ORTHOGONALABS · PREPRINT

# Compute by Contraction

*Determinism in Energy-Based Models*

Francis A. Ukpeh

orthogonaLabs

*May 2026*

ABSTRACT

We define a three-term operator on $\mathbb{R}^N$ whose iteration provably converges to a unique fixed point for any input. The operator is state-adaptive: its contraction strength varies with the spectral properties of the current iterate. We derive convergence, uniqueness, stability, and cost-scaling properties from the Banach fixed-point theorem alone.

*Keywords: fixed-point theory · contraction mapping · deterministic computation · energy-based models*

SECTION 1

# Introduction

Modern machine learning systems compute outputs via stochastic iteration: gradient descent on continuous parameters, sampling from learned distributions, or step-wise generation conditioned on training-time priors. The cost of each output is bounded loosely by training scale and inference depth; the output itself carries no analytic guarantee of correctness, repeatability, or termination. When the same input is presented twice, the outputs may differ; when the input is unusual, the system may produce a confident error that is undetectable from the output alone.

This paper studies a different mode of computation. We define a single state-adaptive operator $\mathcal{T}$ on the unit hypersphere $S^{N-1} \subset \mathbb{R}^N$ whose iteration, under a bound derived from the Banach fixed-point theorem [1], converges to a unique fixed point for any input. The operator has three terms — a retention term, an input-injection term, and a state-dependent linear transformation — and the bound that guarantees contraction is a single inequality on the operator norms of those terms. From this inequality, geometric convergence rate, uniqueness, stability, cost-scaling, and compositional properties follow as direct consequences. No training distribution, no sampling, no termination heuristic.

The goal of the paper is to characterize this operator at the level of pure mathematics: to prove that the conditions for global contraction can be stated cleanly, that the resulting fixed-point map admits the standard interpretations of retrieval and generation, and that the convergence rate is determined by the input's distance to the attractor rather than by any external schedule. We do not claim that this operator subsumes existing neural architectures; we claim only that, under the stated bound, computation by deterministic contraction is well-posed, terminates with provable error decay, and yields a fixed-point representation of constant width.

SECTION 2

# Preliminaries

**Definition 1 (Banach contraction).** Let $(X, d)$ be a complete metric space. A map $f : X \to X$ is a *contraction* if there exists $\gamma \in [0, 1)$ such that

$$d(f(x), f(y)) \leq \gamma \cdot d(x, y) \quad \forall x, y \in X \tag{1}$$

**Theorem 1 (Banach, 1922).** Every contraction on a complete metric space has a unique fixed point $z^* = f(z^*)$, obtained as

$$z^* = \lim_{n \to \infty} f^n(x_0) \quad \text{for any } x_0 \in X \tag{2}$$

with geometric convergence $d(f^n(x_0), z^*) \leq \frac{\gamma^n}{1-\gamma} \cdot d(x_0, f(x_0))$.

**Definition 2 (Spectral norm).** For $M \in \mathbb{R}^{N \times N}$, the spectral norm is $\sigma_{\max}(M) = \max_{\|v\|=1} \|Mv\|$.

**Definition 3 (State space).** Let $X = S^{N-1} \subset \mathbb{R}^N$ be the unit hypersphere with geodesic metric $d(x, y) = \arccos(\langle x, y\rangle)$. Equip $X$ with the subspace topology of $\mathbb{R}^N$ and project all iterates back to $S^{N-1}$ via $\pi(z) = z / \|z\|$.

SECTION 3

## The Operator

**Definition 4.** Let $\alpha, \beta \in \mathbb{R}_{\geq 0}$ be scalar weights, $H : \{0, ..., 255\}^* \to \mathbb{R}^N$ a fixed (trained) map from byte sequences to $\mathbb{R}^N$, and $K : \mathbb{R}^N \to \mathbb{R}^{N \times N}$ a matrix-valued function. Define

$$\mathcal{T}(z) = \pi(\alpha z + \beta \cdot H(b) + K(z) \cdot z) \tag{3}$$

where $b \in \{0, ..., 255\}^*$ is the input byte sequence and $\pi$ is the $\ell^2$ projection onto $S^{N-1}$.

The three terms:

- $\alpha z$ — *retention*: scaled carry of the current state.
- $\beta \cdot H(b)$ — *injection*: projection of raw input into the state space.
- $K(z) \cdot z$ — *adaptation*: state-dependent linear transformation.

