representations that make it easy to answer inferential questions about time. The precise objective and parametrization we studied is not much different from that used in practice, suggesting that either our theoretical results might be adapted to the existing methods, or that practitioners might adopt these details so they can use the closed-form solutions to inference questions. Our work may also have implications for studying the structure of learned representations. While prior work often studies the geometry of representations as a post-hoc check, our analysis provides tools for studying *when* interpolation properties are guaranteed to emerge, as well as *how* to learn representations with certain desired geometric properties.

Limitations. Our analysis hinges on the two assumptions mentioned in Section 3.1, and it remains open how errors in those approximations translate into errors in our analysis. One important open question is whether it is always possible to satisfy these assumptions using sufficiently-expressive representations.

Acknowledgments

We thank Seohong Park, Gautam Reddy, Chongyi Zheng, and anonymous reviewers for feedback and discussions that shaped this project. This work was supported by Princeton Research Computing resources at Princeton University. This work was partially supported by AFOSR FA9550-22-1-0273.

References

[1] Evangelos Theodorou, Jonas Buchli, and Stefan Schaal. A Generalized Path Integral Control Approach to Reinforcement Learning. *Journal of Machine Learning Research*, 11:3137–3181, 2010.

[2] Hannes Jónsson, Greg Mills, and Karsten W Jacobsen. Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions. *Classical and Quantum Dynamics in Condensed Phase Simulations*, pp. 385–404. World Scientific, 1998.

[3] Wouter Saelens, Robrecht Cannoodt, Helena Todorov, and Yvan Saeys. A Comparison of Single-Cell Trajectory Inference Methods. *Nature Biotechnology*, 37(5):547–554, 2019.

[4] Steven R Majewski, Ricardo P Schiavon, Peter M Frinchaboy, Carlos Allende Prieto, Robert Barkhouser, Dmitry Bizyaev, Basil Blank, Sophia Brunner, Adam Burton, Ricardo Carrera, et al. The Apache Point Observatory Galactic Evolution Experiment (APOGEE). *Astronomical Journal*, 154(3):94, 2017.

[5] Shengjia Zhao, Jiaming Song, and Stefano Ermon. Towards Deeper Understanding of Variational Autoencoding Models. arXiv:1702.08658, 2017.

[6] Yizhe Zhu, Martin Renqiang Min, Asim Kadav, and Hans Peter Graf. S3vae: Self-Supervised Sequential VAE for Representation Disentanglement and Data Generation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6538–6547, 2020.

[7] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial Autoencoders. arXiv:1511.05644, 2015.

[8] Vincent Dumoulin, Ishmael Belghazi, Ben Poole, Alex Lamb, Martin Arjovsky, Olivier Mastropietro, and Aaron Courville. Adversarially Learned Inference. *International Conference on Learning Representations*, 2016.

[9] Ali Razavi, Aaron Van den Oord, and Oriol Vinyals. Generating Diverse High-Fidelity Images With Vq-VAE-2. *Neural Information Processing Systems*, 32, 2019.

[10] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation Learning With Contrastive Predictive Coding. arXiv:1807.03748, 2018.

[11] Kihyuk Sohn. Improved Deep Metric Learning With Multi-Class N-Pair Loss Objective. *Neural Information Processing Systems*, 29, 2016.

[12] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved Baselines With Momentum Contrastive Learning. arXiv:2003.04297, 2020.

[13] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive Multiview Coding. *Computer Vision–ECCV 2020: 16th European Conference*, pp. 776–794, 2020.

[14] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised Feature Learning via Non-Parametric Instance Discrimination. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3733–3742, 2018.

[15] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A Simple Framework for Contrastive Learning of Visual Representations. *International Conference on Machine Learning*, pp. 1597–1607, 2020.

[16] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781, 2013.

[17] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning Transferable Visual Models From Natural Language Supervision. *International Conference on Machine Learning*, pp. 8748–8763, 2021.

[18] Matthew Botvinick and Marc Toussaint. Planning as Inference. *Trends in Cognitive Sciences*, 16(10):485–488, 2012.

[19] Hagai Attias. Planning by Probabilistic Inference. *International Workshop on Artificial Intelligence and Statistics*, pp. 9–16, 2003.

[20] Tongzhou Wang and Phillip Isola. Understanding Contrastive Representation Learning Through Alignment and Uniformity on the Hypersphere. *International Conference on Machine Learning*, pp. 9929–9939, 2020.

[21] George E Uhlenbeck and Leonard S Ornstein. On the Theory of the Brownian Motion. *Physical Review*, 36(5):823, 1930.

[22] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2015.

[23] Dmitry M Malioutov, Jason K Johnson, and Alan S Willsky. Walk-Sums and Belief Propagation in Gaussian Graphical Models. *Journal of Machine Learning Research*, 7:2031–2064, 2006.

[24] Yair Weiss and William Freeman. Correctness of Belief Propagation in Gaussian Graphical Models of Arbitrary Topology. *Neural Information Processing Systems*, 12, 1999.

[25] Rui Qian, Tianjian Meng, Boqing Gong, Ming-Hsuan Yang, Huisheng Wang, Serge Belongie, and Yin Cui. Spatiotemporal Contrastive Video Representation Learning. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6964–6974, 2021.

[26] Benjamin Eysenbach, Ruslan Salakhutdinov, and Sergey Levine. C-Learning: Learning to Achieve Goals via Recursive Classification. *International Conference on Learning Representations*, 2020.

[27] Siddharth Karamcheti, Suraj Nair, Annie S. Chen, Thomas Kollar, Chelsea Finn, Dorsa Sadigh, and Percy Liang. Language-Driven Representation Learning for Robotics. arXiv:2302.12766, 2023.

[28] Seongmin Park and Jihwa Lee. Finetuning Pretrained Transformers Into Variational Autoencoders. arXiv:2108.02446, 2021.

[29] Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *NAACL-Hlt*, pp. 4171–4186,

2019.

[30] Micah Carroll, Orr Paradise, Jessy Lin, Raluca Georgescu, Mingfei Sun, David Bignell, Stephanie Milani, Katja Hofmann, Matthew Hausknecht, Anca Dragan, et al. UniMASK: Unified Inference in Sequential Decision Problems. arXiv:2211.10869, 2022.

[31] Junyoung Chung, Kyle Kastner, Laurent Dinh, Kratarth Goel, Aaron C Courville, and Yoshua Bengio. A Recurrent Latent Variable Model for Sequential Data. *Neural Information Processing Systems*, volume 28, 2015.

[32] Rui Shu, Tung Nguyen, Yinlam Chow, Tuan Pham, Khoat Than, Mohammad Ghavamzadeh, Stefano Ermon, and Hung Bui. Predictive Coding for Locally-Linear Control. *International Conference on Machine Learning*, pp. 8862–8871, 2020.

