NeurIPS 2026

NEURAL INFORMATION PROCESSING SYSTEMS

Beyond Isotropy in JEPAs: Hamiltonian Geometry and Symplectic Prediction

Robert Jenkinson Alvarez

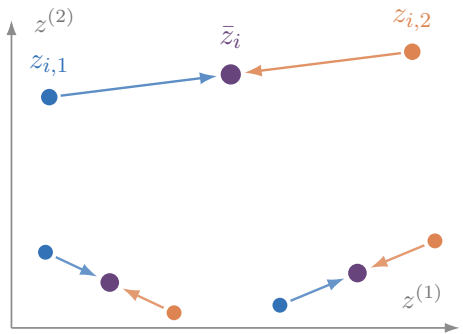
Main-Tack Poster

LeJEPA: views and latent representations

$$X \sim \mathcal{D}, \qquad \tau_1, \tau_2 \overset{\text{iid}}{\sim} \mathcal{A}$$

Views		Latent representations
$V_1 = \tau_1(X)$	$\xrightarrow{f_\theta}$	$Z_1 = f_\theta(V_1) \in \mathbb{R}^d$
$V_2 = \tau_2(X)$	$\xrightarrow{f_\theta}$	$Z_2 = f_\theta(V_2) \in \mathbb{R}^d$

LeJEPA: alignment



Per-image mean

$$\bar{z}_i = \frac{z_{i,1} + z_{i,2}}{2}.$$

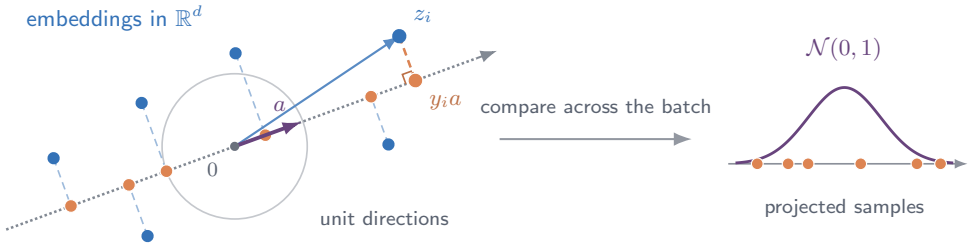
Alignment objective

$$\mathcal{L}_{\text{align}} = \frac{1}{4Bd} \sum_{i=1}^{B} \sum_{v=1}^{2} \|z_{i,v} - \bar{z}_i\|_2^2$$

$$= \frac{1}{8Bd} \sum_{i=1}^{B} \|z_{i,1} - z_{i,2}\|_2^2.$$

B images, two global views, d latent coordinates. Normalization used in our experiments.

LeJEPA: SIGReg



$$a \sim \mathcal{U}(\mathbb{S}^{d-1})$$

$$y_{iv} = a^\top z_{iv}$$

$$\mathcal{L}_{\text{SIGReg}} = \frac{1}{2}\sum_{v=1}^{2} \mathbb{E}_a\left[D_{\text{EP}}\left(\{y_{iv}\}_{i=1}^{B}, \mathcal{N}(0,1)\right)\right].$$

$$\mathcal{L}_{\text{LeJEPA}} = (1-\lambda)\mathcal{L}_{\text{align}} + \lambda\mathcal{L}_{\text{SIGReg}}.$$

D_{EP}: Epps–Pulley discrepancy. Expectation approximated with random directions.

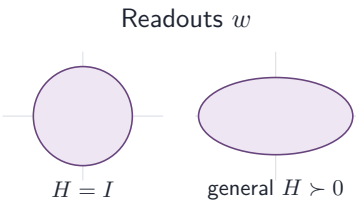
Downstream geometry

$$\widetilde{Z} = Z - \mathbb{E}[Z], \qquad \Sigma = \mathrm{Cov}(Z), \qquad H \succ 0.$$

Linear tasks

$$y = w^\top \widetilde{Z}$$

$$\mathcal{W}_H = \{w \in \mathbb{R}^d : \ w^\top H^{-1} w \le 1\}.$$



Energy budget

$$\mathbb{E}[\widetilde{Z}^\top H \widetilde{Z}] = \mathrm{tr}(H\Sigma) = c, \qquad c > 0.$$

