

The X-Factor

Deterministic Manifold Embedding for Native LLM Inference

Klenio Araujo Padilha — Winnex Brasil Soluções Empresariais LTDA-ME
August 2026 — BSL 1.1 — Supplement to Zenodo 10.5281/zenodo.21861809

Abstract

The X-Factor is a deterministic, training-free operator that embeds arbitrary text into the latent space of a pretrained language model, enabling native inference without fine-tuning. It is computed structurally from a model's embedding matrix — never by running the model. This document gives the complete mathematics.

1. Problem Statement

A pretrained language model has an embedding matrix $E \in \mathbb{R}^{V \times D}$ mapping V tokens to D -dimensional hidden vectors. To perform inference, text must be mapped to a sequence of hidden states in $\mathbb{R}^D$. The X-Factor constructs a deterministic projection from an arbitrary text representation into the model's latent space, using only the geometry of E .

2. The Manifold Hypothesis

The V rows of E do not fill $\mathbb{R}^D$; they concentrate near a lower-dimensional structure. Assume the token embeddings lie on a Riemannian manifold M embedded in $\mathbb{R}^D$. The X-Factor captures the linearized tangent structure of M — the principal subspace where embeddings concentrate.

2.1 Centering

The mean embedding is $\mu = (1/V) \sum_{i=1..V} E_i \in \mathbb{R}^D$.

The centered matrix is $E = E - \mu \cdot \mathbf{1}^T \in \mathbb{R}^{V \times D}$.

2.2 Covariance

The empirical covariance: $C = (1/V) E^T E \in \mathbb{R}^{D \times D}$, symmetric positive semi-definite.

3. Eigendecomposition

3.1 Spectral decomposition

C admits an orthonormal eigendecomposition $C = U \Lambda U^T$ where $U = [u_1, u_2, \dots, u_D] \in \mathbb{R}^{D \times D}$ with $U^T U = I_D$, and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_D)$ with $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_D \geq 0$.

3.2 Computational method (power iteration with deflation)

The full eigendecomposition is $O(D^3)$. Following the Madhava principle — bound-guided pruning without exhaustive computation — the top- r eigenpairs are computed by power

iteration with deflation:

```
initialize  $v_j$  (deterministic, unit norm)
repeat:  $w = C \cdot v_j$ ;  $\lambda_j = v_j^T \cdot C \cdot v_j$  (Rayleigh quotient);  $v_j = w / \|w\|$ 
until  $|\Delta \lambda_j| < \varepsilon$ 
deflate:  $C \leftarrow C - \lambda_j \cdot v_j v_j^T$ 
```

Cost $O(D^2 \cdot r \cdot \text{iters})$ versus $O(D^3)$ — $O(D/r)$ faster. The Rayleigh quotient plays the same role as the Cauchy-Schwarz bound in the Madhava engine: a provable selection criterion.

4. The Effective Rank

The effective rank at variance fraction τ is the smallest r such that $\sum_{k=1..r} \lambda_k \geq \tau \cdot \sum_{k=1..D} \lambda_k$, with $\sum \lambda_k = \text{tr}(C)$. The X-Factor uses $\tau = 0.95$.

5. The X-Factor Operator

5.1 The basis

The X-Factor is $X = [u_1, u_2, \dots, u_r] \in \mathbb{R}^{D \times r}$, the top- r principal directions.

5.2 The projector

The orthogonal projection onto the manifold subspace is $P = XX^T \in \mathbb{R}^{D \times D}$.

$P^2 = P$ (idempotent) and $P^T = P$ (symmetric).

For any $h \in \mathbb{R}^D$, $\|h - Ph\|$ is the distance to the manifold subspace, and Ph is the best r -rank approximation of h in the least-squares sense (Eckart-Young).

6. Text Embedding via the X-Factor

6.1 Spectral tokenization (character histogram)

A text $T = c_1 c_2 \dots c_n$ maps to a deterministic feature vector, the character histogram:

$$h_j(T) = (1/\| \cdot \|) \cdot \sum_{i=1..n} 1[c_i \equiv j \pmod{B}], \quad j = 0..B-1$$

normalized to unit L^2 norm. The histogram is chosen over a raw spectral average because the latter collapses (pairwise cosine between texts ≈ 0.99); the histogram separates texts (cosine spread ≈ 0.13 - 0.60).

6.2 Dimensionality expansion

The B -dimensional histogram is embedded into $\mathbb{R}^D$ by zero-padding and L^2 normalization:

$$\psi = [h_0, \dots, h_{B-1}, 0, \dots, 0] / \| \cdot \|$$

6.3 Projection

The hidden state is $h = P\psi = XX^T\psi \in \mathbb{R}^D$, lying exactly on the model's embedding manifold.

7. The Complete Operator

$$\Phi_X(T) = X \cdot X^T \cdot \text{pad_L2}(\text{histogram}(T))$$

Training-free: X depends only on E (a structural read). Deterministic: same text $\rightarrow$ same h .
Manifold-respecting: $\Phi_X(T) \in \text{span}(X)$.

8. Complexity

Covariance C : $O(VD^2)$ or $O(SD^2)$ sampled. Power iteration: $O(D^2r \cdot \text{iters})$. Projector P : $O(D^2r)$.
Embedding a text: $O(n + Dr)$. For $D=2048$, $r \approx 1700$, $S=10^4$: computed in ~ 73 s on one CPU.

9. Relation to the Madhava Engine

The X-Factor follows the Madhava design philosophy: bound-guided selection (the Rayleigh quotient as a Cauchy-Schwarz-style bound) instead of exhaustive computation; deterministic, provable operators; and structural reads of the model — no forward pass, no training.