AOBE: ANCHORED ORDER-ISOMORPHIC BI-LIPSCHITZ EMBEDDINGS

Anonymous authors
Paper under double-blind review

ABSTRACT

A representation that faithfully encodes a partially monotonic input order should preserve prescribed comparisons within fixed-context fibers without introducing unsupported comparisons between inputs that are incomparable under that order. Forward partial monotonicity alone does not ensure this: it permits input collapse, comparability between incomparable inputs, and uncontrolled metric distortion. We introduce anchored order-isomorphic bi-Lipschitz embeddings (AOBE), constructed by concatenating a coordinate-wise anchor with bounded positive scales and a norm-controlled partially monotonic network. The anchor retains a positively scaled copy of every direction-aligned constrained coordinate and a signed pair of every context coordinate. For every admissible embedding parameter value, the embedding is an exact order isomorphism from the conditional product order onto its image, is injective, and satisfies parameter-uniform bi-Lipschitz bounds determined by fixed hyperparameters. Across six benchmarks spanning binary classification, regression, and ordinal classification, AOBE achieves competitive predictive performance. In a parameter-matched representation comparison, AOBE is the only evaluated embedding that combines zero audited order-preservation and false-comparability violation rates with positive normalized incomparability margins and a parameter-uniform bi-Lipschitz guarantee. Frozen-channel readouts support task-dependent utility of the nonlinear features, while direct-collapse and task-output audits examine input separation and bounded prediction changes. These results demonstrate that faithful order representation and geometric stability can be enforced while retaining predictive utility.

1 INTRODUCTION

Monotonicity constraints encode qualitative domain knowledge that, after direction alignment, increasing designated features while holding the remaining inputs fixed should not decrease a prediction (Duivesteijn & Feelders, 2008; Zhao et al., 2024; Polo-Molina et al., 2026). In partially monotonic problems, directions are prescribed only for a designated subset of coordinates, while unrestricted context variables may affect the response without a specified direction (Daniels & Velikova, 2010; You et al., 2017; Nolte et al., 2023). Inputs are therefore compared by coordinate-wise increases in the aligned constrained coordinates at fixed context, avoiding unjustified ordering across contexts.

Many monotonic neural architectures enforce only a forward condition at the prediction level: increasing the aligned constrained coordinates at fixed context cannot decrease the output (Nolte et al., 2023; Kim & Lee, 2024; Sartor et al., 2025). For an intermediate representation intended to encode the conditional order, this is insufficient: distinct inputs may collapse, incomparable inputs may become comparable, and distances may be distorted without quantitative control. An upper Lipschitz bound controls expansion but not contraction, injectivity alone does not guarantee a positive global lower bound (Amir & Dym, 2025), and bi-Lipschitzness alone does not imply preservation or reflection of the conditional product order (Behrmann et al., 2019; Kinoshita & Toyoizumi, 2024). We therefore ask whether a learnable map can preserve and reflect this order, remain injective, and control distortion without ordering the context variables.

To address this question, we introduce anchored order-isomorphic bi-Lipschitz embeddings (AOBE), which concatenate a norm-controlled partially monotonic branch with an explicit

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

coordinate-wise anchor. Positively scaled constrained coordinates reflect their product order, while signed context pairs force context equality whenever representations are comparable. The anchor is intentionally information-preserving and provides the lower distance bound; the nonlinear branch provides task-dependent score features, and its norm control, together with bounded anchor scales, yields a uniform upper bound. Thus, every admissible embedding is injective, preserves and reflects the conditional product order, and is parameter-uniformly bi-Lipschitz. A positive, norm-controlled readout produces a conditionally monotone Lipschitz score for binary classification, regression, and ordinal classification. Our experiments examine both faithful input representation and the predictive utility and stability of the resulting scores.

Our main contributions are as follows:

- We introduce an anchored representation that is an exact order isomorphism from the input space, equipped with the conditional product order, onto its image, equipped with the product order. It preserves exactly the prescribed comparisons, introduces no unsupported comparability within or across context fibers, and is injective.
- We establish parameter-uniform lower and upper Euclidean distance bounds for the complete embedding, controlling contraction and expansion through fixed architectural hyperparameters.
- We develop a positive, norm-controlled scalar readout that transfers conditional monotonicity and upper Lipschitz control to a latent score. Task-specific monotone Lipschitz links propagate these guarantees to binary probabilities, original-scale regression predictions, and ordinal upper-tail probabilities.

2 Related Work

Monotonic neural networks. Early architectures guarantee monotonicity in each constrained coordinate through sign-constrained weights with monotone activations or min–max constructions of sign-constrained affine functions (Sill, 1997; Daniels & Velikova, 2010). Calibrated interpolated look-up tables, lattice ensembles, and deep lattice networks impose linear inequalities on their component parameters to obtain flexible partial monotonicity (Gupta et al., 2016; Milani Fard et al., 2016; You et al., 2017). An alternative scalar construction integrates the strictly positive output of an otherwise unrestricted neural network (Wehenkel & Louppe, 2019). Certification and counterexample-guided methods retain more general piecewise-linear networks but rely on verification, regularization, counterexample augmentation, or prediction-time correction procedures (Liu et al., 2020; Sivaraman et al., 2020).

Recent architectures enforce monotonicity by construction: the Lipschitz monotonic network (LMN) adds a monotone linear residual to a base network with a controlled ℓ_1-Lipschitz bound with respect to the constrained inputs (Nolte et al., 2023); the scalable monotonic neural network (SMNN) uses a partially connected architecture whose scalable monotonic hidden layer comprises exponentiated, ReLU, and confluence units (Kim & Lee, 2024); the smooth min–max monotonic network (SMM) replaces hard minimum and maximum operators with strictly increasing smooth approximations (Igel, 2024); and constrained monotonic networks combine a base unsaturated monotone activation with two derived variants under sign-constrained weights, yielding universal approximation (Runje & Shankaranarayana, 2023). At the order-theoretic level, these methods primarily target forward monotonicity of a predictor or scalar map, although several also establish properties such as upper Lipschitz control or universal approximation. AOBE instead targets an intermediate full-input embedding that is an order isomorphism onto its image, preserving and reflecting the conditional product order under the representation-space product order.

Lipschitz and bi-Lipschitz representations. Upper Lipschitz control has been pursued through approximate Parseval tight-frame regularization, spectral normalization, norm-constrained GroupSort networks, projected per-layer norm constraints, Cayley-parameterized orthogonal convolutional layers, and direct parameterizations of networks satisfying global bounds (Cisse et al., 2017; Miyato et al., 2018; Anil et al., 2019; Gouk et al., 2021; Trockman & Kolter, 2021; Wang & Manchester, 2023). A complementary line derives upper bounds for fixed networks through semidefinite relaxations and scalable compositional methods (Fazlyab et al., 2019; Xu & Sivaranjani, 2024). Such

upper control limits expansion but, by itself, neither ensures injectivity nor provides a positive lower Lipschitz bound against collapse. Contractive residual blocks provide invertible, bi-Lipschitz architectures (Behrmann et al., 2019); nevertheless, invertibility alone need not ensure numerically stable inversion (Behrmann et al., 2021), and bi-Lipschitz restrictions can limit the expressivity of normalizing flows (Verine et al., 2023). In a related construction, Kinoshita & Toyoizumi (2024) use convex neural networks and the Legendre–Fenchel transform to construct the bi-Lipschitz neural network (BLNN), whose overall metric bounds can be controlled directly. Replacing the input-convex network with a partially input-convex network (Amos et al., 2017) yields the partially bi-Lipschitz neural network (PBLNN), which restricts the BLNN metric construction to selected variables at each fixed setting of the remaining variables and was evaluated on partially monotone prediction tasks. This fiberwise metric control does not by itself imply preservation or reflection of the conditional product order or provide the full-input order-isomorphism guarantee studied here.

3 METHOD

3.1 PROBLEM SETUP

Let $\mathbf{x} = (\mathbf{x}_M, \mathbf{x}_U) \in \mathbb{R}^d$ comprise a priori designated constrained coordinates $\mathbf{x}_M \in \mathbb{R}^{d_M}$ and unrestricted context coordinates $\mathbf{x}_U \in \mathbb{R}^{d_U}$, where $d_M \geq 1$, $d_U \geq 0$, and $d = d_M + d_U$. We align the prescribed directions of $\mathbf{x}_M$ to be increasing. With $\preceq$ denoting coordinate-wise order, define

$$\mathbf{x} \preceq_M \mathbf{x}' \quad \Longleftrightarrow \quad \mathbf{x}_M \preceq \mathbf{x}'_M \quad \text{and} \quad \mathbf{x}_U = \mathbf{x}'_U. \tag{1}$$

Each fixed context defines a product-ordered fiber, and inputs from distinct contexts are incomparable. This domain-level order does not require repeated observed contexts; when $d_U = 0$, it reduces to the usual product order.

3.2 AOBE CONSTRUCTION AND GUARANTEES

3.2.1 EMBEDDING CONSTRUCTION

Let $D_M = \operatorname{diag}(d_1^M, \ldots, d_{d_M}^M)$ and, when $d_U > 0$, let $D_U = \operatorname{diag}(d_1^U, \ldots, d_{d_U}^U)$. Given fixed constants $0 < \underline{d} < \overline{d}$, every diagonal entry is parameterized as

$$d_j^G = \underline{d} + (\overline{d} - \underline{d})\sigma(\alpha_j^G), \qquad G \in \{M, U\}, \quad j = 1, \ldots, d_G, \tag{2}$$

where the parameters α_j^G are unconstrained and σ is the logistic function, so $\underline{d} < d_j^G < \overline{d}$ for every parameter value. The order anchor is

$$a(\mathbf{x}) = \begin{bmatrix} D_M \mathbf{x}_M \\ D_U \mathbf{x}_U \\ -D_U \mathbf{x}_U \end{bmatrix}, \tag{3}$$

where the last two blocks are omitted when $d_U = 0$. The complete embedding concatenates the anchor with a nonlinear representation h_θ:

$$z_\theta(\mathbf{x}) = \begin{bmatrix} a(\mathbf{x}) \\ h_\theta(\mathbf{x}) \end{bmatrix} \in \mathbb{R}^{d_M + 2d_U + q}. \tag{4}$$

Here θ collects all unconstrained anchor and nonlinear-branch parameters.