The optimal covariance

Minimax task variance

$$\min_{\substack{\Sigma \succeq 0 \\ \operatorname{tr}(H\Sigma)=c}} \underbrace{\sup_{w \in \mathcal{W}_H} w^\top \Sigma w}_{V(\Sigma)}$$

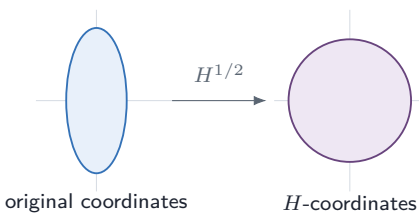
Unique minimizer

$$\Sigma^\star = \frac{c}{d} H^{-1}, \qquad V(\Sigma^\star) = \frac{c}{d}.$$

Maximum entropy

$$\widetilde{Z}^\star \sim \mathcal{N}\left(0, \frac{c}{d} H^{-1}\right).$$

Isotropic in H-coordinates



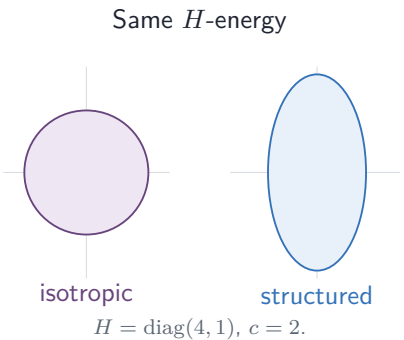
$$H^{1/2} \Sigma^\star H^{1/2} = \frac{c}{d} I.$$

The price of isotropy

$$\Sigma_{\text{iso}} = \frac{c}{\operatorname{tr}(H)} I.$$

$$\rho(H) = \frac{V(\Sigma_{\text{iso}})}{V(\Sigma^\star)} = \frac{d\lambda_{\max}(H)}{\operatorname{tr}(H)} \geq 1.$$

$$\rho(H) = 1 \quad \Longleftrightarrow \quad H \propto I.$$



Unknown H: no universal covariance shape.

Marginals leave the predictive coupling undetermined.

From marginal geometry to predictive structure

Unknown geometry

$$\left\{ \frac{c}{d} H^{-1} : H \succ 0 \right\} = \{ \Sigma \succ 0 \}.$$

Scalar example: $\Sigma = \sigma^2$



Same marginals

$$Z_a, Z_b \sim \mathcal{N}(0, \Sigma), \qquad \Sigma \succ 0.$$

Different predictors

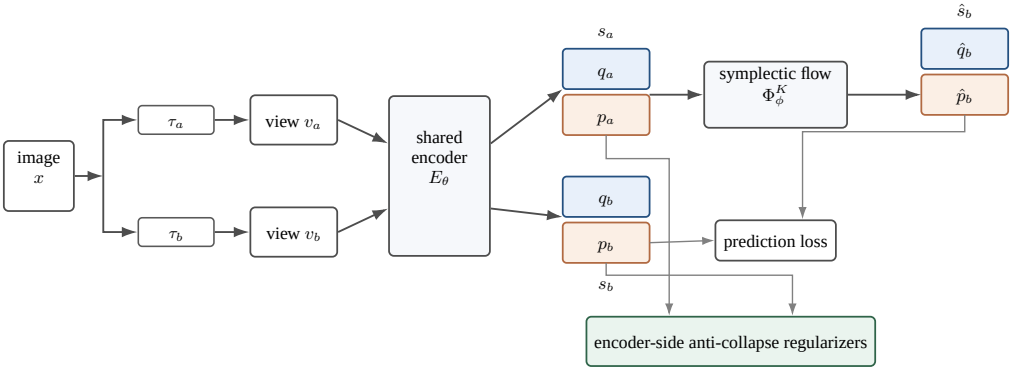
$$\mathbb{E}[Z_b \mid Z_a] = C\Sigma^{-1} Z_a.$$

Jointly Gaussian views; $C = \mathrm{Cov}(Z_b, Z_a)$.