[33] Manuel Watter, Jost Springenberg, Joschka Boedecker, and Martin Riedmiller. Embed to Control: A Locally Linear Latent Dynamics Model for Control From Raw Images. *Neural Information Processing Systems*, 28, 2015.

[34] Ershad Banijamali, Rui Shu, Hung Bui, Ali Ghodsi, et al. Robust Locally-Linear Controllable Embedding. *International Conference on Artificial Intelligence and Statistics*, pp. 1751–1759, 2018.

[35] Brandon Cui, Yinlam Chow, and Mohammad Ghavamzadeh. Control-Aware Representations for Model-Based Reinforcement Learning. *International Conference on Learning Representations*, 2020.

[36] Tung D Nguyen, Rui Shu, Tuan Pham, Hung Bui, and Stefano Ermon. Temporal Predictive Coding for Model-Based Planning in Latent Space. *International Conference on Machine Learning*, pp. 8130–8139, 2021.

[37] Tung Nguyen, Rui Shu, Tuan Pham, Hung Bui, and Stefano Ermon. Non-Markovian Predictive Coding for Planning in Latent Space. 2020.

[38] Ross Goroshin, Michael F Mathieu, and Yann LeCun. Learning to Linearize Under Uncertainty. *Neural Information Processing Systems*, 28, 2015.

[39] Dinesh Jayaraman and Kristen Grauman. Slow and Steady Feature Analysis: Higher Order Temporal Coherence in Video. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3852–3861, 2016.

[40] Michael Gutmann and Aapo Hyvärinen. Noise-Contrastive Estimation: A New Estimation Principle for Unnormalized Statistical Models. *International Conference on Artificial Intelligence and Statistics*, pp. 297–304, 2010.

[41] Xinlei Chen and Kaiming He. Exploring Simple Siamese Representation Learning. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15750–15758, 2021.

[42] Pierre Sermanet, Corey Lynch, Yevgen Chebotar, Jasmine Hsu, Eric Jang, Stefan Schaal, Sergey Levine, and Google Brain. Time-Contrastive Networks: Self-Supervised Learning From Video. *IEEE International Conference on Robotics and Automation*, pp. 1134–1141, 2018.

[43] Benjamin Eysenbach, Tianjun Zhang, Sergey Levine, and Russ R Salakhutdinov. Contrastive Learning as Goal-Conditioned Reinforcement Learning. *Neural Information Processing Systems*, 35:35603–35620, 2022.

[44] Mengda Xu, Zhenjia Xu, Cheng Chi, Manuela Veloso, and Shuran Song. Xskill: Cross Embodiment Skill Discovery. *Conference on Robot Learning*, pp. 3536–3555, 2023.

[45] Laurenz Wiskott and Terrence J. Sejnowski. Slow Feature Analysis: Unsupervised Learning of Invariances. *Neural Computation*, 14(4):715–770, 2002.

[46] Jia-Wei Yan, Ci-Siang Lin, Fu-En Yang, Yu-Jhe Li, and Yu-Chiang Frank Wang. Semantics-Guided Representation Learning With Applications to Visual Synthesis. *International Conference on Pattern Recognition*, pp. 7181–7187, 2021.

[47] Alon Oring, Zohar Yakhini, and Yacov Hel-Or. Autoencoder Image Interpolation by Shaping the Latent Space. *International Conference on Machine Learning*, pp. 8281–8290, 2021.

[48] Ying-Cong Chen, Xiaogang Xu, Zhuotao Tian, and Jiaya Jia. Homomorphic Latent Space Interpolation for Unpaired Image-to-Image Translation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2403–2411, 2019.

[49] Xiaofeng Liu, Yang Zou, Lingsheng Kong, Zhihui Diao, Junliang Yan, Jun Wang, Site Li, Ping Jia, and Jane You. Data Augmentation via Latent Space Interpolation for Image Classification. *International Conference on Pattern Recognition*, pp. 728–733, 2018.

[50] Omer Levy and Yoav Goldberg. Linguistic Regularities in Sparse and Explicit Word Representations. *Computational Natural Language Learning*, pp. 171–180, 2014.

[51] Sanjeev Arora, Yuanzhi Li, Yingyu Liang, Tengyu Ma, and Andrej Risteski. A Latent Variable Model Approach to Pmi-Based Word Embeddings. *Transactions of the Association for Computational Linguistics*, 4:385–399, 2016.

[52] Carl Allen and Timothy Hospedales. Analogies Explained: Towards Understanding Word Embeddings. *International Conference on Machine Learning*, pp. 223–231, 2019.

[53] Tatsunori B Hashimoto, David Alvarez-Melis, and Tommi S Jaakkola. Word Embeddings as Metric Recovery in Semantic Spaces. *Transactions of the Association for Computational Linguistics*, 4:273–286, 2016.

[54] Allen Newell, John C Shaw, and Herbert A Simon. Report on a General Problem Solving Program. *IFIP Congress*, volume 256, p. 64, 1959.

[55] John E Laird, Allen Newell, and Paul S Rosenbloom. Soar: An Architecture for General Intelligence. *Artificial Intelligence*, 33(1):1–64, 1987.

[56] Lili Chen, Kevin Lu, Aravind Rajeswaran, Kimin Lee, Aditya Grover, Misha Laskin, Pieter Abbeel, Aravind Srinivas, and Igor Mordatch. Decision Transformer: Reinforcement Learning via Sequence Modeling. *Neural Information Processing Systems*, 34:15084–15097, 2021.

[57] Elliot Chane-Sane, Cordelia Schmid, and Ivan Laptev. Goal-Conditioned Reinforcement Learning With Imagined Subgoals. *International Conference on Machine Learning*, pp. 1430–1440, 2021.

[58] Cédric Colas, Tristan Karch, Olivier Sigaud, and Pierre-Yves Oudeyer. Intrinsically Motivated Goal-Conditioned Reinforcement Learning: A Short Survey. *Preprint*, 2021.

[59] Rui Yang, Yiming Lu, Wenzhe Li, Hao Sun, Meng Fang, Yali Du, Xiu Li, Lei Han, and Chongjie Zhang. Rethinking Goal-Conditioned Supervised Learning and Its Connection to Offline RL. *International Conference on Learning Representations*, 2021.

[60] Yecheng Jason Ma, Jason Yan, Dinesh Jayaraman, and Osbert Bastani. How Far I'll Go: Offline Goal-Conditioned Reinforcement Learning via f-Advantage Regression. arXiv:2206.03023, 2022.

[61] Yannick Schroecker and Charles Isbell. Universal Value Density Estimation for Imitation Learning and Goal-Conditioned Reinforcement Learning. arXiv:2002.06473, 2020.

[62] Michael Janner, Qiyang Li, and Sergey Levine. Offline Reinforcement Learning as One Big Sequence Modeling Problem. *Neural Information Processing Systems*, 34:1273–1286, 2021.

[63] Joey Hejna, Jensen Gao, and Dorsa Sadigh. Distance Weighted Supervised Learning for Offline Interaction Data. arXiv:2304.13774, 2023.