The nonlinear branch is an L-layer feedforward network with positive-integer widths $q_0 = d, q_1, \ldots, q_L = q$, where $L \geq 1$. The first-layer columns associated with $\mathbf{x}_M$ are constrained to be nonnegative, whereas those associated with $\mathbf{x}_U$ remain unrestricted:

$$A_0 = \begin{bmatrix} \operatorname{softplus}(V_0^M) & V_0^U \end{bmatrix}, \qquad A_\ell = \operatorname{softplus}(V_\ell), \quad \ell = 1, \ldots, L-1, \tag{5}$$

where $\operatorname{softplus}$ is applied entrywise and V_0^U is omitted when $d_U = 0$. For fixed layer bounds $\gamma_\ell > 0$, define

$$W_\ell = \frac{\gamma_\ell A_\ell}{\max\{\gamma_\ell, \|A_\ell\|_F\}}, \qquad \ell = 0, \ldots, L-1. \tag{6}$$

This positive rescaling preserves the required signs and guarantees $\|W_\ell\|_2 \le \|W_\ell\|_F \le \gamma_\ell$. Frobenius rescaling avoids singular-value estimation, but can be more restrictive than spectral-norm control. Starting from $u^{(0)} = \mathbf{x}$, the hidden states and branch output are

$$u^{(\ell+1)} = \tanh(W_\ell u^{(\ell)} + b_\ell), \qquad h_\theta(\mathbf{x}) = u^{(L)}. \tag{7}$$

The biases are unconstrained, and $\tanh$ is coordinate-wise increasing and 1-Lipschitz. Deeper layers monotonically aggregate first-layer features rather than form an unrestricted context subnetwork; unrestricted first-layer context directions nevertheless permit nonmonotonic context responses. An embedding parameter value is admissible when all raw parameters are finite and the fixed hyperparameters satisfy $0 < \underline{d} < \overline{d}$ and $\gamma_\ell > 0$.

Proposition 1 (Partial monotonicity and Lipschitz continuity)**.** *For every admissible embedding parameter value and all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$, $\mathbf{x} \preceq_M \mathbf{x}'$ implies $h_\theta(\mathbf{x}) \preceq h_\theta(\mathbf{x}')$. Moreover,*

$$\|h_\theta(\mathbf{x}) - h_\theta(\mathbf{x}')\|_2 \le \kappa \|\mathbf{x} - \mathbf{x}'\|_2, \qquad \kappa = \prod_{\ell=0}^{L-1} \gamma_\ell. \tag{8}$$

The proof is given in Appendix A.1.

3.2.2 Order isomorphism

Partial monotonicity of h_θ guarantees only forward order preservation: it need not be injective and may map inputs from distinct contexts to comparable outputs. The positive constrained block reflects the order of $\mathbf{x}_M$, while the signed context pair forces equality of $\mathbf{x}_U$ whenever representations are comparable. For any changed context coordinate, its two signed anchor entries change in opposite directions, making the full representations incomparable regardless of the nonlinear branch.

Within a fixed context, positive diagonal scaling preserves conflicting signs in constrained-coordinate differences. These coordinate conflicts remain present after concatenating the nonlinear branch, preventing unsupported comparisons within the same context fiber.

Theorem 1 (Anchored order isomorphism)**.** *For every admissible embedding parameter value and all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$,*

$$\mathbf{x} \preceq_M \mathbf{x}' \quad \Longleftrightarrow \quad z_\theta(\mathbf{x}) \preceq z_\theta(\mathbf{x}'), \tag{9}$$

where the representation space is equipped with the product order. Consequently, z_θ is injective and is an order isomorphism from the input partially ordered set onto its image equipped with the induced product order.

The proof is given in Appendix A.2.

The classical characterization of nonnegative-orthant automorphisms identifies square linear product-order isomorphisms as positive coordinate scalings up to permutation (Ito & Lourenço, 2023). The diagonal choice retains constrained feature identities; Appendix A.2 explains the argument.

3.2.3 Bi-Lipschitz bounds

The Euclidean pullback distance $d_\theta(\mathbf{x}, \mathbf{x}') := \|z_\theta(\mathbf{x}) - z_\theta(\mathbf{x}')\|_2$ is a metric by Theorem 1.

Theorem 2 (Uniform bi-Lipschitz bound)**.** *For every admissible embedding parameter value and all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$,*

$$\underline{d}\,\|\mathbf{x} - \mathbf{x}'\|_2 \le d_\theta(\mathbf{x}, \mathbf{x}') \le \sqrt{2\overline{d}^{\,2} + \kappa^2}\,\|\mathbf{x} - \mathbf{x}'\|_2. \tag{10}$$

If $d_U = 0$, the upper constant tightens to $\sqrt{\overline{d}^{\,2} + \kappa^2}$. The same sharper bound holds for fixed-context pairs $\mathbf{x}_U = \mathbf{x}'_U$ when $d_U > 0$.

The proof is given in Appendix A.3. Let U_z denote the global upper constant in Theorem 2. The lower bound makes z_θ^{-1} on its image $1/\underline{d}$-Lipschitz, while U_z controls expansion. With the architectural bounds fixed, the same constants apply to every admissible parameter value throughout training.

For fixed admissible parameters, the anchor retains a recoverable scaled copy of every input coordinate and is itself an order isomorphism with parameter-uniform bi-Lipschitz bounds. Appending the nonlinear branch changes the representation geometry and supplies learned score features without requiring that branch to be injective.

3.3 Monotonic scalar readout

For an unconstrained vector $v \in \mathbb{R}^{d_M+2d_U+q}$, let $\bar{w} = \mathrm{softplus}(v)$. Given a fixed bound $\gamma_s > 0$, define the positive, norm-controlled readout and scalar score by

$$w = \frac{\gamma_s \bar{w}}{\max\{\gamma_s, \|\bar{w}\|_2\}}, \qquad s_\phi(\mathbf{x}) = w^\top z_\theta(\mathbf{x}) + \beta, \tag{11}$$

where $\phi = (v, \beta)$, $\beta \in \mathbb{R}$ is unconstrained, and the dependence of s_ϕ on the embedding parameters θ is suppressed for brevity. This parameterization gives $w > 0$ entrywise and $\|w\|_2 \le \gamma_s$. Partitioning $w = (w_M, w_U^+, w_U^-, w_h)$ according to Eq. (4), the score decomposes as

$$s_\phi(\mathbf{x}) = w_M^\top D_M \mathbf{x}_M + (w_U^+ - w_U^-)^\top D_U \mathbf{x}_U + w_h^\top h_\theta(\mathbf{x}) + \beta. \tag{12}$$

The context term is omitted when $d_U = 0$; otherwise, it permits linear effects of either sign. The anchor alone gives an affine score in the aligned inputs, while the nonlinear branch can add context-dependent nonlinear responses and interactions that remain nondecreasing in $\mathbf{x}_M$ at fixed $\mathbf{x}_U$.

Proposition 2 (Monotonicity and Lipschitz continuity of the scalar score)**.** *For every admissible embedding parameter value, every finite readout parameter value ϕ, and all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$, $\mathbf{x} \preceq_M \mathbf{x}'$ implies $s_\phi(\mathbf{x}) \le s_\phi(\mathbf{x}')$. Moreover,*

$$|s_\phi(\mathbf{x}) - s_\phi(\mathbf{x}')| \le \gamma_s U_z \|\mathbf{x} - \mathbf{x}'\|_2. \tag{13}$$

The proof is given in Appendix B.1.

Writing $L_s = \gamma_s U_z$, any nondecreasing L_g-Lipschitz scalar link g makes $g \circ s_\phi$ conditionally monotone and $L_g L_s$-Lipschitz. Injectivity, order reflection, and the lower metric bound remain embedding guarantees and are not generally inherited by scalar task outputs.

Task instantiations. For binary classification, the score is used as a logit; for regression, it predicts the training-standardized response and is mapped back to the original scale by a positive affine transformation; for ordinal classification, it serves as the latent score of an ordered-logistic head. Appendix B.2 gives the outputs, guarantees, and losses. The construction uses only feedforward operations, without enumerating ordered sample pairs or forming pairwise matrices. Constructing the coordinate-wise anchor requires $O(d)$ operations per input; full complexity is given in Appendix B.3.

Algorithm 1 summarizes the AOBE forward construction, from the coordinate-wise anchor and norm-controlled partially monotonic branch to their concatenated embedding and the positive, norm-controlled scalar readout.

4 Experiments

4.1 Experimental setup

Datasets. We evaluate six partially monotonic tabular benchmarks: COMPAS and Heart Disease for binary classification, BlogFeedback and Auto MPG for regression, and Wine Quality and Cleveland for ordinal classification. Dataset construction, preprocessing, and feature partitions are given in Appendix C.1.

Comparison methods. We compare AOBE with an unconstrained multilayer perceptron (MLP) (Hornik et al., 1989), a positive-weight partially monotonic variant (Pos-MLP) (Daniels & Velikova, 2010), LMN (Nolte et al., 2023), SMNN (Kim & Lee, 2024), SMM (Igel, 2024), and a source-aligned post-activation-switch MLP (PAS-MLP) (Sartor et al., 2025). All included methods share task heads, losses, and data partitions; Appendix C.1 gives the training protocol and comparison scope, and Table 4 in Appendix C.3 reports parameter counts.

Algorithm 1: AOBE embedding and monotonic readout

Input: Aligned input $\mathbf{x} = (\mathbf{x}_M, \mathbf{x}_U)$; raw parameters θ and $\phi = (v, \beta)$; fixed bounds $0 < \underline{d} < \overline{d}$, $\gamma_0, \ldots, \gamma_{L-1} > 0$, and $\gamma_s > 0$.