Learn a structured coupling; allow anisotropic marginals.

HamJEPA: a Hamiltonian predictor

A phase-space encoder produces a content coordinate q and an auxiliary momentum p.



The predictor models the view-to-view transition. Encoder-side penalties control collapse.

A canonical Hamiltonian coupling

Continuous linear flow F_t

Stationary oracle law

$$(Q,P) \sim \mathcal{N}\left(0, \frac{c}{d}\begin{pmatrix} H^{-1} & 0 \\ 0 & I \end{pmatrix}\right).$$

Momentum reversal

$$\mathsf{R}F_t\mathsf{R} = F_{-t}, \qquad \mathsf{R}(Q,P) = (Q,-P).$$

Unit kinematics

$$\dot{Q} = P$$

Unique dynamics

$$\dot{Q} = P, \qquad \dot{P} = -HQ.$$

$$\mathcal{H}_H(q,p) = \tfrac{1}{2} q^\top H q + \tfrac{1}{2}\|p\|^2.$$

Learned energy

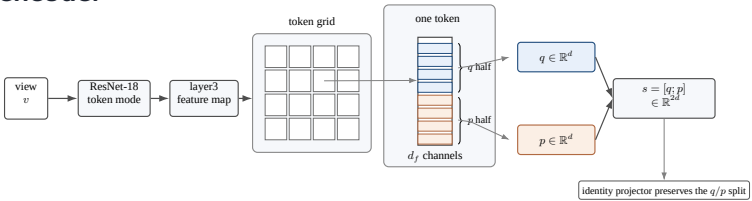
$$\mathcal{H}_\phi(q,p) = \tfrac{1}{2}\|p\|^2 + V_\phi(q).$$

Symplectic, reversible leapfrog

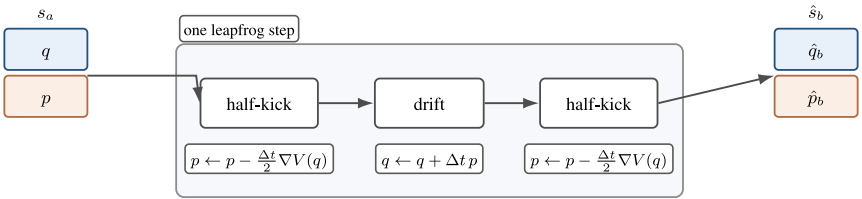
$$(D\Phi)^\top J (D\Phi) = J, \quad \det D\Phi = 1, \quad J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

The encoder and the leapfrog step

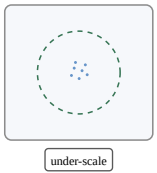
Phase-space encoder

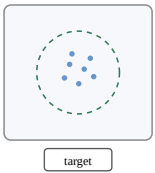


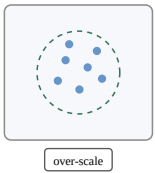
Leapfrog: $K = 2$ steps with $\Delta t = 0.05$

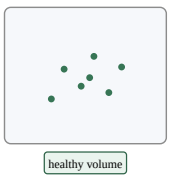


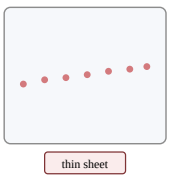
Anti-collapse: scale and volume

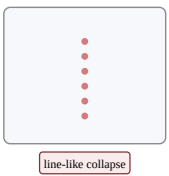












Scale targets

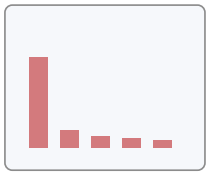
$$\frac{\mathbb{E}\|q\|_2^2}{d_0} \approx \alpha_q, \qquad \frac{\mathbb{E}\|p\|_2^2}{d_0} \approx \alpha_p.$$

Volume floor

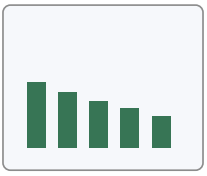
$$\mathcal{L}_{\text{vol}} = \left[\tau - \frac{1}{k}\log\det C\right]_+^2.$$

$C = R^\top \widehat{\Sigma} R + \varepsilon I_k$: centered, projected covariance. $[u]_+ = \max(0, u)$. Soft penalties.