[64] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis With Latent Diffusion Models. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.

[65] OpenAI. GPT-4 Technical Report. arXiv:2303.08774, 2023.

[66] Chongyi Zheng, Benjamin Eysenbach, Homer Rich Walke, Patrick Yin, Kuan Fang, Ruslan Salakhutdinov, and Sergey Levine. Stabilizing Contrastive RL: Techniques for Robotic Goal Reaching From Offline Data. *International Conference on Learning Representations*, 2024.

[67] Stephen Tian, Suraj Nair, Frederik Ebert, Sudeep Dasari, Benjamin Eysenbach, Chelsea Finn, and Sergey Levine. Model-Based Visual Planning With Self-Supervised Functional Distances. *International Conference on Learning Representations*, 2020.

[68] Alekh Agarwal, Nan Jiang, Sham M Kakade, and Wen Sun. Reinforcement Learning: Theory and Algorithms. *CS Dept*, pp. 10–4, 2019.

[69] Yecheng Jason Ma, Shagun Sodhani, Dinesh Jayaraman, Osbert Bastani, Vikash Kumar, and Amy Zhang. VIP: Towards Universal Visual Reward and Representation via Value-Implicit Pre-Training. *International Conference on Learning Representations*, 2022.

[70] Yecheng Jason Ma, Vikash Kumar, Amy Zhang, Osbert Bastani, and Dinesh Jayaraman. LIV: Language-Image Representations and Rewards for Robotic Control. *International Conference on Machine Learning*, 2023.

[71] Suraj Nair, Aravind Rajeswaran, Vikash Kumar, Chelsea Finn, and Abhinav Gupta. R3M: A Universal Visual Representation for Robot Manipulation. *Conference on Robot Learning*, pp. 892–909, 2023.

[72] Bo Liu, Yihao Feng, Qiang Liu, and Peter Stone. Metric Residual Network for Sample Efficient Goal-Conditioned Reinforcement Learning. *AAAI Conference on Artificial Intelligence*, volume 37, pp. 8799–8806, 2023.

[73] Tongzhou Wang, Antonio Torralba, Phillip Isola, and Amy Zhang. Optimal Goal-Reaching Reinforcement Learning via Quasimetric Learning. *International Conference on Machine Learning*, pp. 36411–36430, 2023.

[74] Max Schwarzer, Ankesh Anand, Rishab Goel, R Devon Hjelm, Aaron Courville, and Philip Bachman. Data-Efficient Reinforcement Learning With Self-Predictive Representations. *International Conference on Learning Representations*, 2020.

[75] Yunhao Tang, Zhaohan Daniel Guo, Pierre Harvey Richemond, Bernardo Avila Pires, Yash Chandak, Rémi Munos, Mark Rowland, Mohammad Gheshlaghi Azar, Charline Le Lan, Clare Lyle, et al. Understanding Self-Predictive Learning for Reinforcement Learning. *International Conference on Machine Learning*, pp. 33632–33656, 2023.

[76] Adam Stooke, Kimin Lee, Pieter Abbeel, and Michael Laskin. Decoupling Representation Learning From Reinforcement Learning. *International Conference on Machine Learning*, pp. 9870–9879, 2021.

[77] Homanga Bharadhwaj, Mohammad Babaeizadeh, Dumitru Erhan, and Sergey Levine. Information Prioritization Through Empowerment in Visual Model-Based RL. *International Conference on Learning Representations*, 2021.

[78] Ankesh Anand, Evan Racah, Sherjil Ozair, Yoshua Bengio, Marc-Alexandre Côté, and R Devon Hjelm. Unsupervised State Representation Learning in Atari. *Neural Information Processing Systems*, 32, 2019.

[79] Pablo Samuel Castro, Tyler Kastner, P. Panangaden, and Mark Rowland. MICo: Improved Representations via Sampling-Based State Similarity for Markov Decision Processes. *Neural Information Processing Systems*, 2021.

[80] Raj Ghugare, Homanga Bharadhwaj, Benjamin Eysenbach, Sergey Levine, and Russ Salakhutdinov. Simplifying Model-Based RL: Learning Representations, Latent-Space Models, and Policies With One Objective. *International Conference on Learning Representations*, 2022.

[81] Cameron S. Allen. Learning Markov State Abstractions for Deep Reinforcement Learning. *Neural Information Processing Systems*, 2021.

[82] Bogdan Mazoure, Benjamin Eysenbach, Ofir Nachum, and Jonathan Tompson. Contrastive Value Learning: Implicit Models for Simple Offline RL. *Conference on Robot Learning*, pp. 1257–1267, 2023.

[83] Zhaohan Guo, Shantanu Thakoor, Miruna Pîslar, Bernardo Avila Pires, Florent Altché, Corentin Tallec, Alaa Saade, Daniele Calandriello, Jean-Bastien Grill, Yunhao Tang, et al. Byol-Explore: Exploration by Bootstrapped Prediction. *Neural Information Processing Systems*, 35:31855–31870, 2022.

[84] Yilun Du, Chuang Gan, and Phillip Isola. Curious Representation Learning for Embodied Intelligence. *IEEE/CVF International Conference on Computer Vision*, pp. 10388–10397, 2021.

[85] Peter Dayan. Improving Generalization for Temporal Difference Learning: The Successor Representation. *Neural Computation*, 5:613–624, 1993.

[86] André Barreto, Will Dabney, Rémi Munos, Jonathan J Hunt, Tom Schaul, Hado P van Hasselt, and David Silver. Successor Features for Transfer in Reinforcement Learning. *Neural Information Processing Systems*, 30, 2017.

[87] Lydia E Kavraki, Petr Svestka, J-C Latombe, and Mark H Overmars. Probabilistic Roadmaps for Path Planning in High-Dimensional Configuration Spaces. *IEEE Transactions on Robotics and Automation*, 12(4):566–580, 1996.

[88] Steven M LaValle, James J Kuffner, BR Donald, et al. Rapidly-Exploring Random Trees: Progress and Prospects. *Algorithmic and Computational Robotics: New Directions*, 5:293–308, 2001.

[89] Sep Thijssen and H. J. Kappen. Path Integral Control and State-Dependent Feedback. *Physical Review E*, 91(3):032104, 2015.

[90] Grady Williams, Andrew Aldrich, and Evangelos Theodorou. Model Predictive Path Integral Control Using Covariance Variable Importance Sampling. arXiv:1509.01149, 2015.

[91] Kuan Fang, Patrick Yin, Ashvin Nair, Homer Rich Walke, Gengchen Yan, and Sergey Levine. Generalization With Lossy Affordances: Leveraging Broad Offline Data for Learning Visuomotor Tasks. *Conference on Robot Learning*, pp. 106–117, 2023.

[92] Ben Eysenbach, Russ R Salakhutdinov, and Sergey Levine. Search on the Replay Buffer: Bridging Planning and Reinforcement Learning. *Neural Information Processing Systems*, volume 32, 2019.