Output: Embedding $z_\theta(\mathbf{x})$ and score $s_\phi(\mathbf{x})$.

Generate D_M and, if $d_U > 0$, D_U with bounded positive diagonal entries;

Construct the order anchor $a(\mathbf{x})$;

$A_0 \leftarrow \begin{bmatrix}\mathrm{softplus}(V_0^M) & V_0^U\end{bmatrix}$ (omit V_0^U when $d_U = 0$); when $L > 1$, set $A_\ell \leftarrow \mathrm{softplus}(V_\ell)$ for $\ell = 1, \ldots, L-1$;

$u^{(0)} \leftarrow \mathbf{x}$;

for $\ell = 0, \ldots, L-1$ **do**

  $W_\ell \leftarrow \gamma_\ell A_\ell / \max\{\gamma_\ell, \|A_\ell\|_F\}$;

  $u^{(\ell+1)} \leftarrow \tanh(W_\ell u^{(\ell)} + b_\ell)$;

$h_\theta(\mathbf{x}) \leftarrow u^{(L)}$; $z_\theta(\mathbf{x}) \leftarrow \begin{bmatrix} a(\mathbf{x}) \\ h_\theta(\mathbf{x}) \end{bmatrix}$;

$\bar{w} \leftarrow \mathrm{softplus}(v)$; $w \leftarrow \gamma_s \bar{w} / \max\{\gamma_s, \|\bar{w}\|_2\}$;

$s_\phi(\mathbf{x}) \leftarrow w^\top z_\theta(\mathbf{x}) + \beta$;

return $z_\theta(\mathbf{x}), s_\phi(\mathbf{x})$.

Table 1: Primary performance (mean ± sample standard deviation over five seed-level split averages). Bold and underline mark the best and second-best means; ranks are descriptive.

Method	COMPAS AUROC ↑	Heart Disease AUROC ↑	Blog Feedback RMSE ↓	Auto MPG RMSE ↓	Wine Quality RPS ↓	Cleveland RPS ↓	Average rank ↓
MLP	0.7331±0.0012	0.8950±0.0122	**25.869±0.241**	2.845±0.066	0.0665±0.0011	0.1036±0.0052	3.50
Pos-MLP	0.7292±0.0096	0.8829±0.0310	28.837±0.811	2.914±0.043	0.0659±0.0006	0.1507±0.0141	6.00
LMN	<u>0.7337±0.0006</u>	<u>0.9053±0.0089</u>	26.572±0.257	2.958±0.069	0.0657±0.0007	<u>0.0973±0.0018</u>	<u>3.33</u>
SMNN	0.7325±0.0018	**0.9095±0.0040**	26.339±0.736	3.000±0.089	0.0662±0.0007	0.0999±0.0029	3.83
SMM	0.7319±0.0009	0.8905±0.0036	27.741±0.651	**2.788±0.053**	**0.0649±0.0007**	0.1157±0.0080	4.17
PAS-MLP	**0.7360±0.0007**	0.8865±0.0256	26.486±0.675	<u>2.828±0.063</u>	0.0678±0.0007	0.1045±0.0093	4.17
AOBE	0.7324±0.0009	0.8993±0.0066	<u>26.256±0.161</u>	2.945±0.035	<u>0.0651±0.0007</u>	**0.0957±0.0010**	**3.00**

Evaluation criteria. Primary metrics for binary classification, regression, and ordinal classification are area under the receiver operating characteristic curve (AUROC), original-scale root mean squared error (RMSE), and normalized ranked probability score (RPS), respectively. Coordinate-wise and joint monotonicity violation rates (MVRs) audit finite-sample scalar-score monotonicity. For four parameter-matched embeddings, deterministic label-free pair banks measure the order-preservation violation rate (OPVR), which counts failures to preserve comparable fixed-context pairs, and false-comparability rates (FCRs), which count incomparable inputs becoming comparable in either representation direction. Context-FCR uses identical constrained coordinates but distinct coordinate-wise ordered contexts; Mixed-FCR holds context fixed while increasing one constrained coordinate and decreasing another. A positive normalized pairwise incomparability margin indicates separation from both product-order cones. For positive finite distance ratios $r = \|\Delta z\|_2 / \|\Delta \mathbf{x}\|_2$, the central spread $D_{90} = Q_{0.95}(r)/Q_{0.05}(r)$ is closer to one when empirical distance scaling is more uniform. Complete definitions and protocols are in Appendix C.2.

4.2 Predictive performance and scalar monotonicity

Table 1 shows that AOBE achieves the best average rank (3.00), ranking first on Cleveland, second on BlogFeedback and Wine Quality, third on Heart Disease, and fifth on COMPAS and Auto MPG. These results demonstrate competitive prediction across the three task types alongside AOBE's structural guarantees.

Every constrained method records zero coordinate-wise and joint MVRs on the audited perturbations in all final runs, whereas the corresponding MLP dataset means range from 0.177 to 0.446 and from 0.002 to 0.319, respectively. Both audits support forward monotonicity but do not distinguish the

Table 2: Parameter-matched representation comparison, averaged uniformly over six datasets. Predictive ranks are descriptive; margins are normalized fifth percentiles. Guarantees concern order properties and parameter-uniform Lipschitz bounds.

Embedding	Average rank ↓	OPVR ↓	Context-FCR ↓	Mixed-FCR ↓	Context margin ↑	Mixed margin ↑	Structural guarantees
Unconstrained	2.33	0.909	0.064	0.086	0.054	0.059	None
Partially monotonic	3.50	**0.000**	0.719	0.933	−0.018	−0.018	Order preservation
Norm-controlled monotonic	2.33	**0.000**	0.832	0.953	−0.042	−0.046	Order preservation plus an upper Lipschitz bound
AOBE	**1.83**	**0.000**	**0.000**	**0.000**	**0.141**	**0.189**	Order isomorphism plus a parameter-uniform bi-Lipschitz bound

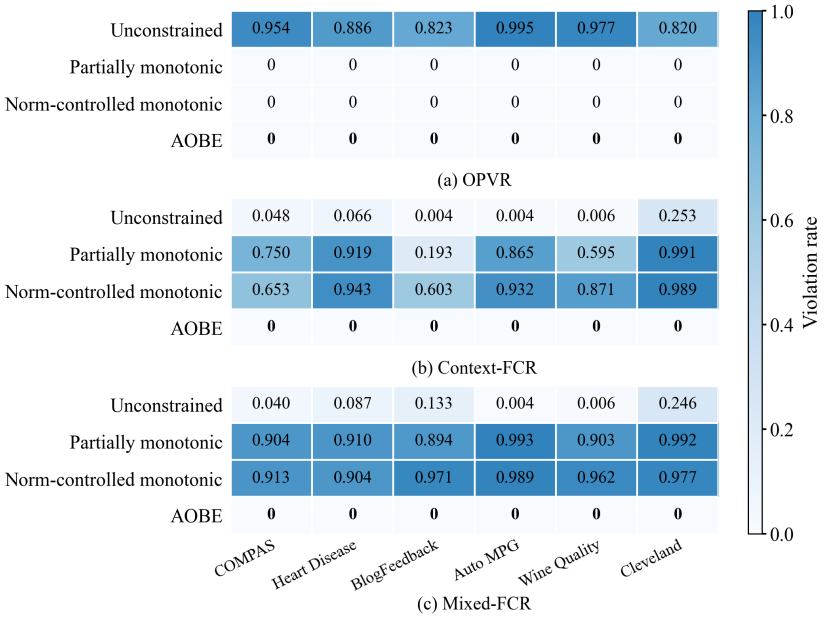


Figure 1: Dataset-level order-fidelity audits for the four parameter-matched embeddings: (a) OPVR, (b) Context-FCR, and (c) Mixed-FCR. Cells average five seed-level split averages; lower is better, with a common $[0, 1]$ color scale.

constrained methods, motivating the matched representation study below. Dataset-level results are in Table 5 of Appendix C.3.

4.3 Parameter-matched representation comparison

The matched study fixes representation dimension, scalar readout, task head, loss, and pair bank within each dataset, with parameter counts within 1.10% of AOBE. Selection uses predictive validation criteria only; Appendix C.3 describes the parameter-matched protocol.

Among the four matched embeddings in Table 2, AOBE alone combines zero audited OPVR, Context-FCR, and Mixed-FCR, positive margins, analytic order isomorphism, and a parameter-uniform bi-Lipschitz guarantee. The unconstrained embedding has OPVR 0.909, while the two unanchored monotonic embeddings preserve comparable pairs but incur high FCRs and negative margins. AOBE also achieves the best average predictive rank under the shared representation and capacity controls; dataset-level predictive results are reported in Table 6 of Appendix C.3.

Figure 1 shows zero OPVR, Context-FCR, and Mixed-FCR for AOBE on every dataset. Unconstrained OPVR ranges from 0.820 to 0.995; the two unanchored monotonic embeddings have zero OPVR but Context- and Mixed-FCR ranges of 0.193–0.991 and 0.894–0.993, respectively. These results expose two distinct failures: unconstrained representations can violate prescribed compar-

isons, while forward-monotonic representations can introduce unsupported comparisons between incomparable inputs. AOBE avoids both failures across the audited datasets, preserving the prescribed comparisons while keeping the tested context and mixed pairs incomparable.

On positive finite secants, AOBE has the lowest median local and global D_{90} (6.88 and 6.93), compared with 63.13 and 48.84 for the norm-controlled unanchored embedding. It also attains the smallest local and global D_{90} on five and four datasets, respectively. These results show that AOBE combines faithful order representation with more uniform empirical distance scaling across sampled input pairs. The substantially larger spread of the norm-controlled unanchored embedding further shows that upper norm control alone does not ensure comparable geometric regularity. Dataset-level values are reported in Table 7 of Appendix C.3.

