Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

MAT-3: A Bounded Averaged-Projection Operator for Pre-Inference Cognitive-Affective Guidance

Tiago Aguioncio Vieira

tiago.avieira@usp.br

Universidade de São Paulo

Article

Keywords: affective computing, averaged projections, pre-inference intervention, representation geometry, Lyapunov stability, safe abstention, large language models, MDEI

Posted Date: September 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10626455/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

MAT-3: A Bounded Averaged-Projection Operator for Pre-Inference Cognitive-Affective Guidance

Tiago Aguioncio Vieira (Corresponding Author)
Universidade de São Paulo, São Paulo, Brazil
Correspondence: tiago.avieira@usp.br

Abstract

Language-model affective behaviour is commonly influenced through weight adaptation, prompting, or direct manipulation of internal activations. We formulate MAT-3 as an external pre-inference operator that instead constructs a complementary vector while preserving the original input and embedding. Local covariance geometry defines an affine set, and MAT-3 is an averaged projection toward that set with an intervention radius ε. Under frozen geometry, the operator is non-expansive in the full space and strictly contractive in transverse directions, with the exact rate $\mathrm{dist}(M_3^n(z), A_z) = (1-\eta)^n \mathrm{dist}(z, A_z)$. A sufficient perturbation bound extends transverse contraction to adaptive geometry when $\varepsilon_adapt < \eta$. The intervention is selectively withheld when neighbourhood size, conditioning, eigengap, out-of-distribution distance, or other declared validity conditions fail, yielding safe abstention z_aux = z. A coercive Lyapunov condition connects the MDEI state dynamics to bounded states and absorbing sets. Synthetic tests validate the algebra, implementation, covariance convention, abstention logic, and numerical identities. They do not establish downstream language-model benefit. Evaluation on frozen LLMs remains the principal empirical limitation and the next falsifiable step.

Keywords: affective computing; averaged projections; pre-inference intervention; representation geometry; Lyapunov stability; safe abstention; large language models; MDEI

1. Introduction

Affective computing has represented emotion using categorical labels and low-dimensional continuous coordinates. The circumplex and pleasure–arousal–dominance traditions provide influential examples of the latter approach.[1,2] These psychological models are references for dimensional representation, not identities imposed on the computational state used here. MDEI (Modelo de Dinâmica de Estados Internos; Internal State Dynamics Model) treats cognitive-affective descriptions as continuous state trajectories whose stability, perturbation and geometry can be studied with dynamical-systems tools. This programme follows the broader objective of affective computing: to give computational systems explicit, testable representations of affect-related information without equating representation with subjective experience.[3,4]

Recent work strengthens two parts of this motivation. Trajectory-based evaluation shows that single-turn labels can miss temporal changes in emotional-support dialogue.[5] Mechanistic analysis has also identified geometrically organised emotion-concept directions in a large language model and demonstrated causal effects under internal activation steering, while cautioning against interpreting such representations as persistent subjective states.[6] These results motivate geometric and temporal tests; they do not prove the MDEI assumptions or the benefit of MAT-3.

The present work addresses a narrower question: can a local geometric correction be defined outside a target model's weights, bounded explicitly, audited before inference, and withheld when its geometric basis is unreliable? MAT-3 answers this question at the operator level. Given an original input x and an auxiliary embedding z, the method estimates a local affine geometry and produces a complementary vector z_aux. The original x and z are preserved. The target interface receives an object equivalent to I(x) = (x, z, z_aux), subject to the capabilities of the chosen model interface.

The mathematical contribution follows one chain:

local geometry → averaged projection → bounded intervention → transverse contraction → adaptive stability condition → safe abstention.

This chain separates what is proved from what remains empirical. The paper proves boundedness and exact transverse convergence under frozen geometry, gives a sufficient adaptive perturbation condition, establishes local semantic and angular guarantees under explicit assumptions, and implements abstention tests. It does not report downstream LLM accuracy, emotional competence, safety improvement, reasoning improvement, or computational savings.

Earlier MDEI publications introduced the state representation, Lyapunov-based stability programme, and latent-space dynamical perspective.[7,8,9,10] They establish provenance, not independent validation of the present operator. MAT-4 is outside the scope of this article and appears only as future work.