[93] Tianjun Zhang, Benjamin Eysenbach, Ruslan Salakhutdinov, Sergey Levine, and Joseph E Gonzalez. C-Planning: An Automatic Curriculum for Learning Goal-Reaching Tasks. *International Conference on Learning Representations*, 2021.

[94] Nikunj Saunshi, Orestis Plevrakis, Sanjeev Arora, Mikhail Khodak, and Hrishikesh Khandeparkar. A Theoretical Analysis of Contrastive Unsupervised Representation Learning. *International Conference on Machine Learning*, pp. 5628–5637, 2019.

[95] Claude Elwood Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[96] E. T. Jaynes. Information Theory and Statistical Mechanics. *Physical Review*, 106(4):620–630, 1957.

[97] Keith Conrad. Probability Distributions and Maximum Likelihood. 2010.

[98] Zhuang Ma and Michael Collins. Noise Contrastive Estimation and Negative Sampling for Conditional Models: Consistency and Statistical Efficiency. *Empirical Methods in Natural Language Processing*, pp. 3698–3707, 2018.

[99] Ben Poole, Sherjil Ozair, Aaron Van Den Oord, Alex Alemi, and George Tucker. On Variational Bounds of Mutual Information. *International Conference on Machine Learning*, pp. 5171–5180, 2019.

[100] Nick Higham. What Is the Second Difference Matrix? `https://nhigham.com/2022/01/31/what-is-the-second-difference-matrix/`, 2022.

[101] Morris Newman and John Todd. The Evaluation of Matrix Inversion Programs. *Journal of the Society for Industrial and Applied Mathematics*, 6(4):466–476, 1958.

[102] Justin Fu, Aviral Kumar, Ofir Nachum, George Tucker, and Sergey Levine. D4RL: Datasets for Deep Data-Driven Reinforcement Learning. arXiv:2004.07219, 2020.

[103] Laurens Van der Maaten and Geoffrey Hinton. Visualizing Data Using T-SNE. *Journal of Machine Learning Research*, 9(11), 2008.

A Proofs

This section contains the proofs omitted from the main text.

A.1 Marginal Distribution over Representations is Gaussian

Recall Assumption 1, which states that the marginal distribution over representations is Gaussian.

Assumption 1. *Regularized, temporal contrastive learning acquires representations whose marginal distribution representations $p(\psi) \triangleq \int p(x)\, \mathbb{1}(\psi(x) = \psi)\, \mathrm{d}x$ is an isotropic Gaussian distribution:*

$$p(\psi) = \mathcal{N}(\psi; \mu = 0, \sigma = c \cdot I). \tag{4}$$

We will motivate this statement by connecting the optimal contrastive infoNCE objective to the maximum entropy marginal distribution over the representations.

The infoNCE objective (Eq. 2) can be decomposed into an alignment term and a uniformity term [20], where the uniformity term can be simplified as follows:

$$\begin{aligned}
&\mathbb{E}_{x \sim p(x)} \left[\log \mathbb{E}_{x^- \sim p(x)} \left[e^{-\frac{1}{2}\|A\psi(x^-) - \psi(x)\|_2^2} \right] \right] \\
&= \frac{1}{N} \sum_{i=1}^{N} \log \left(\frac{1}{N-1} \sum_{j=1 \cdots N, j \neq i} e^{-\frac{1}{2}\|A\psi(x_i) - \psi(x_j)\|_2^2} \right) \\
&= \frac{1}{N} \sum_{i=1}^{N} \log \left(\frac{1}{N-1} \sum_{j=1 \cdots N, j \neq i} \underbrace{\frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}\|A\psi(x_i) - \psi(x_j)\|_2^2}}_{\mathcal{N}(\mu = \psi(x_j); \Sigma = I)} \right) + \frac{k}{2} \log(2\pi) \\
&= \frac{1}{N} \sum_{i=1}^{N} \log \hat{p}_{\text{GMM}}(\psi(x_i)) + \frac{k}{2} \log(2\pi) \\
&= -\hat{\mathcal{H}}[\psi(x)] + \frac{k}{2} \log(2\pi).
\end{aligned}$$

The derivation above extends that in Wang and Isola [20] by considering a Gaussian distribution rather than a von Mises Fisher distribution. We are implicitly making the assumption that the marginal distributions satisfy $p(x) = p(x^-)$. This difference corresponds to our choice of using a negative squared L2 distance in the infoNCE loss rather than an inner product, a difference that will be important later in our analysis. A second difference is that we do not use the resubstitution estimator (i.e., we exclude data point x_i from our estimate of $\hat{p}_{\text{GMM}}$ when evaluating the likelihood of x_i), which we found hurt performance empirically. The takeaway from this identity is that maximizing the uniformity term corresponds to maximizing (an estimate of) the entropy of the representations.

We next prove that the maximum entropy distribution with an expected L2 norm constraint is a Gaussian distribution. Variants of this result are well known [95–97], but we include a full proof here for transparency.

Lemma 4. *The maximum entropy distribution satisfying the expected L2 norm constraint in Eq. (3) is a multivariate Gaussian distribution with mean $\mu = 0$ and covariance $\Sigma = c \cdot I$*

Proof. We start by defining the corresponding Lagrangian, with the second constraint saying that $p(x)$ must be a valid probability distribution.

$$\mathcal{L}(p) = \mathcal{H}_p[x] + \lambda_1 \left(\mathbb{E}_{p(x)} \left[\|x\|_2^2 \right] - c \cdot k \right) + \lambda_2 \left(\int p(x)\, \mathrm{d}x - 1 \right)$$

We next take the derivative w.r.t. $p(x)$:

$$\frac{\partial \mathcal{L}}{\partial p(x)} = -p(x)/p(x) - \log p(x) + \lambda_1 \|x\|_2^2 + \lambda_2$$

Setting this derivative equal to 0 and solving for $p(x)$, we get

$$p(x) = e^{-1 + \lambda_2 + \lambda_1 \|x\|_2^2}.$$

We next solve for λ_1 and λ_2 to satisfy the constraints in the Lagrangian. Note that $x \sim \mathcal{N}(\mu = 0, \Sigma = c \cdot I)$ has an expected norm $\mathbb{E}[\|x\|_2^2] = c \cdot k$, so we must have $\lambda_1 = -\frac{1}{2c}$. We determine λ_1 as the normalizing constant for a Gaussian, finally giving us:

$$p(x) = \frac{1}{(2c\pi)^{k/2}} e^{\frac{-1}{2c}\|x\|_2^2}$$

corresponding to an isotropic Gaussian distribution with mean $\mu = 0$ and covariance $\Sigma = c \cdot I$. □

A.2 Proof of Lemma 1

Below we present the proof of Lemma 1.