A direct collapse audit complements D_{90} by retaining zero distance ratios and testing raw and calibration-normalized contraction thresholds. Among the 600 matched checkpoints, both the full and anchor views of all 150 AOBE checkpoints exhibit zero collision and near-collapse rates at the evaluated thresholds. The full embedding maintains separation even for sampled pairs that collide numerically in its nonlinear view, demonstrating how the anchor safeguards input distinguishability alongside the learned nonlinear features. Audit thresholds are specified in Appendix C.2; Table 8 in Appendix C.3 reports fixed-context and local collision and normalized near-collapse rates for AOBE and the baselines.

4.4 Ablation studies

Real-data ablations. Figure 5 in Appendix C.4 distinguishes the roles of the anchor, signed context pair, and norm control. Full AOBE and anchor-only attain zero Context-FCR and Mixed-FCR on all six datasets, whereas branch-only yields rates of 0.312–0.987 and 0.867–0.983, respectively, and larger global D_{90} on every dataset. These contrasts show that the anchor prevents unsupported comparability beyond forward monotonicity and improves empirical distance scaling in these comparisons. Replacing the signed context pair with a one-sided anchor introduces Context-FCRs of 0.082–0.831 while retaining zero Mixed-FCR, isolating the signed pair's role in cross-context incomparability. Removing branch norm control retains zero audited FCRs, order isomorphism, and the parameter-uniform lower bound, but removes the uniform upper Lipschitz guarantee. Its global D_{90} is lower than full AOBE on four datasets and higher on two; norm control supplies a uniform expansion certificate, while its effect on empirical secant spread is dataset-dependent. Full AOBE also improves on no-norm-control and anchor-only in mean primary performance on five and four tasks, respectively (Table 10, Appendix C.4).

Controlled synthetic ablations. We evaluate the same five variants across 12 synthetic settings under a common training budget to examine predictive and geometric effects.

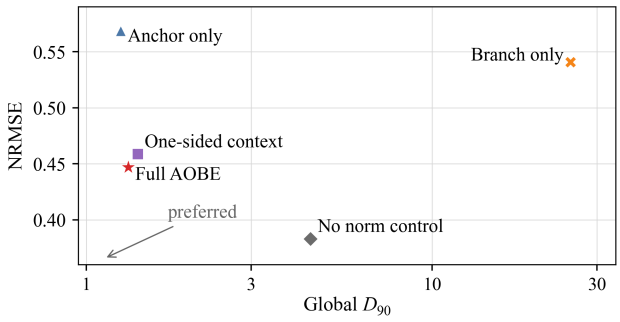


Figure 2: Synthetic predictive-error–geometry trade-off under a common maximum training budget. Points average five seeds within each setting and then the 12 settings uniformly; both axes are minimized, with logarithmic D_{90}.

Figure 2 shows that adding the nonlinear branch to anchor-only reduces normalized RMSE (NRMSE) from 0.569 to 0.447, with global D_{90} increasing modestly from 1.257 to 1.323. Removing branch norm control further reduces NRMSE to 0.383, but raises global D_{90} to 4.453

and removes the parameter-uniform upper Lipschitz guarantee. These results illustrate the nonlinear branch's predictive contribution and the predictive-error–geometry trade-off; Table 11 in Appendix C.4 gives complete component results.

4.5 Representation and score diagnostics

Frozen-channel readouts. To assess channel contributions within a fixed trained encoder, we freeze each of the 150 main-AOBE encoders and refit positive, norm-controlled readouts on its anchor, nonlinear, and concatenated views, using inherited optimization settings.

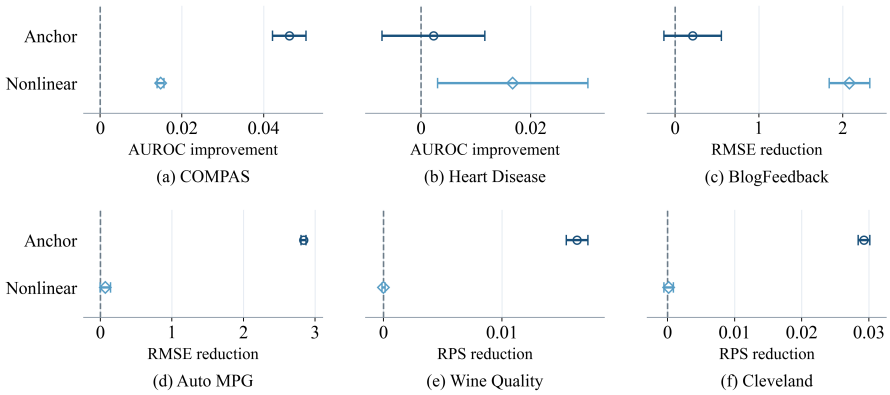


Figure 3: Frozen-channel predictive improvements: paired means and 95% Student-t intervals across five seed-level split averages. Improvements are full minus reference for AUROC and reference minus full for RMSE/RPS; positive values favor full.

Figure 3 shows that paired 95% confidence intervals favor full over anchor on four datasets and full over nonlinear on three. Full has the best mean on all six datasets (Table 12, Appendix C.5). On Auto MPG, adding nonlinear features reduces RMSE from 5.812 to 2.971; on BlogFeedback, adding anchor channels reduces RMSE from 28.259 to 26.179. These contrasts show task-dependent predictive contributions from both learned nonlinear features and retained input coordinates.

Certificate audits. All 150 main-comparison AOBE checkpoints record zero OPVR and FCRs and positive normalized fifth-percentile margins on the audit banks, and pass all available embedding- and score-bound checks (Table 9, Appendix C.3).

Score and task-output stability. We audit context-only, fixed-context constrained, joint local, and joint global perturbations. All 150 main-AOBE checkpoints pass the score- and task-output upper-bound checks in all four families. Across the 24 dataset–family combinations, AOBE has lower sensitivity than separately selected no-norm-control variants in 17 score comparisons and 16 task-output comparisons; BlogFeedback has higher sensitivity in all four families. These results connect certified bounds to bounded prediction changes and frequently lower observed sensitivity (Table 13, Appendix C.5).

5 Conclusion

AOBE addresses the gap between forward monotonicity and faithful conditional-order representation: preserving admissible comparisons alone does not prevent collapse, false comparability, or uncontrolled metric distortion. Its signed anchor and norm-controlled partially monotonic branch yield an injective conditional-order isomorphism with parameter-uniform bi-Lipschitz bounds. A positive, norm-controlled readout provides a conditionally monotone, Lipschitz score. Across six benchmarks, AOBE combines these guarantees with competitive prediction. Parameter-matched comparisons and direct-collapse audits show zero audited order-preservation and false-comparability violations and maintained input separation, while component studies and frozen-channel readouts support the task-dependent predictive utility of combining anchor and nonlinear features. Future work will explore reducing context duplication and tightening norm certificates to improve scalability.

AI USE STATEMENT

In this work, we used generative AI tools to edit the manuscript for clarity and readability. Additionally, we used these tools to assist with experimental code development and debugging and refinement of the synthetic experiment design. We did not use generative AI tools for other applicable tasks requiring disclosure; the remaining disclosure categories are not relevant to this work. We have reviewed all AI-assisted work, including testing the code and independently assessing the synthetic experimental design. We take responsibility for the final content of this work, including all text, claims, results, and artifacts produced with the assistance of generative AI.

REPRODUCIBILITY STATEMENT

Code and prepared data splits are available in an anonymous review repository at `https://anonymous.4open.science/r/AOBE-ICLR2027-review-4D36/`. The repository includes the prepared evaluation splits, selected configurations, tuning seed 2026, final seeds 42–46, and scripts for the main comparison, parameter-matched representation study, ablations, synthetic experiments, frozen-channel readouts, direct-collapse audits, and score and task-output stability audits. Preprocessing, evaluation metrics, and experimental protocols are detailed in Section 4 and Appendix C.

REFERENCES

Tal Amir and Nadav Dym. Fourier sliced-wasserstein embedding for multisets and measures. In *The Thirteenth International Conference on Learning Representations*, pp. 24590–24629, 2025.

Brandon Amos, Lei Xu, and J. Zico Kolter. Input convex neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 146–155, 2017.

Cem Anil, James Lucas, and Roger Grosse. Sorting out Lipschitz function approximation. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 291–301, 2019.

Jens Behrmann, Will Grathwohl, Ricky T. Q. Chen, David Duvenaud, and Jörn-Henrik Jacobsen. Invertible residual networks. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 573–582, 2019.

Jens Behrmann, Paul Vicol, Kuan-Chieh Wang, Roger Grosse, and Jörn-Henrik Jacobsen. Understanding and mitigating exploding inverses in invertible neural networks. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, pp. 1792–1800, 2021.

Moustapha Cisse, Piotr Bojanowski, Edouard Grave, Yann Dauphin, and Nicolas Usunier. Parseval networks: Improving robustness to adversarial examples. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 854–863, 2017.

Hennie Daniels and Marina Velikova. Monotone and partially monotone neural networks. *IEEE Transactions on Neural Networks*, 21(6):906–917, 2010.

Wouter Duivesteijn and Ad Feelders. Nearest neighbour classification with monotonicity constraints. In *Machine Learning and Knowledge Discovery in Databases*, pp. 301–316, 2008.

Mahyar Fazlyab, Alexander Robey, Hamed Hassani, Manfred Morari, and George J. Pappas. Efficient and accurate estimation of Lipschitz constants for deep neural networks. In *Advances in Neural Information Processing Systems*, 2019.

Henry Gouk, Eibe Frank, Bernhard Pfahringer, and Michael J. Cree. Regularisation of neural networks by enforcing Lipschitz continuity. *Machine Learning*, 110:393–416, 2021.

Maya Gupta, Andrew Cotter, Jan Pfeifer, Konstantin Voevodski, Kevin Canini, Alexander Mangylov, Wojciech Moczydlowski, and Alexander van Esbroeck. Monotonic calibrated interpolated look-up tables. *Journal of Machine Learning Research*, 17(109):1–47, 2016.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

Christian Igel. Smooth min-max monotonic networks. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 20908–20923, 2024.

Masaru Ito and Bruno F. Lourenço. Automorphisms of rank-one generated hyperbolicity cones and their derivative relaxations. *SIAM Journal on Applied Algebra and Geometry*, 7(1):236–263, 2023.