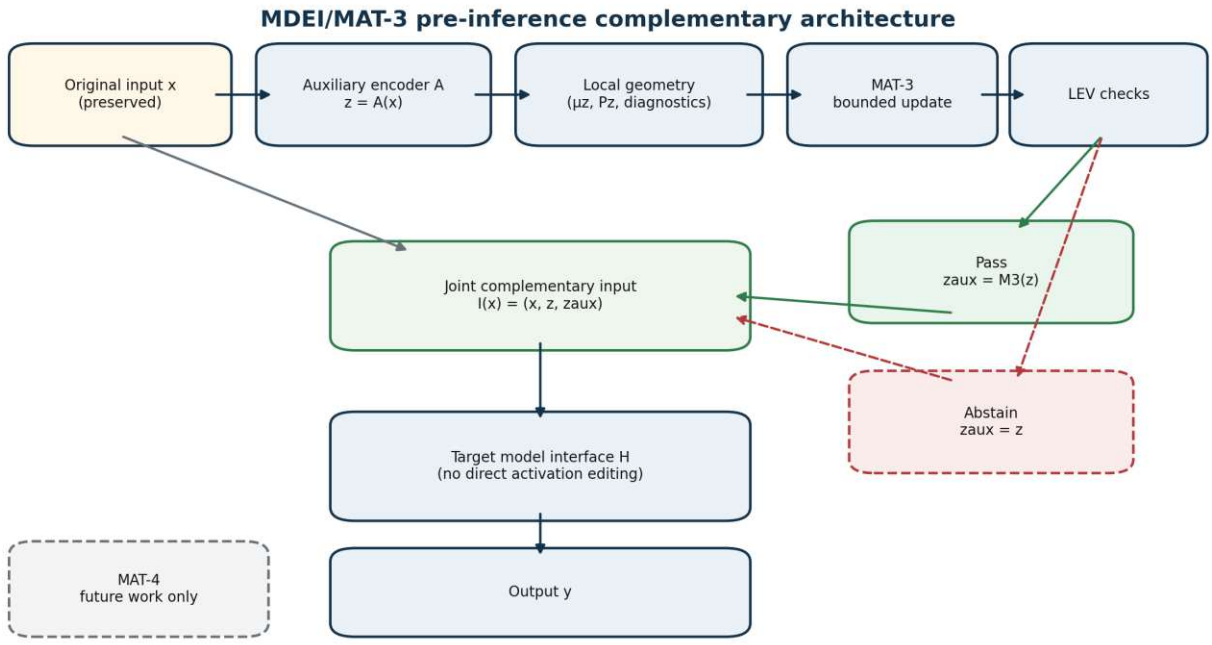


Figure 1 | MDEI/MAT-3 pre-inference complementary architecture. The original input x is preserved. An auxiliary encoder constructs z, local geometry supplies the local mean and tangent projector, and MAT-3 produces a bounded complementary vector only when all declared validity checks pass. If a check fails, the method abstains and returns the unchanged auxiliary vector. The interface H consumes the joint complementary input without direct editing of the target model's internal activations. MAT-4 is shown only as an out-of-scope future extension.

2. MDEI state and model interface

2.1 Causal chain

Let x denote the original input. A deterministic auxiliary encoder A maps x to an embedding z ∈ R^d,

$$x \xrightarrow{A} z \xrightarrow{E} u \qquad (1)$$

The map E: R^d → R^3 produces an MDEI state

$$u = (c, \iota, \tau)^{\mathsf{T}} \in \mathcal{M} \subset \mathbb{R}^3 \qquad (2)$$

where c is a cognitive-valence coordinate, ι is an intensity coordinate and τ is a cognitive-tension coordinate. These coordinates are modelling variables, not direct psychological measurements. The finite-dimensional space R^3 is sufficient for every result in this paper. Extensions to Hilbert spaces are future work.

The MDEI state has an operational control role. A task context T and state u may calibrate the permitted radius and the nominal intervention weight,

$$\varepsilon = \varepsilon(u, \mathcal{T}), \qquad \lambda_{\text{ctrl}} = \lambda_{\text{ctrl}}(u, \mathcal{T}) \qquad (3)$$

subject to validation-set calibration and predeclared bounds. Thus MDEI is not a parallel narrative attached to local PCA: it controls the allowable magnitude and selectivity of the geometric intervention.

2.2 Deterministic dynamics and coercive stability

The deterministic layer is an ordinary differential equation

$$\frac{du}{dt} = F(u, t), \qquad u(0) = u_0 \qquad (4)$$

If F is locally Lipschitz in u and has an appropriate growth bound, the initial-value problem has a unique maximal solution. This continuous-state formulation is compatible with, but does not require, a Neural ODE implementation.[11] Let V: R^3 → [0,∞) be continuously differentiable and coercive, meaning V(u) → ∞ as ||u|| → ∞, equivalently that every sublevel set {u: V(u) ≤ q} is compact. Suppose constants α > 0 and β ≥ 0 satisfy

$$\frac{dV}{dt} = \nabla V(u) \cdot F(u, t) \leq -\alpha V(u) + \beta \qquad (5)$$

The comparison inequality gives

$$V\big(u(t)\big) \leq e^{-\alpha t} V(u_0) + \frac{\beta}{\alpha}(1 - e^{-\alpha t}) \qquad (6)$$

Because V is coercive, bounded V implies bounded u. For any q > β/α, the compact sublevel set B_q = {u: V(u) ≤ q} is absorbing and positively invariant. If the flow is a continuous asymptotically compact semiflow, the standard omega-limit construction yields a compact global attractor.[12] The coercivity hypothesis is essential: an upper bound on a non-proper V alone would not imply a bounded state.

2.3 Stochastic extension

The stochastic layer is stated separately as the Itô equation

$$du_t = b(u_t, t)\, dt + \Sigma(u_t, t)\, dW_t \qquad (7)$$

Global Lipschitz and linear-growth assumptions on b and Σ are sufficient for a unique strong solution. Stability is formulated through the Itô generator L. If a twice continuously differentiable coercive V satisfies

$$\mathcal{L}V(u, t) \leq -\alpha V(u) + \beta \qquad (8)$$

then E[V(u_t)] obeys the analogue of Eq. (6). This expectation bound is distinct from deterministic pathwise stability. No stochastic term is used inside the ODE proofs.