Anti-collapse: spectral guarantees



one dominant spike



variance spread across modes

Participation ratio

$$\mathrm{PR}(C) = \frac{\mathrm{tr}(C)^2}{\mathrm{tr}(C^2)} \geq r_0.$$

Conditional bounds

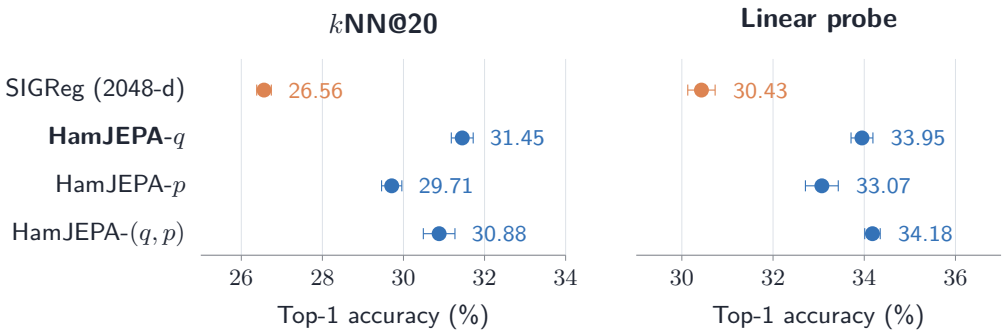
$$C \succ 0, \quad \mathrm{tr}(C) = m, \quad \frac{1}{k} \log\det C \geq \tau, \quad \mathrm{PR}(C) \geq r_0, \quad 1 < r_0 \leq k$$

$$\Longrightarrow \quad \lambda_{\max}(C) \leq \frac{m}{\sqrt{r_0}}, \qquad \lambda_{\min}(C) \geq e^{k\tau} \left(\frac{\sqrt{r_0}}{m} \right)^{k-1}.$$

Conditional on satisfied constraints. Training uses soft penalties on projected, ridged covariances.

CIFAR-100: 30 epochs

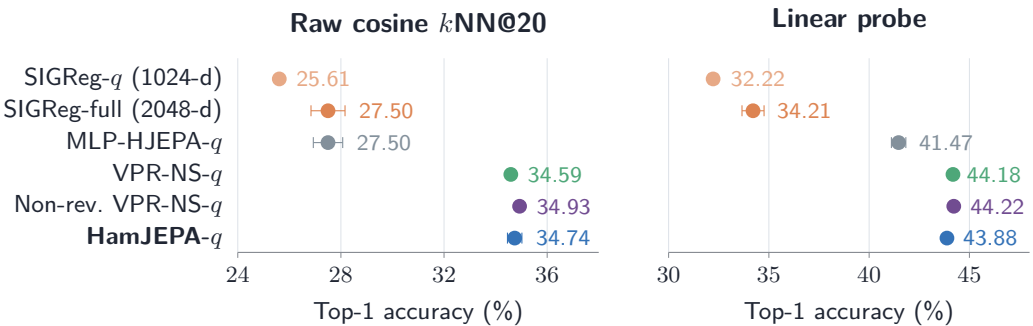
HamJEPA-q: $+4.89$ kNN@20, $+3.52$ probe points over SIGReg.



3 training seeds. Mean ± reported SD. q, p: 1,024-d. Full: 2,048-d.