Hyunho Kim and Jong-Seok Lee. Scalable monotonic neural networks. In *International Conference on Learning Representations*, pp. 35160–35176, 2024.

Yuri Kinoshita and Taro Toyoizumi. A provable control of sensitivity of neural networks through a direct parameterization of the overall bi-Lipschitzness. In *Advances in Neural Information Processing Systems*, pp. 85241–85309, 2024.

Xingchao Liu, Xing Han, Na Zhang, and Qiang Liu. Certified monotonic neural networks. In *Advances in Neural Information Processing Systems*, pp. 15427–15438, 2020.

Mahdi Milani Fard, Kevin Canini, Andrew Cotter, Jan Pfeifer, and Maya Gupta. Fast and flexible monotonic functions with ensembles of lattices. In *Advances in Neural Information Processing Systems*, 2016.

Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, 2018.

Niklas Nolte, Ouail Kitouni, and Mike Williams. Expressive monotonic neural networks. In *The Eleventh International Conference on Learning Representations*, 2023.

Alejandro Polo-Molina, David Alfaya, and Jose Portela. MonoKAN: Certified monotonic kolmogorov-arnold network. *Neural Networks*, 196:108278, 2026.

Davor Runje and Sharath M. Shankaranarayana. Constrained monotonic neural networks. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 29338–29353, 2023.

Davide Sartor, Alberto Sinigaglia, and Gian Antonio Susto. Advancing constrained monotonic neural networks: Achieving universal approximation beyond bounded activations. In *Proceedings of the 42nd International Conference on Machine Learning*, pp. 52995–53018, 2025.

Joseph Sill. Monotonic networks. In *Advances in Neural Information Processing Systems*, 1997.

Aishwarya Sivaraman, Golnoosh Farnadi, Todd Millstein, and Guy Van den Broeck. Counterexample-guided learning of monotonic neural networks. In *Advances in Neural Information Processing Systems*, 2020.

Asher Trockman and J. Zico Kolter. Orthogonalizing convolutional layers with the cayley transform. In *International Conference on Learning Representations*, 2021.

Alexandre Verine, Benjamin Negrevergne, Yann Chevaleyre, and Fabrice Rossi. On the expressivity of bi-Lipschitz normalizing flows. In *Proceedings of the 14th Asian Conference on Machine Learning*, pp. 1054–1069, 2023.

Ruigang Wang and Ian R. Manchester. Direct parameterization of Lipschitz-bounded deep networks. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 36093–36110, 2023.

Antoine Wehenkel and Gilles Louppe. Unconstrained monotonic neural networks. In *Advances in Neural Information Processing Systems*, 2019.

Yuezhu Xu and S. Sivaranjani. Eclipse: Efficient compositional Lipschitz constant estimation for deep neural networks. In *Advances in Neural Information Processing Systems*, pp. 10414–10441, 2024.

Seungil You, David Ding, Kevin Canini, Jan Pfeifer, and Maya R. Gupta. Deep lattice networks and partial monotonic functions. In *Advances in Neural Information Processing Systems*, 2017.

Jiachi Zhao, Hongwen Zhang, Yue Wang, Yiteng Zhai, and Yao Yang. Deep isotonic embedding network: A flexible monotonic neural network. *Neural Networks*, 171:457–465, 2024.

A REPRESENTATION GUARANTEES

A.1 PARTIAL MONOTONICITY AND LIPSCHITZ CONTINUITY

Proof of Proposition 1. Fix an admissible embedding parameter value and write

$$c_\ell := \frac{\gamma_\ell}{\max\{\gamma_\ell, \|A_\ell\|_F\}} > 0, \qquad W_\ell = c_\ell A_\ell. \tag{14}$$

Partition the first-layer matrix as $W_0 = [W_0^M \; W_0^U]$ according to the input blocks. Eqs. (5) and (14) give $W_0^M = c_0 \operatorname{softplus}(V_0^M) \geq 0$ entrywise and $W_\ell = c_\ell \operatorname{softplus}(V_\ell) \geq 0$ entrywise for every $\ell = 1, \ldots, L-1$. The block W_0^U may have arbitrary signs.

Partial monotonicity. Fix $\mathbf{x}_U \in \mathbb{R}^{d_U}$ and let $\mathbf{x}_M \preceq \mathbf{x}'_M$. Let $u^{(\ell)}$ and $v^{(\ell)}$ denote the states generated from $\mathbf{x} = (\mathbf{x}_M, \mathbf{x}_U)$ and $\mathbf{x}' = (\mathbf{x}'_M, \mathbf{x}_U)$, respectively. At the first layer,

$$\big(W_0 v^{(0)} + b_0\big) - \big(W_0 u^{(0)} + b_0\big) = W_0^M(\mathbf{x}'_M - \mathbf{x}_M) \succeq 0. \tag{15}$$

Here $\mathbf{x}'_M - \mathbf{x}_M \succeq 0$ and $W_0^M \geq 0$ entrywise, while the context terms and biases cancel. Since $\tanh$ is coordinate-wise increasing, Eq. (15) implies $u^{(1)} \preceq v^{(1)}$. Suppose that $u^{(\ell)} \preceq v^{(\ell)}$ for some $\ell \in \{1, \ldots, L-1\}$. The entrywise nonnegativity of W_ℓ gives

$$\big(W_\ell v^{(\ell)} + b_\ell\big) - \big(W_\ell u^{(\ell)} + b_\ell\big) = W_\ell\big(v^{(\ell)} - u^{(\ell)}\big) \succeq 0. \tag{16}$$

Applying the coordinate-wise monotonicity of $\tanh$ again yields $u^{(\ell+1)} \preceq v^{(\ell+1)}$. Thus, by induction (with no induction step required when $L = 1$),

$$h_\theta(\mathbf{x}_M, \mathbf{x}_U) = u^{(L)} \preceq v^{(L)} = h_\theta(\mathbf{x}'_M, \mathbf{x}_U). \tag{17}$$

When $d_U = 0$, the same argument applies with the context block omitted.

Lipschitz continuity. For every $\ell = 0, \ldots, L-1$, Eq. (6) gives

$$\|W_\ell\|_F = \frac{\gamma_\ell \|A_\ell\|_F}{\max\{\gamma_\ell, \|A_\ell\|_F\}} \leq \gamma_\ell, \qquad \|W_\ell\|_2 \leq \|W_\ell\|_F. \tag{18}$$

The latter inequality follows from the singular values. For any nonzero matrix A, $1 \leq \|A\|_F / \|A\|_2 \leq \sqrt{\operatorname{rank}(A)}$. Thus, Frobenius control can be conservative relative to spectral-norm control, with the gap depending on the matrix rather than its width alone. Because $|\tanh'(t)| = \operatorname{sech}^2(t) \leq 1$, the mean value theorem gives $|\tanh(\xi_j) - \tanh(\xi'_j)| \leq |\xi_j - \xi'_j|$ for each coordinate. Hence, for arbitrary vectors ξ and ξ' of the same dimension,

$$\|\tanh(\xi) - \tanh(\xi')\|_2^2 = \sum_j |\tanh(\xi_j) - \tanh(\xi'_j)|^2 \leq \sum_j |\xi_j - \xi'_j|^2 = \|\xi - \xi'\|_2^2. \tag{19}$$

Let $u^{(\ell)}$ and $v^{(\ell)}$ be the states generated from arbitrary $\mathbf{x}$ and $\mathbf{x}'$, respectively. Using Eq. (19), cancellation of the common bias, the operator-norm inequality, and Eq. (18), for each $\ell = 0, \ldots, L-1$, we obtain

$$\begin{aligned} \|u^{(\ell+1)} - v^{(\ell+1)}\|_2 &\leq \|W_\ell(u^{(\ell)} - v^{(\ell)})\|_2 \\ &\leq \|W_\ell\|_2 \|u^{(\ell)} - v^{(\ell)}\|_2 \\ &\leq \gamma_\ell \|u^{(\ell)} - v^{(\ell)}\|_2. \end{aligned} \tag{20}$$

Iterating Eq. (20) and using $u^{(0)} = \mathbf{x}$ and $v^{(0)} = \mathbf{x}'$ yields

$$\begin{aligned} \|h_\theta(\mathbf{x}) - h_\theta(\mathbf{x}')\|_2 &= \|u^{(L)} - v^{(L)}\|_2 \\ &\leq \left(\prod_{\ell=0}^{L-1} \gamma_\ell\right) \|u^{(0)} - v^{(0)}\|_2 \\ &= \kappa \|\mathbf{x} - \mathbf{x}'\|_2. \end{aligned} \tag{21}$$

□

A.2 ANCHORED ORDER ISOMORPHISM

Proof of Theorem 1. Fix an admissible embedding parameter value and let $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$. Suppose that $\mathbf{x} \preceq_M \mathbf{x}'$. Then $\mathbf{x}_M \preceq \mathbf{x}'_M$ and $\mathbf{x}_U = \mathbf{x}'_U$. Since every diagonal entry of D_M is strictly positive, $D_M\mathbf{x}_M \preceq D_M\mathbf{x}'_M$.

The two signed context blocks in $a(\mathbf{x})$ and $a(\mathbf{x}')$ are equal. Proposition 1 also gives

$$h_\theta(\mathbf{x}_M, \mathbf{x}_U) \preceq h_\theta(\mathbf{x}'_M, \mathbf{x}_U) = h_\theta(\mathbf{x}'_M, \mathbf{x}'_U). \tag{22}$$

Thus every block of $z_\theta(\mathbf{x})$ is coordinate-wise no greater than the corresponding block of $z_\theta(\mathbf{x}')$, proving $z_\theta(\mathbf{x}) \preceq z_\theta(\mathbf{x}')$.