2.4 Complementary LLM interface

The external construction is

$$z \to (P_z, \mu_z) \overset{\text{MAT-3}}{\to} z_{\text{aux}} \qquad (9)$$

The model-side object is

$$\mathcal{J}(x) = (x, z, z_{\text{aux}}) \qquad (10)$$

MAT-3 itself does not retrain target-model weights. A continuous-prefix implementation may nevertheless require an independently trained adapter when the target model does not natively accept continuous vectors. Changing the input naturally changes downstream activations; the claim is therefore that MAT-3 does not directly edit internal activations, not that activations remain numerically unchanged.

For a nonlinear interface G, the relevant response is

$$\Delta h = G(z_{\text{aux}}) - G(z) \qquad (11)$$

or, for an explicitly complementary interface,

$$\Delta h = H(x, [z, z_{\text{aux}}]) - H(x, [z, z]) \qquad (12)$$

A directional test such as cos(Δh,v_e) > ρ is a future empirical diagnostic. The invalid shortcut G(z_aux−z) is not used.

3. Local geometry and the MAT-3 operator

3.1 Neighbourhood and covariance convention

Let N_k(z) = {z_(1),…,z_(k)} be the k nearest reference vectors. Define

$$\mu_z = \frac{1}{k}\sum_{j=1}^{k} z_{(j)} \qquad (13)$$

and use one covariance convention throughout the theory and code: the unbiased sample covariance

$$C_z = \frac{1}{k-1}\sum_{j=1}^{k}\left(z_{(j)} - \mu_z\right)\left(z_{(j)} - \mu_z\right)^{\mathsf{T}} \qquad (14)$$

Regularisation is controlled only by λ_cov,

$$C_z^{\lambda_{\text{cov}}} = C_z + \lambda_{\text{cov}} I, \qquad \lambda_{\text{cov}} > 0 \qquad (15)$$

Let U_r contain the r leading orthonormal eigenvectors and P_z = U_rU_r^T. Local PCA requires an approximately lower-dimensional neighbourhood and a separated spectrum; regularisation ensures invertibility but cannot repair an inadequate sample or vanishing eigengap.[13,14,15]

Define the frozen affine set

$$\mathcal{A}_z = \mu_z + \text{range}(P_z) \qquad (16)$$

with orthogonal projection

$$\Pi_{\mathcal{A}_z}(z) = \mu_z + P_z(z - \mu_z) \qquad (17)$$

3.2 Averaged projection and parameter separation

Write a = z−μ_z and a_perp = (I−P_z)a. The nominal control weight is determined by a separate parameter λ_ctrl,

$$\eta_0 = \frac{1}{1 + \lambda_{\text{ctrl}}} \qquad (18)$$

The effective weight is

$$\eta = \min\left(\eta_0, \frac{\varepsilon}{\| a_\perp \|}\right), \quad \| a_\perp \| > 0 \qquad (19)$$

and η = η_0 when ||a_perp|| = 0. Because the correction is then zero, this guard is safe. MAT-3 is

$$M_3(z) = z_{\text{aux}} = z - \eta a_\perp = (1-\eta)z + \eta \Pi_{\mathcal{A}_z}(z) \qquad (20)$$

Thus λ_cov controls geometric regularisation, whereas λ_ctrl controls intervention cost. They are neither symbolically nor operationally interchangeable.

Equation (20) is the solution of the constrained problem

$$\min_{\delta} \frac{1}{2} \| (I - P_z)(a+\delta) \|^2 + \frac{\lambda_{\text{ctrl}}}{2} \| \delta \|^2 \quad \text{s.t.} \quad \| \delta \| \leq \varepsilon \qquad (21)$$

The unconstrained stationary solution is δ = −η_0a_perp. If its norm exceeds ε, the Karush–Kuhn–Tucker multiplier activates the boundary and gives δ = −εa_perp/||a_perp||. Both cases are exactly Eq. (19).