CIFAR-100: 80 epochs and matched controls

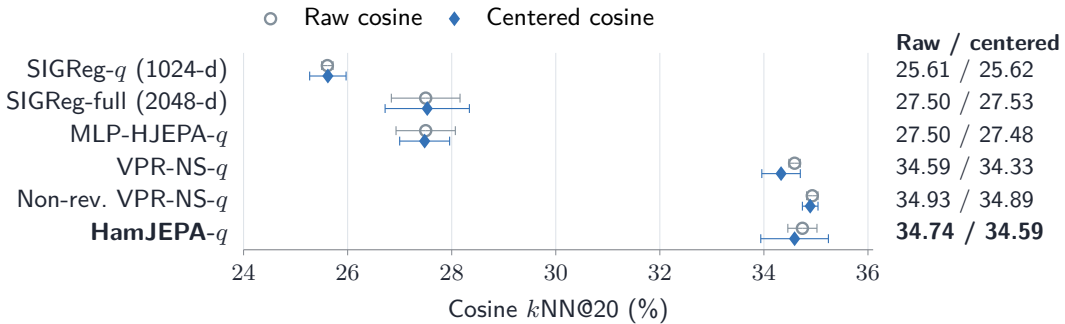
HamJEPA-q vs SIGReg-full: $+7.24$ kNN@20, $+9.67$ probe points.



2 training seeds. Error bars: sample SD. Probe: mean of 5 runs per seed.
Matched non-symplectic and non-reversible controls retain the gain.

CIFAR-100: centering

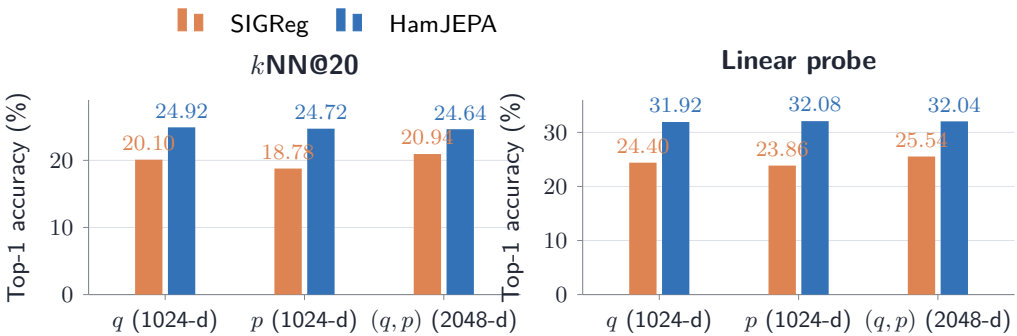
HamJEPA-q vs SIGReg-q: 9.13 raw, 8.97 centered kNN points.



80 epochs. 2 training seeds. Mean ± sample SD. Cosine kNN@20.

ImageNet-100: 45 epochs

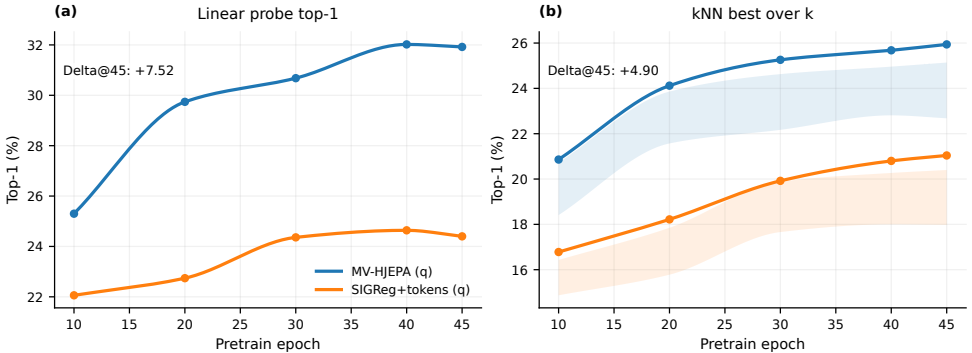
HamJEPA-q vs SIGReg-q: $+4.82$ kNN@20, $+7.52$ probe points.



1 training seed. No seed error bars. SIGReg q, p denote feature blocks.