Lemma 1. *Under the assumptions from Section 3, the distribution over representations of future states follows a Gaussian distribution with mean parameter given by the initial state representation:*

$$p(\psi_{t+} = \psi \mid \psi_0) = \mathcal{N}\Big(\mu = \frac{c}{c+1} A\psi_0, \Sigma = \frac{c}{c+1} I\Big). \tag{6}$$

Proof. Our proof technique will be similar to that of the law of the unconscious statistician:

$$\begin{aligned}
p(\psi_{t+} \mid \psi_0) &\overset{(a)}{=} \frac{p(\psi_{t+}, \psi_0)}{\cancel{p(\psi_0)}} \propto \iint p(\psi_{t+}, x_{t+}, \psi_0, x_0)\, \mathrm{d}x_{t+}\, \mathrm{d}x_0 \\
&\overset{(b)}{=} \iint p(\psi_{t+} \mid x_{t+}) p(\psi_0 \mid x_0) p(x_{t+} \mid x_0) p(x_0)\, \mathrm{d}x_{t+}\, \mathrm{d}x_0 \\
&\overset{(c)}{\propto} \iint \mathbb{1}(\psi(x_{t+}) = \psi_{t+})\, \mathbb{1}(\psi(x_0) = \psi_0) p(x_{t+}) e^{-\frac{1}{2}\|A\psi(x_0) - \psi(x_{t+})\|_2^2} p(x_0)\, \mathrm{d}x_{t+}\, \mathrm{d}x_0 \\
&\overset{(d)}{=} e^{-\frac{1}{2}\|A\psi_0 - \psi_{t+}\|_2^2} \iint \mathbb{1}(\psi(x_{t+}) = \psi_{t+})\, \mathbb{1}(\psi(x_0) = \psi_0) p(x_{t+}) p(x_0)\, \mathrm{d}x_{t+}\, \mathrm{d}x_0 \\
&\overset{(e)}{=} e^{-\frac{1}{2}\|A\psi_0 - \psi_{t+}\|_2^2} \underbrace{\left(\int p(x_{t+})\, \mathbb{1}(\psi(x_{t+}))\, \mathrm{d}x_{t+} \right)}_{p(\psi_{t+})} \underbrace{\left(\int p(x_0)\, \mathbb{1}(\psi(x_0)\, \mathrm{d}x_0 \right)}_{p(\psi_0)} \\
&\overset{(f)}{\propto} e^{-\frac{1}{2}\|A\psi_0 - \psi_{t+}\|_2^2} e^{-\frac{1}{2c}\|\psi_{t+}\|_2^2} e^{-\frac{1}{2c}\|\psi_0\|_2^2} \\
&\overset{(g)}{\propto} e^{-\frac{1+\frac{1}{c}}{2}\left\| \frac{1}{1+\frac{1}{c}} A\psi_0 - \psi_{t+} \right\|_2^2} \\
&\propto \mathcal{N}\left(\psi_{t+}; \mu = \tfrac{c}{c+1} A\psi_0, \Sigma = \tfrac{c}{c+1} I\right).
\end{aligned}$$

In *(a)* we applied Bayes' Rule and removed the denominator, which is a constant w.r.t. ψ_{t+}. In *(b)* we factored the joint distribution, noting that ψ_{t+} and ψ_0 are deterministic functions of x_{t+} and x_0 respectively, so they are conditionally independent from the other random variables. In *(c)* we used Assumption 2 after solving for $p(x_{t+} \mid x_0) = p(x_{t+}) e^{-\frac{1}{2}\|A\psi(x_0) - \psi(x)\|_2^2}$. In *(d)* we noted that when the integrand is nonzero, it takes on a constant value of $e^{-\frac{1}{2}\|A\psi_0 - \psi_{t+}\|_2^2}$, so we can move that constant outside the integral. In *(e)* we used the definition of the marginal representation distribution (Eq. 6). In *(f)* we used Assumption 1 to write the marginal distributions $p(\psi_{t+})$ and $p(\psi_0)$ as Gaussian distributions. We removed the normalizing constants, which are independent of ψ_{t+}. In *(g)* we completed the square and then recognized the expression as the density of a multivariate Gaussian distribution. □

A.3 Proof of Theorem 2: Waypoint Distribution

Theorem 2. *Under Assumptions 1 and 2, the posterior distribution over waypoint representations is a Gaussian whose mean and covariance are linear functions of the initial and final state representations:*

$$p(\psi_w \mid \psi_0, \psi_{t+}) = \mathcal{N}\Big(\psi_w; \mu = \Sigma(A^T \psi_{t+} + A\psi_0), \Sigma^{-1} = \tfrac{c}{c+1} A^T A + \tfrac{c+1}{c} I\Big).$$

Proof.

$$
\begin{aligned}
p(\psi_w \mid \psi_0, \psi_{t+}) &\overset{(a)}{=} \frac{p(\psi_{t+} \mid \psi_w) p(\psi_w \mid \psi_0)}{\cancel{p(\psi_{t+} \mid \psi_0)}} \\
&\overset{(b)}{\propto} e^{-\frac{1+\frac{1}{c}}{2} \| \frac{c}{c+1} A\psi_w - \psi_{t+} \|_2^2} e^{-\frac{1+\frac{1}{c}}{2} \| \frac{c}{c+1} A\psi_0 - \psi_w \|_2^2} \\
&\overset{(c)}{\propto} e^{-\frac{1}{2}(\psi_w - \mu)^T \Sigma^{-1} (\psi_w - \mu)} = \mathcal{N}(\psi_w; \mu, \Sigma)
\end{aligned}
$$

where $\Sigma^{-1} = \frac{c}{c+1} A^T A + \frac{c+1}{c} I$ and $\mu = \Sigma(A^T \psi_{t+} + A\psi_0)$. □

In line *(a)* we used the definition of the conditional distribution and then simplified the numerator using the Markov property. Line *(b)* uses the Lemma 1. Line *(c)* completes the square, the details of which are below:

$$
\begin{aligned}
&\frac{1}{2} \cdot \frac{c+1}{c} \left(\left\| \frac{c}{c+1} A\psi_w - \psi_{t+} \right\|_2^2 + \left\| \frac{c}{c+1} A\psi_0 - \psi_w \right\|_2^2 \right) \\
&\quad = \frac{1}{2} \cdot \frac{c+1}{c} \left(\psi_w^T \Big(\frac{c}{c+1} A \Big)^T \Big(\frac{c}{c+1} A \Big) \psi_w - 2\psi_{t+}^T \Big(\frac{c}{c+1} A \Big) \psi_w + \cancel{\psi_{t+}^T \psi_{t+}} \cdot \right. \\
&\qquad \left. + \cancel{\psi_0^T \Big(\frac{c}{c+1} A \Big)^T \Big(\frac{c}{c+1} A \Big) \psi_0} - 2\psi_0^T \Big(\frac{c}{c+1} A \Big)^T \psi_w + \psi_w^T \psi_w \right) \\
&\quad \overset{\text{const.}}{=} \frac{1}{2} \cdot \frac{c+1}{c} \left(\psi_w^T \left(\Big(\frac{c}{c+1} \Big)^2 A^T A + I \right) \psi_w - 2 \cdot \frac{c}{c+1} \big(A^T \psi_{t+} + A\psi_0 \big)^T \psi_w \right) \\
&\quad = \frac{1}{2} \psi_w^T \Bigg(\underbrace{\frac{c}{c+1} A^T A + \frac{c+1}{c} I}_{\Sigma^{-1}} \Bigg) \psi_w - \big(A^T \psi_{t+} + A\psi_0 \big)^T \psi_w \\
&\quad \overset{\text{const.}}{=} (\psi_w - \mu)^T \Sigma^{-1} (\psi_w - \mu),
\end{aligned}
$$

where $\Sigma^{-1} = \frac{c}{c+1} A^T A + \frac{c+1}{c} I$ and $\mu = \Sigma(A^T \psi_{t+} + A\psi_0)$. Above, we have used $\overset{\text{const.}}{=}$ to denote equality up to an additive constant that is independent of ψ_w.