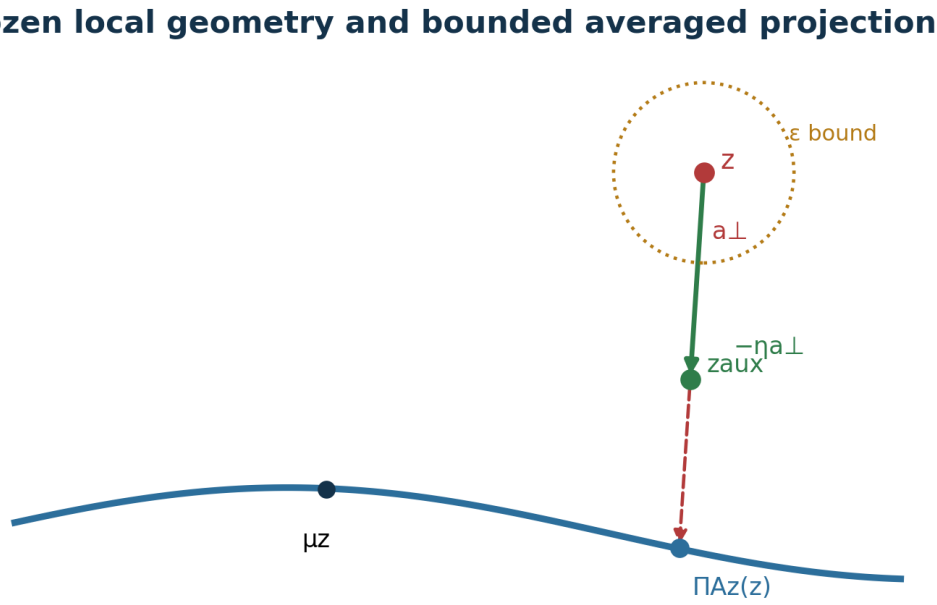


Figure 2 | Frozen local geometry and bounded averaged projection. The local affine set A_z is defined by μ_z and the tangent projector P_z. MAT-3 leaves the tangential component unchanged and reduces only the transverse residual a_perp. The update is limited by the intervention radius ε.

3.3 Theorem 1: bounded intervention

For every z with finite a_perp, Eq. (19) gives

$$\| M_3(z) - z \| = \eta \| a_\perp \| \leq \varepsilon \qquad (22)$$

This guarantee concerns the complementary vector only. It does not imply semantic preservation or downstream improvement.

3.4 Theorem 2: averaged-projection identity

For fixed μ_z, P_z and η ∈ [0,1],

$$\text{dist}(M_3(z), \mathcal{A}_z) = (1-\eta)\text{dist}(z, \mathcal{A}_z) \qquad (23)$$

Proof. Since I−P_z is the projector onto the normal complement,

$$(I - P_z)(M_3(z) - \mu_z) = (1 - \eta)(I - P_z)(z - \mu_z) \qquad (24)$$

Taking norms proves Eq. (23). Repeated application with the same frozen geometry yields the exact rate

$$\mathrm{dist}(M_3^n(z), \mathcal{A}_z) = (1 - \eta)^n \mathrm{dist}(z, \mathcal{A}_z) \qquad (25)$$

3.5 Proposition 1: frozen-geometry non-expansiveness and transverse contraction

When μ_z, P_z and η are frozen,

$$DM_3 = (1 - \eta)I + \eta P_z \qquad (26)$$

Its eigenvalue is 1 on range(P_z) and 1−η on the normal complement. Therefore ρ(DM_3)=1: MAT-3 is non-expansive, not strictly contractive, on the full space. For $0 < \eta \leq 1$ it is strictly contractive in every transverse direction. This distinction prevents a tangential identity direction from being misreported as full-space contraction.

3.6 Theorem 3: sufficient condition under adaptive geometry

When neighbourhoods and projectors vary with z, decompose the derivative as

$$DM_3(z) = DM_3^{\mathrm{frozen}}(z) + R(z), \qquad \| R(z) \| \leq \varepsilon_{\mathrm{adapt}} \qquad (27)$$

For a unit transverse vector v_perp,

$$\| DM_3(z) v_\perp \| \leq \left(1 - \eta + \varepsilon_{\mathrm{adapt}}\right) \| v_\perp \| \qquad (28)$$

Hence ε_adapt < η is sufficient to preserve transverse contraction. This is a conditional local result: the paper does not claim that every k-nearest-neighbour geometry satisfies the bound. A Davis–Kahan perturbation estimate can contribute to bounding projector variation when the eigengap is positive, but neighbourhood-boundary discontinuities and changes in η must also be controlled.[13]

3.7 Conditional angular reduction

Let r_z ∈ range(P_z), p = P_z(z−μ_z), and q = (I−P_z)(z−μ_z). If

$$\langle p, r_z \rangle \geq 0 \qquad (29)$$

then the centred angle does not increase:

$$\angle(M_3(z) - \mu_z, r_z) \leq \angle(z - \mu_z, r_z) \qquad (30)$$

Indeed, the numerator <p,r_z> is unchanged while ||p+(1−η)q|| ≤ ||p+q||. The non-negative numerator makes the cosine non-decreasing. If r_z is not tangential or the acute-alignment condition fails, Eq. (30) is not guaranteed and becomes an empirical check that may trigger abstention.

3.8 Local semantic bound

Let S: R^d → R^m be locally L_S-Lipschitz on a trusted neighbourhood. Then

$$\| S(z_{\mathrm{aux}}) - S(z) \| \leq L_S \| z_{\mathrm{aux}} - z \| \leq L_S \varepsilon \qquad (31)$$

Therefore the sufficient calibration condition

$$\varepsilon \leq \frac{\tau_{\mathrm{sem}}}{L_S} \qquad (32)$$

guarantees a local semantic-distance tolerance τ_sem for that declared S. This is not universal semantic preservation: S, its neighbourhood and L_S must be specified and empirically calibrated.