SECTION 4

## Contraction Guarantee

**Theorem 2 (Main result).** If

$$\beta \cdot \inf_b \|H(b)\| > 2\alpha + 2 \sup_{z \in S^{N-1}} \sigma_{\max}(K(z)) + L_K \tag{4}$$

where $L_K$ is the Lipschitz constant of $K(z)$ under the operator norm, then $\mathcal{T}$ is a strict global contraction on $S^{N-1}$ with respect to the geodesic metric, and has a unique fixed point $z^* = \lim_{n\to\infty} \mathcal{T}^n(z_0)$ for any starting point $z_0 \in S^{N-1}$.

*Proof.* Let $z \in S^{N-1}$ and $w \in T_z S^{N-1}$ be a unit tangent vector (so $\|w\| = 1$ and $\langle z, w\rangle = 0$). Let $v(z) = \alpha z + \beta H(b) + K(z) z$. The derivative of $\mathcal{T}(z) = \pi(v(z))$ in the direction $w$ is:

$$D\mathcal{T}(z) \cdot w = D\pi(v(z)) \cdot Dv(z) \cdot w \tag{5}$$

Using the derivative of the projection map $D\pi(x) = \frac{1}{\|}x\| \left(I - \frac{xx^T}{\|}x\|^2\right)$, we get:

$$D\mathcal{T}(z) \cdot w = \frac{1}{\|}v(z)\| \left(Dv(z) \cdot w - \langle Dv(z) \cdot w, \mathcal{T}(z)\rangle \mathcal{T}(z)\right) \tag{6}$$

This is the projection of $Dv(z) \cdot w$ onto the subspace orthogonal to $v(z)$, scaled by $1/ \|v(z)\|$. Taking the norm:

$$\|D\mathcal{T}(z) \cdot w\| \leq \frac{\|Dv(z) \cdot w\|}{\|} v(z)\| \tag{7}$$

Bound the numerator and denominator:

1. **Numerator.** Since $w$ is a unit tangent vector,

$$Dv(z) \cdot w = \alpha w + K(z)w + (DK(z) \cdot w)z \tag{8}$$

   Applying the triangle inequality and noting $\|z\| = 1$, $\|w\| = 1$:

$$\|Dv(z) \cdot w\| \leq \alpha + \sigma_{\max}(K(z)) + L_K \leq \alpha + \sup_z \sigma_{\max}(K(z)) + L_K \tag{9}$$

2. **Denominator.**

$$\|v(z)\| \geq \beta \; \|H(b)\| - \|\alpha z + K(z)z\| \geq \beta \inf_b \|H(b)\| - \alpha - \sup_z \sigma_{\max}(K(z)) \tag{10}$$

Combining these, the operator norm of the derivative is bounded by:

$$\|D\mathcal{T}(z)\|_{\mathsf{op}} \leq \frac{\alpha + \sup_z \sigma_{\max(K(z))} + L_K}{\beta \inf_b \|H(b)\| - \alpha - \sup_z \sigma_{\max(K(z))}} \tag{11}$$

For $\mathcal{T}$ to be a contraction, we require $\|D\mathcal{T}(z)\|_{\mathsf{op}} \leq \gamma < 1$ for all $z$, which holds if

$$\alpha + \sup_z \sigma_{\max(K(z))} + L_K < \beta \inf_b \|H(b)\| - \alpha - \sup_z \sigma_{\max(K(z))} \tag{12}$$

which simplifies directly to the condition in Equation 4. Under this condition, $\mathcal{T}$ is a contraction with constant $\gamma < 1$ under the geodesic metric, and the Banach Fixed-Point Theorem guarantees the existence and uniqueness of $z^*$. □

**Remark 1 (Uniqueness and the repeller point).** If condition Equation 4 is violated, the mapping can possess multiple fixed points. For instance, when $K(z) = 0$ and $\beta \; \|H(b)\| < \alpha$, both $+H(b)$ and $-H(b)$ are fixed points. However, $-H(b)$ acts as an unstable repeller since its local derivative norm exceeds 1:

$$\|D\mathcal{T}(-H_b)\|_{\mathsf{op}} = \frac{\alpha}{\alpha - \beta \; \|H_b\|} > 1 \tag{13}$$

Enforcing Equation 4 guarantees that $v(z)$ is bounded away from zero for all $z$, eliminating this unstable fixed point and ensuring $+H_b$ is the unique global attractor.