A.4 Proof of Theorem 3: Planning over Many Intermediate States

Theorem 3. *Given observations from a Markov chain $x_0 \to x_1 \cdots x_{t+}$, the joint distribution over representations is a Gaussian distribution. Using $\psi_{1:n} = (\psi_{w_1}, \cdots, \psi_{w_n})$ to denote the concatenated representations of each observation, we can write this distribution as*

$$
p(\psi_{1:n}) \propto \exp\big(-\tfrac{1}{2} \psi_{1:n}^T \Sigma^{-1} \psi_{1:n} + \eta^T \psi_{1:n} \big),
$$

where Σ^{-1} is a tridiagonal matrix

$$
\Sigma^{-1} = \begin{pmatrix}
\frac{c}{c+1} A^T A + \frac{c+1}{c} I & -A^T & & \\
-A & \frac{c}{c+1} A^T A + \frac{c+1}{c} I & -A^T & \\
 & & \ddots &
\end{pmatrix} \quad \text{and } \eta = \begin{pmatrix} A\psi_0 \\ 0 \\ \vdots \\ A^T \psi_{t+} \end{pmatrix}.
$$

Proof. We start by recalling that the waypoints form a Markov chain, so we can express their joint density as a product of conditional densities:

$$
p(\psi_{1:n}) = p(\psi_0) p(\psi_1 \mid \psi_0) p(\psi_2 \mid \psi_1) \cdots p(\psi_n \mid \psi_{n-1}).
$$

The aim of this lemma is to express the joint distribution over multiple waypoints, given an initial and final state representation:

$$
p(\psi_{1:n} \mid \psi_0, \psi_{t+}) \overset{(a)}{=} \frac{p(\psi_{1:n} \mid \psi_0) p(\psi_{t+} \mid \psi_n)}{\cancel{p(\psi_{t+} \mid \psi_0)}}
$$

$$
\begin{aligned}
&\overset{(b)}{\propto} p(\psi_1 \mid \psi_0)p(\psi_2 \mid \psi_1)\cdots p(\psi_{t+} \mid \psi_n) \\
&\overset{(c)}{\propto} \exp\Big(-\tfrac{1}{2}\tfrac{c+1}{c}\|\tfrac{c}{c+1}A\psi_0 - \psi_1\|_2^2 - \tfrac{1}{2}\tfrac{c+1}{c}\|\tfrac{c}{c+1}A\psi_1 \\
&\qquad\qquad - \psi_2\|_2^2 - \cdots - \tfrac{1}{2}\tfrac{c+1}{c}\|\tfrac{c}{c+1}A\psi_n - \psi_{t+}\|_2^2\Big) \\
&\overset{(d)}{=} \exp\Big(\cancel{-\tfrac{1}{2}\tfrac{c}{c+1}\psi_0^T A^T A\psi_0} + \psi_0^T A^T\psi_1 - \tfrac{1}{2}\tfrac{c+1}{c}\psi_1^T\psi_1 \\
&\qquad\qquad - \tfrac{1}{2}\tfrac{c}{c+1}\psi_1^T A^T A\psi_1 + \psi_1^T A^T\psi_2 - \tfrac{1}{2}\tfrac{c+1}{c}\psi_2^T\psi_2 \\
&\qquad\qquad - \tfrac{1}{2}\tfrac{c}{c+1}\psi_2^T A^T A\psi_2 + \psi_2^T A^T\psi_3 - \tfrac{1}{2}\tfrac{c+1}{c}\psi_3^T\psi_3 \\
&\qquad\qquad\qquad\qquad \vdots \\
&\qquad\qquad - \tfrac{1}{2}\tfrac{c}{c+1}\psi_n^T A^T A\psi_n + \psi_n^T A^T\psi_{t+} - \cancel{\tfrac{1}{2}\tfrac{c+1}{c}\psi_{t+}^T\psi_{t+}}\Big) \\
&= \exp\big(-\tfrac{1}{2}\psi_{1:n}^T\Sigma^{-1}\psi_{1:n} + \eta^T\psi_{1:n}\big),
\end{aligned}
$$

where

$$
\Sigma^{-1} = \begin{pmatrix}
\frac{c}{c+1}A^TA + \frac{c+1}{c}I & -A^T & & \\
-A & \frac{c}{c+1}A^TA + \frac{c+1}{c}I & -A^T & \\
 & -A & \frac{c}{c+1}A^TA + \frac{c+1}{c}I & \\
 & & & \ddots
\end{pmatrix} \text{ and } \eta = \begin{pmatrix} A\psi_0 \\ 0 \\ \vdots \\ 0 \\ A^T\psi_{t+} \end{pmatrix}.
$$

In *(a)* we applied Bayes' rule and removed the denominator because it is a constant with respect to $\psi_{1:n}$. In *(b)* we applied the Markov assumption. In *(c)* we used Lemma 1 to express the conditional probabilities as Gaussians, ignoring the proportionality constants (which are independent of ψ. In *(d)* we simplified the exponents, removing terms that do not depend on $\psi_{1:n}$. □

A.5 Formalizing Assumption 2

Assumption 2 relates the learned contrastive critic to a log-likelihood ratio between the positive and negative data distribution.

Assumption 2. *Applying contrastive learning to the symmetrized infoNCE objective results in representations that encode a probability ratio:*

$$
e^{-\frac{1}{2}\|\phi(x_0)-\psi(x)\|_2^2} = \frac{p_{t+}(x_{t+} = x \mid x_0)}{p(x)C}. \tag{5}
$$

We can justify this assumption by analyzing the general solution to the symmetrized version of the Oord et al. [10] infoNCE objective, which we do in Lemma 5. Applying this lemma to our representation learning objective (2) for sufficiently large batch size B then yields Eq. (5), with the function approximator $\|\phi(x) - \psi(x^+)\|^2 \approx f(x, x^+)$.