4. Local validity and safe abstention

Safe abstention is part of MAT-3, not an exceptional recovery path. If a required condition fails,

$$z_{\text{aux}} = z \qquad (33)$$

No correction is forced. Table 1 gives the operational validity set.

Table 1 | MAT-3 local-validity conditions. Every deployment must predeclare thresholds and calibrate them without access to the final test set.

Condition	Requirement	Operational rule
LEV-1	Finite domain	z is finite; the zero-vector case is handled without division.
LEV-2	Sufficient neighbourhood	k ≥ r+1 and the reference set contains at least k+1 vectors.
LEV-3	Conditioning	cond(C_z^(λ_cov)) ≤ κ_max.
LEV-4	Bounded intervention	The correction norm does not exceed ε.
LEV-5	Transverse reduction	Eq. (23) holds in frozen geometry.
LEV-6	Angular condition	Eq. (30) requires tangential reference and acute alignment.
LEV-7	Semantic tolerance	Semantic distance does not exceed τ_sem under the declared local map.
LEV-8	Geometry regularity	λ_r(C_z)−λ_(r+1)(C_z) ≥ γ_min, with controlled neighbourhood and parameter variation.
LEV-9	Safe abstention	Failure of a critical check returns z_aux = z.

4.1 Eigengap and conditioning

The eigengap condition is

$$\gamma_z = \lambda_r(C_z) - \lambda_{r+1}(C_z) \geq \gamma_{\min} \qquad (34)$$

It must be tested in code, not merely calculated. Conditioning and eigengap measure different risks: λ_cov may improve inversion while leaving the leading subspace poorly identified.

4.2 Out-of-distribution abstention

Let d_k(z) = ||z−z_(k)|| be the distance to the k-th nearest reference vector. A calibrated rule requires

$$d_k(z) = \| z - z_{(k)} \| \leq \tau_{\text{OOD}} \qquad (35)$$

If d_k(z) > τ_OOD, the query is outside the trusted local support and MAT-3 returns z_aux = z. Other possible abstention triggers include a semantic-bound failure, a non-finite input, excessive adaptive perturbation, or a deployment-specific turbulence threshold.

5. Relation to intervention methods and current research

MAT-3 is external to target weights, pre-inference, complementary, bounded, local, selective, auditable and capable of abstention. It is geometry- and calibration-dependent. It is not fine-tuning, replacement of the original input, direct activation editing, evidence of emotion or consciousness, a universal model-agnostic algorithm, or a guarantee of improved inference or lower compute.

Table 2 | Intervention locus and evidence boundary. “Training” describes the operator itself; an interface adapter may add a separate training requirement.

Method	Intervention locus	Target weights updated	Internal access	Original input preserved	Native abstention
Fine-tuning	Model weights	Yes	Yes	Not applicable	No
Prompt engineering	Text input	No	No	Yes	No
Soft/prefix prompting	Continuous input	Usually no target-weight update; prompt parameters trained	Interface-dependent	Yes	No
Activation steering	Internal activations	No	Yes	Not applicable	Usually no
GradCuit	Intermediate latent states at test time	No	Yes	Yes	Optimisation-dependent
MAT-3	External complementary vector before inference	No	No for the operator	Yes	Yes

GradCuit optimises instance-specific intermediate latent states at test time while freezing model parameters; it provides a useful contrast because it acts inside a selected Transformer layer and uses outcome gradients, whereas MAT-3 acts before the target forward pass and uses local reference geometry.[16] EmoWorld is not an LLM method, but its decomposition of atmosphere, semantic cues and temporal affect in a frozen video generator is relevant as cross-domain evidence that affective control benefits from separating geometry, semantics and time.[17] Neither work proves MAT-3. They identify future comparison dimensions: intervention locus, temporal evolution, gradient access, boundedness, abstention and downstream utility.

6. Synthetic algebraic validation

6.1 Scope

The validation tests the operator algebra and implementation. It uses no LLM, no emotion-labelled corpus and no downstream task. Consequently, it cannot establish Macro-F1 improvement, emotional competence, semantic safety, reasoning benefit, general model safety or superiority to a baseline.

6.2 Data and parameters

A two-dimensional spiral manifold was embedded in R^10 with isotropic Gaussian noise. The reference set contained n = 2,000 vectors. The run used k = 30, r = 2, λ_cov = 0.001, λ_ctrl = 0.75, ε = 0.06, κ_max = 2×10^5 and γ_min = 10^−4. The OOD threshold τ_OOD = 1.3113578802 was the 99th percentile of within-reference k-th-neighbour distances. Random seed 20260808 was fixed before data generation.

6.3 Main results

Of 2,000 reference queries, 1,967 passed all validity checks and 33 abstained (1.65%): 13 due to the eigengap and 20 due to OOD distance. Among accepted queries, the mean correction was 0.05847184 and the maximum was 0.06000000, equal to ε within floating-point precision. Mean η was 0.44260992, with range 0.23819359 to 0.57142857. The mean observed transverse ratio was 0.55739008, matching the sample average of 1−η. The maximum error in Eq. (23) was 7.08×10^−16. Across ten frozen-geometry iterations, the maximum error in Eq. (25) was 3.89×10^−16.