ImageNet-100: training dynamics

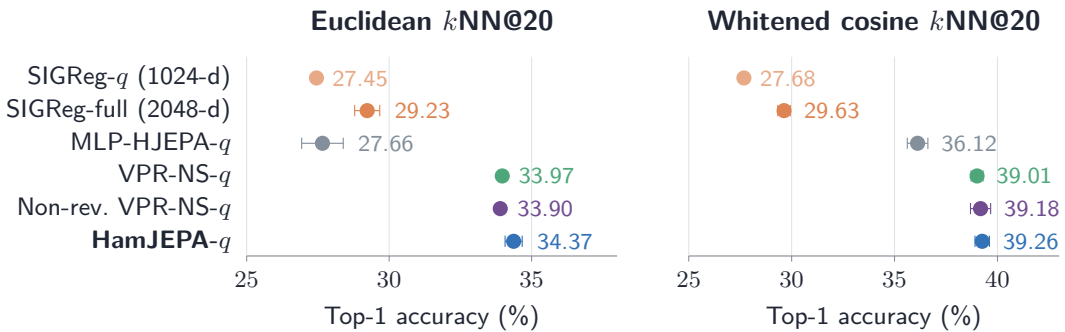
q readout, 1,024 dimensions.



Paper panels (a,b). 1 training seed. Right: best kNN over k. Shading: spread across k. MV-HJEPA = HamJEPA.

CIFAR-100: metric robustness

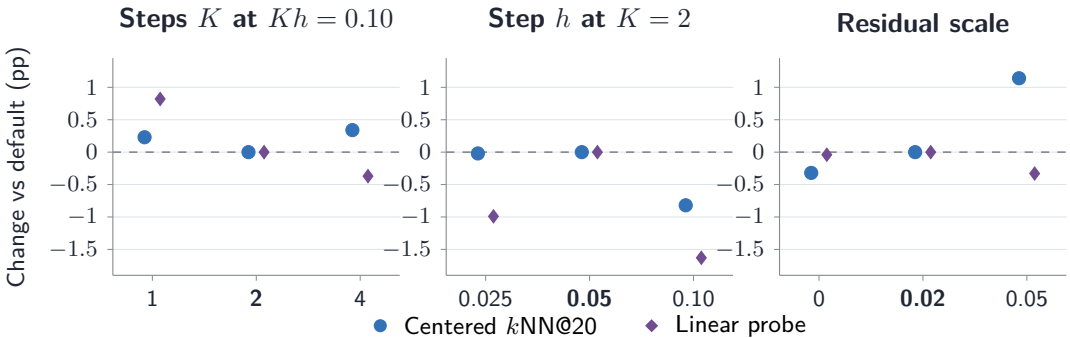
HamJEPA-q remains ahead of SIGReg under both metrics.



80 epochs. 2 training seeds. Mean ± sample SD. kNN@20. Whitening fitted on training features only.

CIFAR-100: rollout and potential sensitivity

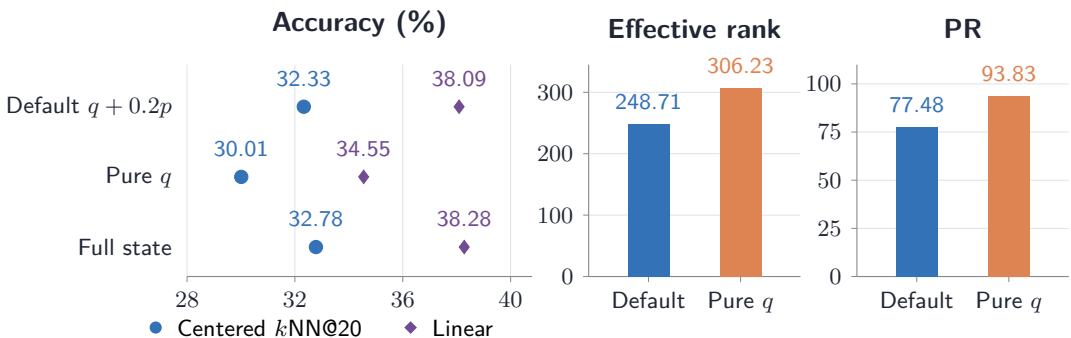
At fixed horizon, $K = 1, 2, 4$ span 0.34 centered-kNN points.



1 training seed. Epoch 30 of the common 80-epoch schedule. Changes relative to default; no seed error bars.

CIFAR-100: auxiliary momentum matching

Removing p matching: -2.32 centered-kNN, -3.54 probe points.



1 training seed. Epoch 30 of the common 80-epoch schedule. Full-spectrum, centered covariance diagnostics.