SECTION 5

## Properties

All of the following are direct consequences of Theorem 2.

### 5.1 · Geometric convergence rate

From Banach, the $n$-th iterate satisfies

$$d(\mathcal{T}^n(z_0), z^*) \leq \frac{\gamma^n}{1-\gamma} \cdot d(z_0, \mathcal{T}(z_0)) \tag{14}$$

where the convergence constant is bounded by

$$\gamma = \frac{\alpha + \sup_z \sigma_{\max(K(z))} + L_K}{\beta \inf_b \|H(b)\| - \alpha - \sup_z \sigma_{\max(K(z))}} < 1 \tag{15}$$

The number of iterations to reach $\varepsilon$-convergence is $O(\log(1/\varepsilon))$.

### 5.2 · State-adaptive contraction strength

$K(z)$ varies across $S^{N-1}$, meaning the local contraction rate $\gamma(z) = \left(\alpha + \sigma_{\max(K(z))} + L_K\right) / \|v(z)\|$ is state-dependent. Regions with low $\sigma_{\max(K(z))}$ contract rapidly to stabilize local structures, while critical regions allow slow contraction for representation routing.

### 5.3 · Phase classification

The spectral norm of the local Jacobian $J(z) = \partial \mathcal{T} / \partial z$ classifies the local dynamics:

| Phase | $\sigma_{\max}(J)$ | Behavior |
|---|---|---|
| Condensed | $< 0.7$ | Rapid contraction, stable representation, robust to noise |
| Critical | $\approx 1.0$ | Marginal stability, maximal information capacity |
| Gaseous | $> 1.0$ | Local expansion (locally violates contraction — dynamics are repelled to stable regions) |

### 5.4 · Inverse cost scaling

**Proposition 1.** Let $V(z) = \left\{v \in S^{N-1} : d(v, z^*) > \varepsilon\right\}$ be the unconverged volume at tolerance $\varepsilon$. The number of iterations to convergence is bounded by

$$n(\varepsilon) \le C \cdot \log(1/\ \mathrm{vol}(V(z))) \tag{16}$$

for a constant $C$ depending on $\gamma$. More constrained inputs require fewer iterations, meaning computational cost decreases with specificity.

*Proof.* A more constrained input starts closer to $z^*$ (smaller initial distance $d_0$). The contraction bound is monotone decreasing in $d_0$. □

### 5.5 · Multi-depth loss

At iteration depth $d$, define the loss $\mathcal{L}_d = \ell\left(\mathcal{T}^d(z_0), y\right)$ for target $y$. The total loss is

$$\mathcal{L} = \sum_{d=1}^{D} w_d \cdot \mathcal{L}_d \tag{17}$$

with learnable weights $w_d \ge 0$, $\sum w_d = 1$. Gradients from every depth simultaneously inform the parameters of $K$ and $H$.

### 5.6 · Fixed-width representation

The fixed point $z^* \in S^{N-1} \subset \mathbb{R}^N$ has exactly $N$ coordinates regardless of input size or modality, creating a resolution-agnostic embedding space of $4N$ bytes (for float32).

### 5.7 · Uniqueness under composition

**Proposition 2.** Let $\mathcal{T}_1, \mathcal{T}_2$ be two contraction operators satisfying Equation 4 with constants $\gamma_1, \gamma_2 < 1$. Then $\mathcal{T}_2 \circ \mathcal{T}_1$ is a contraction with constant $\gamma_1\gamma_2 < 1$ and has a unique fixed point.

### 5.8 · Retrieval via fixed-point distance

Given a database of fixed points $\{z_i^*\}$, retrieval of the nearest neighbor to a query $q^*$ is

$$\mathrm{nn}(q^*) = \arg\min_i d(q^*, z_i^*) \tag{18}$$

where $d$ is the geodesic metric on $S^{N-1}$. Since fixed points are unique, equality $z_i^* = z_j^*$ implies identical inputs (up to the kernel of $H$).

### 5.9 · Generation as inverse iteration

The inflation operator $\mathcal{T}^{-1}$ reconstructs an output sequence $\hat{b}$ from a fixed point $z^*$. Because $z^*$ is unique, the reconstruction is determined by the fixed point alone.