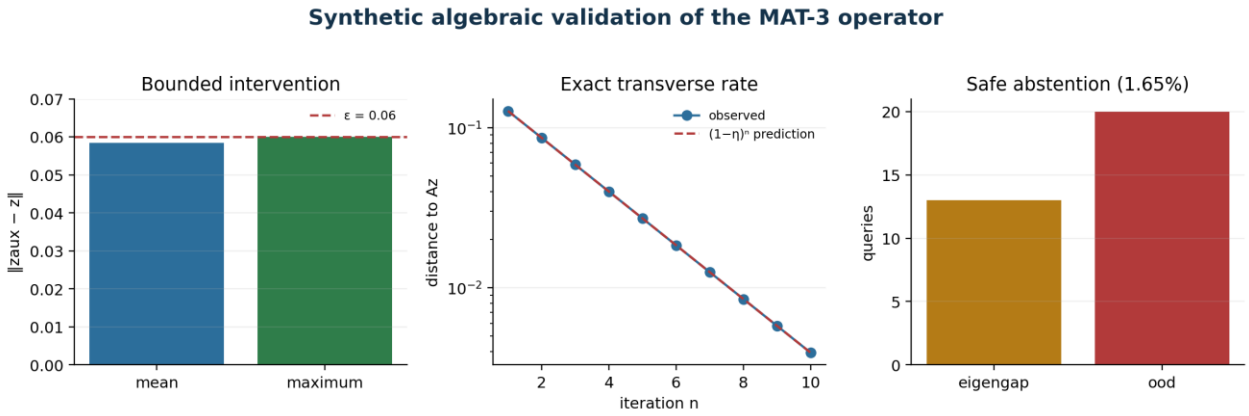


Figure 3 | Synthetic algebraic validation. Left, the intervention remains within ε. Centre, observed transverse distance is indistinguishable from the exact (1−η)^n rate in frozen geometry. Right, the method abstains on declared eigengap and OOD failures. These are implementation checks, not downstream LLM evidence.

6.4 Unit tests

Table 3 | Executed unit tests. Numerical tolerances were 10^−9 unless a Boolean guard was tested.

Test	Result	Interpretation
Projector symmetry	0.00	P_z = P_z^T
Projector idempotency	4.44×10^−16	P_z^2 = P_z
Bounded correction	max 0.06000000000000061	Numerical round-off around ε = 0.06
Averaged-projection identity	max error 7.08×10^−16	Eq. (23)
Iterated (1−η)^n identity	max error 3.89×10^−16	Eq. (25)
Zero-vector guard	Pass	No division by zero
Insufficient-neighbourhood guard	Pass	LEV-2 abstention
Orthogonal equivariance	error 8.38×10^−16	M_3(Qz) = QM_3(z)
λ_cov versus λ_ctrl	Pass	Geometry and intervention roles are distinct
Eigengap abstention	Pass	z_aux = z
OOD abstention	Pass	z_aux = z
Covariance convention	error 2.78×10^−17	Matches ddof = 1
Synthetic Lipschitz bound	positive margin 0.1700	Algebraic bound only
Adaptive sufficient bound	error 1.11×10^−16	Saturates declared operator bound
Routh–Hurwitz check	21/21 agreement	Auxiliary dynamical-system check
Frozen Jacobian	spectral radius 1.0000000000000002	Non-expansive, not full contraction

Orthogonal equivariance follows because Qμ_z, QC_zQ^T and QP_zQ^T are the corresponding quantities after an orthogonal coordinate change. General affine equivariance is not claimed.

7. Future downstream evaluation

The next paper should evaluate a fixed interface on frozen LLMs. Candidate families include Llama, Mistral, Gemma and Qwen, subject to licence and compute constraints. Thresholds and hyperparameters must be selected on a validation split and frozen before test evaluation.

The minimum comparison set is: (i) baseline; (ii) additional textual context; (iii) a random vector matched in norm; (iv) simple local projection; and (v) MAT-3. Fine-tuning may be included as a separate training-based reference. If a learned continuous-prefix adapter is necessary, every condition must use the same adapter architecture and training budget so that the comparison isolates the operator.

Primary outcomes should include Macro-F1, per-emotion precision/recall/F1, balanced accuracy, confusion matrices, calibration and abstention rate. Stress sets should cover opposed emotions, multiple simultaneous emotions, sarcasm, negation and paraphrase robustness. Semantic preservation must be measured independently. Operational measurements should include latency, throughput and memory, without assuming savings in advance. Trajectory measures such as baseline emotional level, emotional trajectory volatility and emotional centroid displacement can evaluate temporal behaviour.[5]

For the complementary interface, report Δh using Eq. (12), measure alignment with preregistered emotion directions where available, and distinguish positive, null and harmful outcomes. A valid conclusion may be that

MAT-3 improves, does not change, or degrades performance. Safe abstention should be evaluated as a coverage–risk curve rather than treated as missing data.

8. Discussion

The operator's strongest result is geometric and deliberately narrow. MAT-3 is an averaged projection onto a locally estimated affine set. Its full-space spectral radius is one because tangential information is preserved; its guaranteed action is transverse. This makes the method interpretable: every nonzero correction can be audited through the neighbourhood, covariance convention, eigengap, projector, ε and η.