Lemma 5. *The solution to the optimization problem*

$$
\max_{f(x,x^+)} \lim_{B\to\infty} \mathbb{E}_{\{(x_i,x_i^+)\}_{i=1}^B \sim p(x,x^+)} \left[\tfrac{1}{B}\sum_{i=1}^B \log \frac{e^{f(x_i,x_i^+)}}{\sum_{j\neq i} e^{f(x_i,x_j^+)}} + \log \frac{e^{f(x_i,x_i^+)}}{\sum_{j\neq i} e^{f(x_j,x_i^+)}} \right] \tag{8}
$$

satisfies

$$
f(x, x^+) = \log\left(\frac{p(x^+ \mid x)}{p(x^+)C}\right) \tag{9}
$$

for some C.

Proof of Lemma 5. We first break down the LHS and RHS of Eq. (2):

$$
\max_{f} \lim_{B\to\infty} \mathbb{E}_{\{(x_i,x_i^+)\}_{i=1}^B \sim p(x,x^+)} \left[\frac{1}{B}\sum_{i=1}^B \underbrace{\log \frac{e^{f(x_i,x_i^+)}}{\sum_{j\neq i} e^{f(x_i,x_j^+)}}}_{\mathcal{J}_1} + \underbrace{\log \frac{e^{f(x_i,x_i^+)}}{\sum_{j\neq i} e^{f(x_j,x_i^+)}}}_{\mathcal{J}_2} \right]
$$

$$\mathcal{J}_1(f) = \lim_{B\to\infty} \mathbb{E}_{\{(x_i,x_i^+)\}_{i=1}^B \sim p(x,x^+)} \left[\frac{1}{B} \sum_{i=1}^{B} \log \frac{e^{f(x_i,x_i^+)}}{\sum_{j\neq i} e^{f(x_i,x_j^+)}} \right]$$

$$\mathcal{J}_2(f) = \lim_{B\to\infty} \mathbb{E}_{\{(x_i,x_i^+)\}_{i=1}^B \sim p(x,x^+)} \left[\frac{1}{B} \sum_{i=1}^{B} \log \frac{e^{f(x_i,x_i^+)}}{\sum_{j\neq i} e^{f(x_j,x_i^+)}} \right]$$

We now use the following result from Ma and Collins [98]:

Lemma 6. *The optimal solutions f_1 and f_2 for $\mathcal{J}_1$ and $\mathcal{J}_2$ satisfy*

$$f_1(x, x^+) = \log p(x \mid x^+) - \log c_1(x) \tag{10}$$

$$f_2(x, x^+) = \log p(x^+ \mid x) - \log c_2(x^+) \tag{11}$$

for arbitrary $c_1(x), c_2(x^+)$.

For any C, when $c_1(x) = Cp(x)$ and $c_2(x^+) = Cp(x^+)$,

$$f_1(x, x^+) = \log\left(\frac{p(x \mid x^+)}{p(x)C}\right) = \log\left(\frac{p(x^+ \mid x)}{p(x^+)C}\right) = f_2(x, x^+). \tag{12}$$

It follows that Eq. (12) maximizes both $\mathcal{J}_1$ and $\mathcal{J}_2$, and is precisely the optimal solution Eq. (9) for Eq. (8). □

What does C represent? From Eq. (9), we can connect C to the mutual information $I(x, x^+)$:

$$C = \frac{\mathbb{E}_{(x,x^+)\sim p(x,x^+)}\big[f(x, x^+)\big]}{I(x, x^+)}. \tag{13}$$

Proof of Lemma 6. We can first consider $\mathcal{J}_1$ without loss of generality. Denoting

$$g(x, x^+) = e^{f(x,x^+)},$$

we take the functional derivative:

$$\begin{aligned}
\delta\mathcal{J}_1(\log g) &= \lim_{B\to\infty} \delta\, \mathbb{E}_{\{(x_i,x_i^+)\}_{i=1}^B \sim p(x,x^+)} \left[\tfrac{1}{B} \sum_{i=1}^{B} \log \frac{g(x_i,x_i^+)}{\sum_{j\neq i} g(x_i,x_j^+)} \right] \\
&= \lim_{B\to\infty} \mathbb{E}_{\{(x_i,x_i^+)\}_{i=1}^B \sim p(x,x^+)} \left[\tfrac{1}{B} \sum_{i=1}^{B} \frac{(\sum_{j\neq i} g(x_i,x_j^+))\delta g(x_i,x_i^+) - g(x_i,x_i^+)\delta(\sum_{j\neq i} g(x_i,x_j^+))}{g(x_i,x_i^+)(\sum_{j\neq i} g(x_i,x_j^+))} \right] \\
&= \lim_{B\to\infty} \mathbb{E}_{\{(x_i,x_i^+)\}_{i=1}^B \sim p(x,x^+)} \left[\tfrac{1}{B} \sum_{i=1}^{B} \frac{\delta g(x_i,x_i^+)}{g(x_i,x_i^+)} - \frac{\delta(\sum_{j\neq i} g(x_i,x_j^+))}{\sum_{j\neq i} g(x_i,x_j^+)} \right] \\
&= \lim_{B\to\infty} \mathbb{E}_{\{(x_i,x_i^+)\}_{i=1}^B \sim p(x_i,x^+)} \Bigg[\tfrac{1}{B} \sum_{i=1}^{B} \int \Big(\big(\tfrac{\delta g(x_i,x^+)}{g(x_i,x^+)}\big) p(x^+ \mid x_i) \\
&\qquad\qquad - \sum_{k\neq i} \big(\tfrac{\delta g(x_i,x^+)}{g(x_i,x^+) - g(x_i,x_k^+) + \sum_{j\neq i} g(x_i,x_j^+)}\big) p(x^+) \Big)\, \mathrm{d}x^+ \Bigg] \\
&= \lim_{B\to\infty} \mathbb{E}_{\{(x_i,x_i^+)\}_{i=1}^B \sim p(x,x^+)} \Bigg[\tfrac{1}{B} \sum_{i=1}^{B} \int \delta g(x_i,x^+) \Big(\tfrac{p(x^+|x_i)}{g(x_i,x^+)} \\
&\qquad\qquad - \underbrace{\mathbb{E}_{\{(x_i,x_i^+)\}_{i=1}^B} \Big[\tfrac{1}{\sum_{j\neq i} g(x_i,x_j^+)} \Big]}_{\text{as } B\to\infty} p(x^+) \Big)\, \mathrm{d}x^+ \Bigg]
\end{aligned}$$