Conversely, suppose that $z_\theta(\mathbf{x}) \preceq z_\theta(\mathbf{x}')$. Comparing the constrained anchor blocks gives $D_M\mathbf{x}_M \preceq D_M\mathbf{x}'_M$. Thus, for each $j = 1, \ldots, d_M$, $d_j^M(x'_{M,j} - x_{M,j}) \geq 0$. Since $d_j^M > 0$, it follows that $x_{M,j} \leq x'_{M,j}$, and hence $\mathbf{x}_M \preceq \mathbf{x}'_M$.

When $d_U > 0$, comparing the signed context blocks gives

$$D_U\mathbf{x}_U \preceq D_U\mathbf{x}'_U, \qquad -D_U\mathbf{x}_U \preceq -D_U\mathbf{x}'_U. \tag{23}$$

The second inequality is equivalent to $D_U\mathbf{x}_U \succeq D_U\mathbf{x}'_U$. Together with the first, it yields $D_U\mathbf{x}_U = D_U\mathbf{x}'_U$. Every diagonal entry of D_U is strictly positive, so D_U is invertible and therefore $\mathbf{x}_U = \mathbf{x}'_U$. When $d_U = 0$, there is a unique empty context vector, so the context equality is automatic. In either case, $\mathbf{x}_M \preceq \mathbf{x}'_M$ and $\mathbf{x}_U = \mathbf{x}'_U$, which by Eq. (1) gives $\mathbf{x} \preceq_M \mathbf{x}'$. This proves order reflection and completes the equivalence in Eq. (9).

It remains to establish the stated consequences. If $z_\theta(\mathbf{x}) = z_\theta(\mathbf{x}')$, then both $z_\theta(\mathbf{x}) \preceq z_\theta(\mathbf{x}')$ and $z_\theta(\mathbf{x}') \preceq z_\theta(\mathbf{x})$ hold. Applying the proved equivalence in both directions gives $\mathbf{x} \preceq_M \mathbf{x}'$ and $\mathbf{x}' \preceq_M \mathbf{x}$. Antisymmetry of $\preceq_M$ yields $\mathbf{x} = \mathbf{x}'$, so z_θ is injective.

Let $\mathcal{Z}_\theta = z_\theta(\mathbb{R}^d)$ be the image. The map $z_\theta : \mathbb{R}^d \to \mathcal{Z}_\theta$ is surjective by definition and injective by the preceding argument, hence bijective. For any $y = z_\theta(\mathbf{x})$ and $y' = z_\theta(\mathbf{x}')$ in $\mathcal{Z}_\theta$, Eq. (9) states

$$\mathbf{x} \preceq_M \mathbf{x}' \quad \Longleftrightarrow \quad y \preceq y'. \tag{24}$$

Thus z_θ and its inverse on $\mathcal{Z}_\theta$ both preserve order, proving that z_θ is an order isomorphism from the input partially ordered set onto its image. □

Square linear anchors. For square linear maps on the constrained coordinates, preserving and reflecting the product order is equivalent to positive coordinate-wise scaling up to permutation. Specifically, for $A \in \mathbb{R}^{d_M \times d_M}$, $x \preceq y \iff Ax \preceq Ay$ for all $x, y \in \mathbb{R}^{d_M}$ holds if and only if $A = PD$, where P is a permutation matrix and D is a strictly positive diagonal matrix.

Order preservation implies $A \geq 0$ entrywise, while order reflection implies injectivity and hence invertibility. Reflection also gives $A^{-1} \geq 0$. Writing $B = A^{-1}$, for $i \neq j$ we have $0 = (AB)_{ij} = \sum_k A_{ik}B_{kj}$, with every summand nonnegative. For each k, choose j with $B_{kj} > 0$; then $A_{ik} = 0$ for every $i \neq j$. Invertibility therefore forces exactly one positive entry in each column and row of A, yielding $A = PD$. Conversely, permutations and positive coordinate-wise scalings preserve and reflect the product order. AOBE retains the constrained feature identities and uses the diagonal case.

A.3 UNIFORM BI-LIPSCHITZ BOUND

Proof of Theorem 2. Fix an admissible embedding parameter value and arbitrary $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$. Let $\delta_M = \mathbf{x}_M - \mathbf{x}'_M$ and $\delta_U = \mathbf{x}_U - \mathbf{x}'_U$. Since z_θ concatenates its anchor and nonlinear blocks, their squared Euclidean distances add:

$$\begin{aligned} d_\theta^2(\mathbf{x}, \mathbf{x}') &= \|D_M\delta_M\|_2^2 + \|D_U\delta_U\|_2^2 + \| - D_U\delta_U\|_2^2 + \|h_\theta(\mathbf{x}) - h_\theta(\mathbf{x}')\|_2^2 \\ &= \|D_M\delta_M\|_2^2 + 2\|D_U\delta_U\|_2^2 + \|h_\theta(\mathbf{x}) - h_\theta(\mathbf{x}')\|_2^2, \end{aligned} \tag{25}$$

where the context terms are omitted when $d_U = 0$.

The diagonal scale bounds in Eq. (2) imply

$$\underline{d}^{\,2}\|\delta_M\|_2^2 \leq \|D_M\delta_M\|_2^2 \leq \overline{d}^{\,2}\|\delta_M\|_2^2, \tag{26}$$

$$\underline{d}^{\,2}\|\delta_U\|_2^2 \leq \|D_U\delta_U\|_2^2 \leq \overline{d}^{\,2}\|\delta_U\|_2^2, \tag{27}$$

with Eq. (27) omitted when $d_U = 0$. Dropping the nonnegative nonlinear term in Eq. (25) gives

$$\begin{aligned} d_\theta^2(\mathbf{x}, \mathbf{x}') &\geq \underline{d}^{\,2}\big(\|\delta_M\|_2^2 + 2\|\delta_U\|_2^2\big) \\ &\geq \underline{d}^{\,2}\big(\|\delta_M\|_2^2 + \|\delta_U\|_2^2\big) \\ &= \underline{d}^{\,2}\|\mathbf{x} - \mathbf{x}'\|_2^2. \end{aligned} \tag{28}$$

For the upper bound, Proposition 1 gives

$$\|h_\theta(\mathbf{x}) - h_\theta(\mathbf{x}')\|_2^2 \leq \kappa^2\|\mathbf{x} - \mathbf{x}'\|_2^2 = \kappa^2\big(\|\delta_M\|_2^2 + \|\delta_U\|_2^2\big). \tag{29}$$

When $d_U > 0$, substituting the upper bounds in Eqs. (26), (27), and (29) into Eq. (25) yields

$$\begin{aligned} d_\theta^2(\mathbf{x}, \mathbf{x}') &\leq (\overline{d}^{\,2} + \kappa^2)\|\delta_M\|_2^2 + (2\overline{d}^{\,2} + \kappa^2)\|\delta_U\|_2^2 \\ &\leq (2\overline{d}^{\,2} + \kappa^2)\big(\|\delta_M\|_2^2 + \|\delta_U\|_2^2\big) \\ &= (2\overline{d}^{\,2} + \kappa^2)\|\mathbf{x} - \mathbf{x}'\|_2^2. \end{aligned} \tag{30}$$

If $d_U = 0$, the context blocks are absent, and the same calculation instead gives

$$d_\theta^2(\mathbf{x}, \mathbf{x}') \leq (\overline{d}^{\,2} + \kappa^2)\|\mathbf{x} - \mathbf{x}'\|_2^2. \tag{31}$$

Taking nonnegative square roots of Eq. (28) together with Eq. (30) or, when $d_U = 0$, Eq. (31) proves the stated bounds. Their constants depend only on the fixed values $\underline{d}$, $\overline{d}$, and $\gamma_0, \ldots, \gamma_{L-1}$ through κ, so they are uniform over admissible embedding parameter values.

For completeness, nonnegativity, symmetry, and the triangle inequality for d_θ follow from the Euclidean norm, while Eq. (28) gives $d_\theta(\mathbf{x}, \mathbf{x}') = 0$ only if $\mathbf{x} = \mathbf{x}'$. Hence d_θ is a metric. Moreover, for $y = z_\theta(\mathbf{x})$ and $y' = z_\theta(\mathbf{x}')$ in the image of z_θ, the lower bound gives

$$\|z_\theta^{-1}(y) - z_\theta^{-1}(y')\|_2 = \|\mathbf{x} - \mathbf{x}'\|_2 \leq \frac{1}{\underline{d}}\|y - y'\|_2, \tag{32}$$

which proves that the inverse is $1/\underline{d}$-Lipschitz.

The same diagonal estimates applied to the anchor alone give $\underline{d}\|\mathbf{x} - \mathbf{x}'\|_2 \leq \|a(\mathbf{x}) - a(\mathbf{x}')\|_2 \leq \sqrt{2\overline{d}}\|\mathbf{x} - \mathbf{x}'\|_2$, with upper constant $\overline{d}$ when $d_U = 0$. Its positive constrained block and signed context pair also give the order equivalence by the argument in Appendix A.2. Thus, input retention, order reflection, and a positive lower bound do not require the nonlinear branch; its partial monotonicity and norm control ensure that concatenation preserves the full construction's guarantees.

For fixed-context pairs with $d_U > 0$, setting $\delta_U = 0$ in the first line of Eq. (30) gives the same sharper upper bound as in the no-context case. □

B Prediction guarantees and task instantiations

B.1 Scalar-score monotonicity and stability

Proof of Proposition 2. If $\mathbf{x} \preceq_M \mathbf{x}'$, Theorem 1 gives $z_\theta(\mathbf{x}) \preceq z_\theta(\mathbf{x}')$. Since $w > 0$ entrywise,

$$s_\phi(\mathbf{x}') - s_\phi(\mathbf{x}) = w^\top(z_\theta(\mathbf{x}') - z_\theta(\mathbf{x})) \geq 0, \tag{33}$$

which proves score monotonicity. For arbitrary $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$, the Cauchy–Schwarz inequality, $\|w\|_2 \leq \gamma_s$, and Theorem 2 yield

$$\begin{aligned} |s_\phi(\mathbf{x}) - s_\phi(\mathbf{x}')| &= \big|w^\top(z_\theta(\mathbf{x}) - z_\theta(\mathbf{x}'))\big| \\ &\leq \|w\|_2\,\|z_\theta(\mathbf{x}) - z_\theta(\mathbf{x}')\|_2 \\ &\leq \gamma_s U_z\|\mathbf{x} - \mathbf{x}'\|_2. \end{aligned} \tag{34}$$

□

Context perturbations. Under the assumptions of Proposition 2, let $L_s = \gamma_s U_z$. For $d_U > 0$, any $\mathbf{x}_M \preceq \mathbf{x}'_M$ and $\mathbf{u}, \mathbf{u}' \in \mathbb{R}^{d_U}$ satisfy

$$s_\phi((\mathbf{x}_M, \mathbf{u})) \leq s_\phi((\mathbf{x}'_M, \mathbf{u}')) + L_s \|\mathbf{u} - \mathbf{u}'\|_2. \tag{35}$$

This follows by applying conditional monotonicity at $\mathbf{u}$ and then the score Lipschitz bound between $(\mathbf{x}'_M, \mathbf{u})$ and $(\mathbf{x}'_M, \mathbf{u}')$. The bound reduces to exact monotonicity when the contexts agree and otherwise controls how much a context change can offset the comparison without ordering the contexts.