The main theoretical vulnerability is adaptive geometry. Nearest-neighbour changes can make P_z discontinuous, and a positive eigengap alone does not eliminate boundary effects. The sufficient condition ε_adapt < η isolates the quantity that future work must estimate rather than declaring the entire adaptive case open. Smoother kernels, neighbourhood margins or differentiable local covariance estimators may make this bound more tractable.

The semantic guarantee is also local. Equation (31) depends on a declared semantic map and Lipschitz constant. Real language semantics are not exhausted by a single embedding metric, so downstream evaluation must use multiple preservation checks and human review where appropriate.

The historical “emotional Reynolds” analogy is not part of the principal proof. If retained operationally, it should be called a drive-to-restoration ratio,

$$R_{\mathrm{DR}} = \frac{\| F_{\mathrm{drive}} \|}{\| F_{\mathrm{restore}} + F_{\mathrm{diss}} \| + \varepsilon_R} \qquad (36)$$

It is dimensionless only after consistent normalisation, depends on a non-unique decomposition, and is not a hydrodynamic quantity. Likewise, a spatial two-component proxy used in synthetic data must not be equated with a full temporal turbulence index. These diagnostics may trigger abstention, but they do not compete with the averaged-projection theorem.

MAT-4 remains outside the present method. No MAT-4 algorithm, experiment or comparative claim is made here.

9. Conclusion

MAT-3 provides a mathematically explicit pre-inference mechanism for constructing a complementary cognitive-affective vector while preserving the original input. The operator separates covariance regularisation from intervention cost, limits every accepted correction by ε, contracts exactly in transverse directions under frozen geometry, admits a sufficient adaptive perturbation condition, and abstains when local geometry is unreliable. Coercive Lyapunov assumptions connect MDEI dynamics to bounded states and absorbing sets, while ODE and SDE claims remain logically separate. Synthetic execution confirms the algebra and implementation. Whether these properties translate into improved language-model behaviour is not established and requires the downstream protocol specified here.

Methods

Synthetic reference geometry

The synthetic manifold was generated as t = (3π/2)(1+2U), U ~ Uniform(0,1), with two-dimensional coordinates (t cos t, t sin t). A random orthonormal basis embedded these coordinates in R^10 and independent Gaussian noise with standard deviation 0.05 was added. Local neighbours were identified with a k-d tree. Exact self-matches were excluded.

Numerical linear algebra

Equation (14) was implemented with divisor k−1. The regularised covariance in Eq. (15) was diagonalised with a symmetric eigensolver. Eigenvalues were ordered from largest to smallest. P_z was constructed from the first r eigenvectors. Condition number, eigengap and OOD distance were tested before the correction. No expression of the form eigs(...)(end) or other non-standard MATLAB syntax was used.

Executed environment

All reported numbers were generated in Python 3.12.13 with NumPy 2.3.5 and SciPy 1.17.0 on a Linux 6.18.35 x86_64 CPU environment. Seed 20260808 was fixed once. No GPU was used. GNU Octave and MATLAB did not generate the reported results.

Algorithm

Input: z, reference set R, k, r, λ_cov, λ_ctrl, ε, κ_max, γ_min, τ_OOD.

1. Reject non-finite inputs; safely return z for a zero vector when required by the interface.
2. Construct N_k(z). If k < r+1 or the reference set is insufficient, abstain.
3. Compute μ_z and C_z with divisor k−1; set C_reg = C_z + λ_cov I.
4. If cond(C_reg) > κ_max, abstain.
5. Compute P_z and γ_z. If γ_z < γ_min, abstain.
6. Compute d_k(z). If d_k(z) > τ_OOD, abstain.
7. Set a_perp = (I−P_z)(z−μ_z), η_0 = 1/(1+λ_ctrl), and η by Eq. (19).
8. Set z_aux = z−ηa_perp.
9. If a declared semantic, angular or adaptive bound fails, abstain; otherwise return z_aux and diagnostics.

Data availability

No external empirical dataset was analysed. All numerical results were generated from synthetic vectors using the procedure stated in Methods and the fixed random seed 20260808.

Code availability

The complete algorithm, parameter values and executed unit-test definitions are reported in this manuscript. The executable Python implementation used to generate the numerical values and figures will be deposited with the article submission or its associated repository.

Funding

This research received no specific external funding.

Acknowledgements

The author thanks colleagues at Universidade de São Paulo and Zenne Tecnologia for discussions that contributed to the development of the research programme.

Author contributions

T.A.V. conceived the framework, developed the mathematical formulation, designed the computational protocol, interpreted the results and wrote the manuscript.

Competing interests

The author declares no competing interests.