SECTION 6

## The Complete Operator

$$z^* = \lim_{n\to\infty} \mathcal{T}^n(z_0)$$

*where*

$$\mathcal{T}(z) = \pi\left(\underbrace{\alpha z}_{\text{retention}} + \underbrace{\beta \cdot H(b)}_{\text{injection}} + \underbrace{K(z) \cdot z}_{\text{adaptation}}\right)$$

$\mathcal{T} : S^{N-1} \to S^{N-1}$ — state-adaptive contraction operator
$H : \{0, ..., 255\}^* \to \mathbb{R}^N$ — byte-to-structure projection (fixed, trained)
$K : \mathbb{R}^N \to \mathbb{R}^{N\times N}$ — state-dependent linear map (Lipschitz constant $L_K$)
$\beta \cdot \inf_b \|H(b)\| > 2\alpha + 2\sup_z \sigma_{\max(K(z))} + L_K$ — contraction guarantee (Banach)

*The operator has three terms. The bound has one inequality. Convergence, uniqueness, stability, retrieval, generation, cost scaling, and composability are all consequences of $\gamma < 1$.*

SECTION 7

## Discussion and Open Questions

The contraction guarantee of Theorem 2 is stated in terms of three quantities held by the operator itself: the retention scalar $\alpha$, the injection scaling $\beta \cdot \inf_b \|H(b)\|$, and the state-dependent norms $\sigma_{\max}(K(z))$ and $L_K$. The inequality is sharp: when it holds strictly, $\mathcal{T}$ is a global contraction with constant $\gamma < 1$ and the fixed point is unique; when it is violated, the system can admit a repeller in addition to its attractor, and the iteration is no longer guaranteed to be globally well-posed (Remark 1).

The state-adaptive structure of $K(z)$ is what distinguishes this construction from a fixed linear contraction. By allowing $\sigma_{\max}(K(z))$ to vary across $S^{N-1}$, the local contraction rate $\gamma(z)$ adapts to the geometry of the current iterate, contracting rapidly in regions where the linear part is small and admitting slow drift in regions where representational routing is required. The cost-scaling proposition (Proposition 1) connects this geometry to computational cost: more constrained inputs begin closer to $z^*$ and converge in fewer iterations, so the iteration count itself reports on how "specified" the input is.

Three open directions follow naturally from the framework. First, the form of $H(b)$ — here taken as a fixed map from byte sequences to $\mathbb{R}^N$ — admits any encoder whose output norm is bounded below by the injection-scaling requirement; characterizing the family of admissible encoders is left open. Second, the family of state-dependent maps $K(z)$ satisfying the Lipschitz bound on $S^{N-1}$ has not been classified; explicit constructions beyond the trivial $K \equiv 0$ case would refine the achievable contraction rates. Third, the composition rule (Proposition 2) suggests that hierarchies of contractions with progressively tighter bounds can be built to layer specificity; whether such hierarchies admit a natural depth-vs-bound trade-off is unresolved.

**SECTION 8**

## Conclusion

We have defined a single state-adaptive operator on the unit hypersphere whose iteration converges to a unique fixed point under a contraction bound that follows directly from the Banach fixed-point theorem. The operator has three terms and the bound is a single inequality. From this one inequality, geometric convergence, uniqueness, stability under composition, fixed-width representation, and input-specificity cost-scaling all follow as direct consequences. The construction admits the standard interpretations of retrieval (via geodesic distance between fixed points) and generation (via inverse iteration from a fixed point), without recourse to stochastic mechanisms or termination heuristics. The result is a well-posed model of deterministic computation by contraction.

**REFERENCES**

[1] **Banach, S.** (1922). *Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales.* Fundamenta Mathematicae 3(1), 133–181.

[2] **Brouwer, L. E. J.** (1911). *Über Abbildung von Mannigfaltigkeiten.* Mathematische Annalen 71(1), 97–115.

[3] **Hutchinson, J. E.** (1981). *Fractals and Self-Similarity.* Indiana University Mathematics Journal 30(5), 713–747.

[4] **Granas, A. & Dugundji, J.** (2003). *Fixed Point Theory.* Springer Monographs in Mathematics. Springer-Verlag, New York.

***Acknowledgments.*** *The author thanks the orthogonaLabs collaborators for the iterative mathematical discussion that sharpened the contraction bound, and the anonymous reviewers whose forthcoming comments will doubtless improve the manuscript further.*

Correspondence: auto@dydact.io · orthogonaLabs · lab.dydact.io