B.2 Task outputs and training objectives

All comparison methods produce a scalar score and use the same task-specific links and losses below. For AOBE, this score is s_ϕ from Eq. (11); the baselines substitute their respective scalar scores. For AOBE, let $L_s = \gamma_s U_z$ denote the parameter-uniform upper Lipschitz bound for the score in Eq. (13).

Binary classification. For binary labels $y_i \in \{0, 1\}$, the score is used as a logit, giving the positive-class probability $p_{\text{bin}}(\mathbf{x}) = \sigma(s_\phi(\mathbf{x}))$. For AOBE, Proposition 2 and monotonicity of the logistic link imply that $\mathbf{x} \preceq_M \mathbf{x}' \Longrightarrow p_{\text{bin}}(\mathbf{x}) \leq p_{\text{bin}}(\mathbf{x}')$. Moreover, since the logistic function is $1/4$-Lipschitz, for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$,

$$|p_{\text{bin}}(\mathbf{x}) - p_{\text{bin}}(\mathbf{x}')| \leq \frac{L_s}{4} \|\mathbf{x} - \mathbf{x}'\|_2. \tag{36}$$

Regression. For a real-valued response, let μ_{tr} and $\varsigma_{\text{tr}} > 0$ be the mean and standard deviation estimated only from the training split, and define $\widetilde{y}_i = (y_i - \mu_{\text{tr}})/\varsigma_{\text{tr}}$. The score predicts the standardized response, so the corresponding original-scale prediction is $\widehat{y}_{\text{reg}}(\mathbf{x}) := \mu_{\text{tr}} + \varsigma_{\text{tr}} s_\phi(\mathbf{x})$. For AOBE, Proposition 2 and $\varsigma_{\text{tr}} > 0$ imply that $\mathbf{x} \preceq_M \mathbf{x}' \Longrightarrow \widehat{y}_{\text{reg}}(\mathbf{x}) \leq \widehat{y}_{\text{reg}}(\mathbf{x}')$. Moreover, for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$,

$$|\widehat{y}_{\text{reg}}(\mathbf{x}) - \widehat{y}_{\text{reg}}(\mathbf{x}')| = \varsigma_{\text{tr}} \, |s_\phi(\mathbf{x}) - s_\phi(\mathbf{x}')| \leq \varsigma_{\text{tr}} L_s \|\mathbf{x} - \mathbf{x}'\|_2. \tag{37}$$

Ordinal classification. For ordinal labels $y_i \in \{1, \ldots, C\}$ ordered as $1 \prec \cdots \prec C$, with $C \geq 2$, unconstrained parameters $\rho_1, \ldots, \rho_{C-1}$ define an ordered-logistic cumulative-link model with ordered thresholds and upper-tail probabilities:

$$\begin{aligned} \tau_1 = \rho_1, \qquad \tau_c = \tau_{c-1} + \text{softplus}(\rho_c), \quad c = 2, \ldots, C-1, \\ q_c(\mathbf{x}) = \Pr(Y > c \mid \mathbf{x}) = \sigma(s_\phi(\mathbf{x}) - \tau_c), \quad c = 1, \ldots, C-1. \end{aligned} \tag{38}$$

When $C = 2$, only $\tau_1 = \rho_1$ is needed; the recursion for $c \geq 2$ is omitted. Both the score bias β and threshold location ρ_1 are trained. A common translation leaves ordinal probabilities and the unregularized task loss unchanged. This location redundancy does not affect threshold ordering or the monotonicity and Lipschitz guarantees. The ordered thresholds imply $1 > q_1(\mathbf{x}) > \cdots > q_{C-1}(\mathbf{x}) > 0$. Setting $q_0(\mathbf{x}) \equiv 1$ and $q_C(\mathbf{x}) \equiv 0$, the predictive class probabilities are

$$\pi_c(\mathbf{x}) = q_{c-1}(\mathbf{x}) - q_c(\mathbf{x}), \qquad c = 1, \ldots, C. \tag{39}$$

Because $q_0(\mathbf{x}) > q_1(\mathbf{x}) > \cdots > q_C(\mathbf{x})$, every $\pi_c(\mathbf{x}) > 0$, and telescoping gives $\sum_{c=1}^{C} \pi_c(\mathbf{x}) = q_0(\mathbf{x}) - q_C(\mathbf{x}) = 1$. Thus $\boldsymbol{\pi}(\mathbf{x}) = (\pi_1(\mathbf{x}), \ldots, \pi_C(\mathbf{x}))$ is a valid ordinal predictive distribution. For AOBE, Proposition 2 and monotonicity of the logistic link give

$$\mathbf{x} \preceq_M \mathbf{x}' \quad \Longrightarrow \quad q_c(\mathbf{x}) \leq q_c(\mathbf{x}'), \qquad c = 1, \ldots, C-1. \tag{40}$$

Because the q_c are upper-tail probabilities, Eq. (40) is precisely the condition that, whenever $\mathbf{x} \preceq_M \mathbf{x}'$, the predictive distribution $\boldsymbol{\pi}(\mathbf{x}')$ first-order stochastically dominates $\boldsymbol{\pi}(\mathbf{x})$. Moreover, since the logistic function is $1/4$-Lipschitz, for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$ and $c = 1, \ldots, C-1$,

$$|q_c(\mathbf{x}) - q_c(\mathbf{x}')| \leq \frac{L_s}{4} \|\mathbf{x} - \mathbf{x}'\|_2. \tag{41}$$

Training objectives. For a training set $\mathcal{D}_{\mathrm{tr}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$, let ψ collect all trainable scalar-scorer parameters, write $s_i = s_\psi(\mathbf{x}_i)$, set $\rho = (\rho_1, \ldots, \rho_{C-1})$, define $q_{ic} = \sigma(s_i - \tau_c)$, and let $\mathtt{t} \in \{\mathrm{bin}, \mathrm{reg}, \mathrm{ord}\}$. For AOBE, $\psi = (\theta, \phi)$. With $\Theta_{\mathrm{bin}} = \Theta_{\mathrm{reg}} = \psi$ and $\Theta_{\mathrm{ord}} = (\psi, \rho)$, the empirical task loss is

$$\mathcal{L}_{\mathtt{t}}(\Theta_{\mathtt{t}}; \mathcal{D}_{\mathrm{tr}}) = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \left[\mathrm{softplus}(s_i) - y_i s_i\right], & \mathtt{t} = \mathrm{bin}, \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \left(s_i - \widetilde{y}_i\right)^2, & \mathtt{t} = \mathrm{reg}, \\ \dfrac{1}{n(C-1)}\displaystyle\sum_{i=1}^{n}\sum_{c=1}^{C-1} \left(q_{ic} - \mathbb{I}[y_i > c]\right)^2, & \mathtt{t} = \mathrm{ord}. \end{cases} \tag{42}$$

B.3 Computational complexity

For AOBE, let $P_h = \sum_{\ell=0}^{L-1} q_\ell q_{\ell+1}$ denote the number of nonlinear-branch weight entries, let $d_z = d_M + 2d_U + q$ be the embedding dimension, and let $K_{\mathtt{t}} = C-1$ for ordinal classification and $K_{\mathtt{t}} = 1$ otherwise. The anchor occupies $d_M + 2d_U = d + d_U$ coordinates, adding d_U coordinates relative to the original input. For a batch of size B, one forward–backward step requires $O(B(P_h + d_z + K_{\mathtt{t}}))$ time, while per-input inference requires $O(P_h + d_z + K_{\mathtt{t}})$ time. The number of trainable parameters is $O(P_h + \sum_{\ell=1}^{L} q_\ell + d_z + K_{\mathtt{t}})$, and E epochs over n samples require $O(En(P_h + d_z + K_{\mathtt{t}}))$ time.

C Experimental details and additional results

C.1 Datasets and comparison scope

The binary and regression tasks follow the target definitions and monotonicity specifications of Nolte et al. (2023) and Kinoshita & Toyoizumi (2024); Table 3 lists the feature partitions, including those specified for the ordinal tasks before optimization. Each task uses five evaluation splits. BlogFeedback retains its official temporal test set and uses five repeated training–validation splits of the official training pool; the other tasks use five outer folds. Inputs are min–max scaled using training statistics only, without clipping validation or test values; decreasing constrained coordinates are replaced by one minus their scaled values. Categorical features are one-hot encoded where specified in the table, while the supplied ordinal folds retain their numeric encodings. Regression targets are standardized using training statistics, and predictions are evaluated in the original response units.

Table 3: Datasets and feature partitions.