References

1. Russell, J. A. A circumplex model of affect. J. Pers. Soc. Psychol. 39, 1161–1178 (1980). doi:10.1037/h0077714.
2. Mehrabian, A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Curr. Psychol. 14, 261–292 (1996). doi:10.1007/BF02686918.
3. Picard, R. W. Affective Computing (MIT Press, Cambridge, Massachusetts, 1997).
4. Scherer, K. R. What are emotions? And how can they be measured? Soc. Sci. Inf. 44, 695–729 (2005). doi:10.1177/0539018405058216.
5. Tan, Z. et al. Detecting emotional dynamic trajectories: an evaluation framework for emotional support in language models. arXiv:2511.09003 (2025).
6. Sofroniew, N. et al. Emotion concepts and their function in a large language model. Transformer Circuits Thread (2026); arXiv:2604.07729.
7. Vieira, T. A. Formalização algébrico-dinâmica aplicada ao modelo MDEI: da tabela vetorial ao sistema cognitivo-afetivo. ARACÊ 7, 40384–40397 (2025). doi:10.56238/arev7n7-300.
8. Vieira, T. A. Fundamentação vetorial e estabilidade emocional computacional: uma arquitetura para IA cognitivo-afetiva baseada em Lyapunov. Contrib. Cienc. Soc. 18, e19608 (2025). doi:10.55905/revconv.18n.7-297.
9. Vieira, T. A. Dinâmica de redes neurais, espaços latentes e suas fundações matemáticas. Cad. Pedagógico 22, e19007 (2025). doi:10.54033/cadpedv22n10-114.
10. Vieira, T. A. Algebraic-dynamic formalization applied to the MDEI model: from the vector table to the cognitive-affective system. In Fundamentals and Advances in Computer Science 2nd edn, 18–30 (Editora Impacto Científico, 2025). doi:10.56238/edimpacto2025.034.
11. Chen, R. T. Q., Rubanova, Y., Bettencourt, J. & Duvenaud, D. K. Neural ordinary differential equations. Adv. Neural Inf. Process. Syst. 31 (2018).
12. Temam, R. Infinite-Dimensional Dynamical Systems in Mechanics and Physics 2nd edn (Springer, New York, 1997). doi:10.1007/978-1-4612-0645-3.
13. Davis, C. & Kahan, W. M. The rotation of eigenvectors by a perturbation. III. SIAM J. Numer. Anal. 7, 1–46 (1970). doi:10.1137/0707001.
14. Belkin, M. & Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. 15, 1373–1396 (2003). doi:10.1162/089976603321780317.
15. Coifman, R. R. & Lafon, S. Diffusion maps. Appl. Comput. Harmon. Anal. 21, 5–30 (2006). doi:10.1016/j.acha.2006.04.006.
16. Yu, Z. et al. GradCuit: credit-assigned gradient flow enables robust and interpretable test-time latent reasoning. arXiv:2608.02585 (2026).
17. Wang, B., Zhyldyzbekov, B., Feng, K. & Wang, Z. EmoWorld: a decoupled affective field for controllable emotional video generation. arXiv:2608.06231 (2026).

Appendix A. Proof details

A.1 Absorbing set

For $q > \beta/\alpha$, Eq. (5) is strictly negative on $V = q$. Thus the sublevel B_q is positively invariant. Equation (6) shows that for every bounded initial set there exists a finite entrance time after which $V(u(t)) \leq q$. Coercivity makes B_q compact in R^3.

A.2 Orthogonal equivariance

For an orthogonal Q, the transformed mean and covariance are $Q\mu_z$ and QC_zQ^T. The leading projector is QP_zQ^T, and the normal residual is $Q(I-P_z)(z-\mu_z)$. Norms, condition numbers and eigengaps are preserved. Hence $M_3(Qz;QR) = QM_3(z;R)$, provided the same thresholds are used.

A.3 Angular condition

With p perpendicular to q and r_z in range(P_z), $\langle q,r_z \rangle = 0$. The numerator of the centred cosine is $\langle p,r_z \rangle$ before and after correction. The denominator changes from $\|p+q\|\|r_z\|$ to $\|p+(1-\eta)q\|\|r_z\|$ and cannot increase. If $\langle p,r_z \rangle \geq 0$, the cosine cannot decrease and the angle cannot increase. The sign condition is necessary for this argument.

A.4 Adaptive bound

For a transverse v_{perp}, $DM_3^{frozen}\, v_{perp} = (1-\eta)v_{perp}$. Subadditivity and $\|R\| \leq \varepsilon_{adapt}$ give Eq. (28). The coefficient is strictly below one when $\varepsilon_{adapt} < \eta$.

Appendix B. Claim taxonomy

Table 4 | Epistemic status of principal statements.

Statement	Status	Evidence in this paper
MDEI state continuity	Modelling axiom	Not empirically established
Coercive Lyapunov boundedness	Theorem under stated assumptions	Proof in Section 2 and Appendix A
Bounded MAT-3 correction	Theorem	Algebra and executed tests
Exact frozen transverse rate	Theorem	Algebra and executed tests
Adaptive transverse contraction	Conditional theorem	Sufficient perturbation bound; universal validity not claimed
Local semantic tolerance	Conditional theorem	Requires declared Lipschitz semantic map
LLM behavioural improvement	Hypothesis	Not tested
Computational saving	Hypothesis	Not tested
MAT-4 operation	Future work	Outside scope

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FUNDINGDECLARATION.pdf