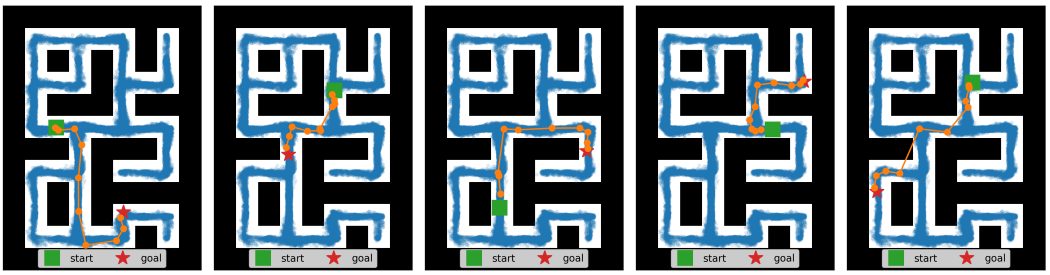
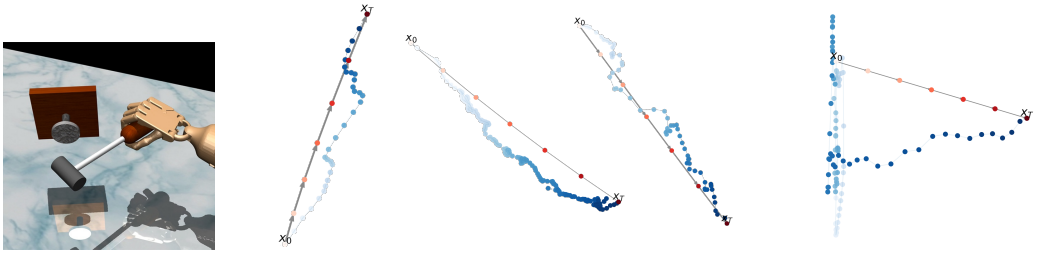


Figure 7: Our approach enables a goal-conditioned policy to reach farther targets (red) from the start (green) by planning over intermediate waypoints (orange).



(a) Contrastive representations (b) PCA representations

Figure 8: Planning for 46-dimensional robotic hammering. *(Left)* A dataset of trajectories demonstrating a hammer knocking a nail into a board [102]. *(Center)* We visualize the learned representations as blue circles, with the transparency indicating the index of that observation along the trajectory. We also visualize the inferred plan (Section 4.3) as red circles connected by arrows. *(Right)* Representations learned by PCA on the same trajectory as *(a, left)*.

$$
\begin{aligned}
&= \lim_{B\to\infty} \mathbb{E}_{\{(x_i,x_i^+)\}_{i=1}^B \sim p(x,x^+)} \Bigg[\frac{1}{B} \sum_{i=1}^{B} \int \delta g(x_i, x^+) \Big(\frac{p(x^+|x_i)}{g(x_i,x^+)} \\
&\qquad\qquad - \underbrace{\mathbb{E}_{\{(x_i,x_i^+)\}_{i=1}^B} \Big[\frac{1}{\sum_{j\neq i} g(x_i,x_j^+)} \Big]}_{\triangleq k(x_i) \text{ indep. of } x^+} p(x^+) \Big) \, \mathrm{d}x^+ \Bigg] \\
&= \lim_{B\to\infty} \mathbb{E}_{\{(x_i,x_i^+)\}_{i=1}^B \sim p(x,x^+)} \Bigg[\frac{1}{B} \sum_{i=1}^{B} \int \delta g(x_i, x^+) \Big(\frac{p(x^+|x_i)}{g(x_i,x^+)} - k(x_i)p(x^+) \Big) \, \mathrm{d}x^+ \Bigg] \\
&= \int \delta g(x, x^+) \big(\tfrac{p(x^+|x)}{g(x,x^+)} - k(x)p(x^+) \big) \, \mathrm{d}x^+ .
\end{aligned}
$$

This is zero when

$$
g(x, x^+) = \frac{p(x \mid x^+)}{k(x)p(x)},
$$

i.e.,

$$
f(x, x^+) = \log p(x \mid x^+) - \log \underbrace{c_1(x)}_{k(x)p(x)}
$$

as in Eq. (10), and Eq. (11) follows similarly, exchanging x and x^+. □

B Additional Experiments

Fig. 7 visualizes the inferred waypoints from the task in Fig. 5. Fig. 8 visualizes the representations learned on a 46-dimensional robotic hammering task (see Section 5.3).

B.1 Stock Prediction

We show results on a stock opening price task in Fig. 9.

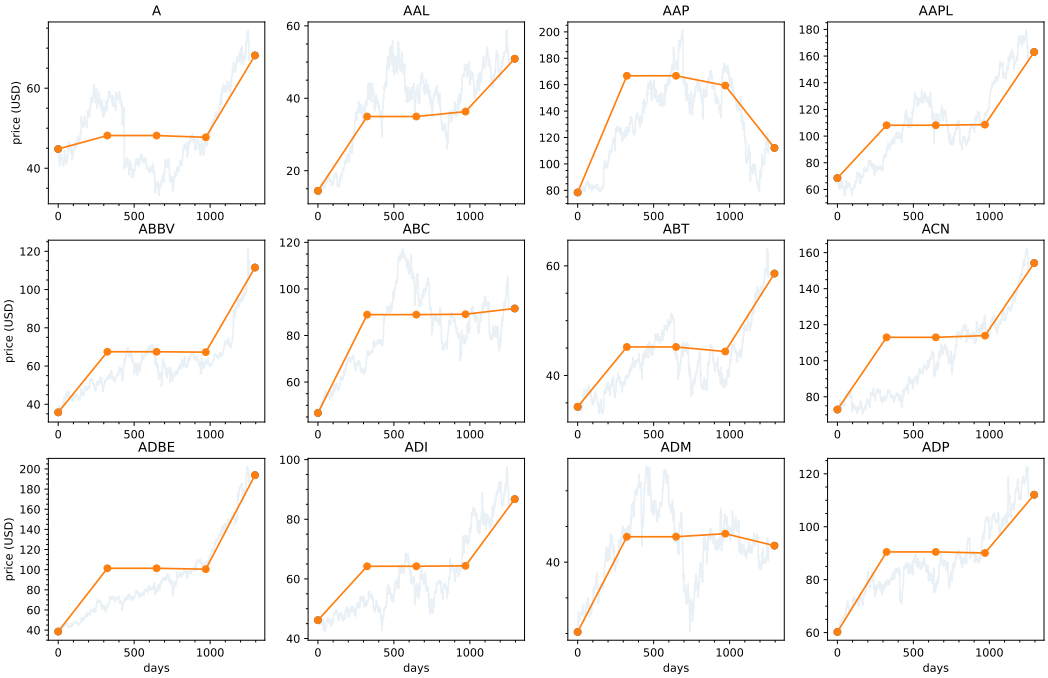


Figure 9: **Stock Prediction.** We apply temporal contrastive learning to time series data of the stock market. Data are the opening prices for the 500 stocks in the S&P 500, over a four year window. We remove 30 stocks that are missing data. For evaluation, we choose a 100 day window from a validation set, and use Theorem 2 to perform "inpainting," predicting the intermediate stock prices *jointly* for all stocks (orange), given the first and last stock price. The true stock prices are shown in blue. While we do not claim that this is a state-of-the-art model for stock prediction, this experiment demonstrates another potential application of our theoretical results.