Dataset	Task	n	$d_M + d_U$	Feature partition
COMPAS	Binary classification	6,172	$4 + 9$	**Constrained:** priors count $(+)$; juvenile felony count $(+)$; juvenile misdemeanor count $(+)$; juvenile other count $(+)$. **Context:** age; race (6 indicators); sex (2 indicators).
Heart Disease	Binary classification	303	$2 + 29$	**Constrained:** resting blood pressure $(+)$; cholesterol $(+)$. **Context:** age; maximum heart rate; oldpeak; one-hot indicators: sex (2); chest pain (5); fasting blood sugar (2); resting ECG (3); exercise angina (2); slope (3); vessels (4); thalassemia (5).
BlogFeedback	Regression	60,021	$8 + 268$	**Constrained:** attributes 051–054 $(+)$; attributes 056–059 $(+)$. **Context:** remaining nonconstant attributes.
Auto MPG	Regression	392	$3 + 6$	**Constrained:** displacement $(-)$; horsepower $(-)$; weight $(-)$. **Context:** cylinders; acceleration; model year; origin (3 indicators).
Wine Quality	Ordinal classification	1,359	$6 + 5$	**Constrained:** volatile acidity $(-)$; citric acid $(+)$; chlorides $(-)$; density $(-)$; sulphates $(+)$; alcohol $(+)$. **Context:** fixed acidity; residual sugar; free sulfur dioxide; total sulfur dioxide; pH.
Cleveland	Ordinal classification	297	$6 + 7$	**Constrained:** age $(+)$; resting blood pressure $(+)$; cholesterol $(+)$; maximum heart rate $(-)$; oldpeak $(+)$; vessels $(+)$. **Context:** sex; chest pain; fasting blood sugar; resting ECG; exercise angina; slope; thalassemia.

Training and selection. For each main-comparison dataset–method–split, we evaluate 16 predeclared configurations with tuning seed 2026. AdamW training uses at most 500 epochs and

validation-loss early stopping with patience 60; checkpoints are selected by validation task loss and configurations by the primary validation metric. Selected configurations are retrained with seeds 42–46; results are averaged over the five splits within each seed, followed by the mean and sample standard deviation across seeds. The candidate designs are specified in the released tuning code.

BLNN and PBLNN. BLNN and PBLNN (Kinoshita & Toyoizumi, 2024) are relevant to AOBE's bi-Lipschitz guarantee: BLNN provides two-sided Euclidean distance bounds for its full-input map, whereas PBLNN provides such bounds for selected variables when the remaining variables are fixed. Neither construction guarantees the full-input conditional-order isomorphism evaluated in our parameter-matched study. The authors describe the released module as a "skeleton" of BLNN and PBLNN and recommend integrating it with a deep-equilibrium framework for an effective implementation and more efficient backpropagation.[1]. Therefore, we compare the scope of their analytic guarantees in Section 2, but do not include either method in our numerical comparison.

C.2 Evaluation metrics and representation audits

Scalar monotonicity metrics. For each run, let s denote the scalar scorer under audit, set $N_p = \min\{256, n_{\text{test}}\}$, and select N_p test inputs deterministically without replacement. For $\delta \in \Delta = \{0.01, 0.05, 0.20\}$, clamp the aligned constrained coordinates to $[0, 1-\delta]$, yielding $\bar{\mathbf{x}}_M^{(i,\delta)}$. With e_j denoting the jth coordinate vector, the coordinate-wise and joint MVRs are

$$\begin{aligned}
\mathrm{MVR}_{\text{coord}} &= \frac{1}{|\Delta| N_p d_M} \sum_{\delta\in\Delta} \sum_{i=1}^{N_p} \sum_{j=1}^{d_M} \mathbb{I}\Big[s\Big((\bar{\mathbf{x}}_M^{(i,\delta)} + \delta e_j, \mathbf{x}_U^{(i)}) \Big) + \epsilon < s\Big((\bar{\mathbf{x}}_M^{(i,\delta)}, \mathbf{x}_U^{(i)}) \Big) \Big], \\
\mathrm{MVR}_{\text{joint}} &= \frac{1}{|\Delta| N_p} \sum_{\delta\in\Delta} \sum_{i=1}^{N_p} \mathbb{I}\Big[s\Big((\bar{\mathbf{x}}_M^{(i,\delta)} + \delta \mathbf{1}, \mathbf{x}_U^{(i)}) \Big) + \epsilon < s\Big((\bar{\mathbf{x}}_M^{(i,\delta)}, \mathbf{x}_U^{(i)}) \Big) \Big].
\end{aligned} \tag{43}$$

Here $\epsilon > 0$ is a numerical tolerance; decreases of magnitude at most ϵ are not counted as violations.

Representation pair banks and metrics. Each evaluation split uses deterministic, feature-only pair banks shared across embeddings and final seeds, with 1,024 pairs per nonempty family. Comparable pairs fix the context and increase one or all varying constrained coordinates; mixed pairs fix the context and increase one constrained coordinate while decreasing another. Context pairs fix $\mathbf{x}_M$ and use distinct coordinate-wise minima and maxima of two held-out context vectors; the resulting contexts need not retain valid categorical encodings. Local secants symmetrically perturb held-out inputs along unit-normalized Gaussian directions, scaled by an independent amplitude drawn uniformly from $[0.0125, 0.05]$ and coordinate-wise training 95th–5th percentile ranges. Endpoints are clipped to the combined training–test coordinate bounds. Global secants join distinct held-out inputs. Constructed pairs probe the ambient input space; their violation rates describe these audit banks rather than violation frequencies under the observed data distribution.

OPVR is the fraction of comparable pairs $(\mathbf{x}_0, \mathbf{x}_1)$ with $\mathbf{x}_0 \preceq_M \mathbf{x}_1$ whose representation order is not preserved, whereas FCR is the fraction of input-incomparable pairs that become comparable in either direction. Both rates use coordinate-wise absolute and relative numerical tolerances of 10^{-6} when testing representation order. Writing $\Delta z = z(\mathbf{x}_1) - z(\mathbf{x}_0)$, the normalized incomparability margin is

$$m(\Delta z) = \begin{cases} \dfrac{\min\{\max_j(-\Delta z_j),\, \max_j(\Delta z_j)\}}{\|\Delta z\|_2}, & \|\Delta z\|_2 > 0, \\ 0, & \|\Delta z\|_2 = 0. \end{cases} \tag{44}$$

A positive value requires a coordinate conflict in both directions. For pair family $F \in \{\text{context}, \text{mixed}\}$, evaluation split f, and final seed r, the run-level margin is the empirical fifth percentile $M_F^{(f,r)} = Q_{0.05}(\{m(\Delta z_k) : k \in F\})$.

For a local or global bank B of N_B generated secants, let $\mathcal{R}_B^+$ be the multiset of positive finite ratios $\|\Delta z\|_2 / \|\Delta \mathbf{x}\|_2$ from pairs with positive finite input distance. When $\mathcal{R}_B^+$ is nonempty, define

[1] https://github.com/yuri-k111/Bi-Lipschitz-Neural-Network

$D_{90}(B) = Q_{0.95}(\mathcal{R}_B^+)/Q_{0.05}(\mathcal{R}_B^+)$; otherwise the run is invalid. Thus, D_{90} measures central spread conditional on a positive finite response and is not a collapse metric. Multiplying all positive distance ratios by a common factor leaves D_{90} unchanged; uniformly tiny positive ratios can therefore still give $D_{90} = 1$. All 150 matched-comparison AOBE runs retain all $N_B = 1{,}024$ ratios in both banks; some unanchored runs retain fewer positive ratios and have the same conditional interpretation. Run-level metrics are averaged over splits within each final seed, and dataset-level means and sample standard deviations are computed across the five final seeds 42–46. Table 2 averages the six dataset-level order-audit means without pooling pairs. Figure 4 reports the median of the six dataset-level D_{90} means in Table 7.

Direct collapse and task-output sensitivity. The collapse audit reuses the 600 matched checkpoints and retains zero ratios $r_k = \|\Delta z_k\|_2/\|\Delta \mathbf{x}_k\|_2$ for positive finite input distances. A numerical collision is recorded when the computed ratio satisfies $r_k = 0$ in floating-point arithmetic; raw near-collapse rates use thresholds 10^{-6} and 10^{-3}. Calibration-normalized rates use $\widetilde{r}_k = r_k/c_{\mathrm{cal}}$ with thresholds 0.01 and 0.001, where c_{cal} is the median finite ratio, including zeros, on the training-input calibration bank. If this median is zero, normalized statistics are undefined rather than assigned zero. The fixed-context bank combines comparable and mixed pairs (2,048 pairs); the other reported families contain 1,024 pairs. Overlapping families are not treated as independent samples. For the anchor and full AOBE, lower-bound violations satisfy $r_k < \underline{d} - 10^{-6}$. For $c_{\mathrm{cal}} > 0$, the raw lower bound gives $\widetilde{r}_k \geq \underline{d}/c_{\mathrm{cal}}$; whether this exceeds a chosen normalized near-collapse threshold also depends on the calibration scale.

The score-stability audit reuses the seven main-comparison scorers and the separately selected no-norm-control AOBE variant, totaling 1,200 checkpoints across six datasets, five splits, and five final seeds. Each checkpoint is evaluated on 1,024 label-free pairs from each of four families: fixed-context constrained changes, context-only changes, joint local changes, and joint global secants. Per-pair score sensitivity is $|\Delta s|/\|\Delta \mathbf{x}\|_2$. Task-output sensitivity replaces the numerator by the absolute change in positive-class probability or original-scale regression prediction, or the maximum absolute change over ordinal upper-tail probabilities. For each sensitivity measure, the run-level statistic is the 95th percentile, averaged over splits within each seed and then over seeds. For AOBE, the analytic sensitivity bounds are L_s for the score, $L_s/4$ for binary and ordinal outputs, and $\varsigma_{\mathrm{tr}} L_s$ for regression. All 1,024 pairs are retained in every evaluated checkpoint–family unit. A per-pair bound violation is recorded when sensitivity exceeds $B(1 + 10^{-5})$, where B is the corresponding analytic bound. Dataset-specific output units are not pooled across tasks.

C.3 Main and parameter-matched audits

Table 4 reports parameter counts for the main comparison; the largest deviation from the common target $P_{\mathrm{target}}(d) = 3\max\{3{,}000, 32(d+1)\}$ is 3.82%. Relative to each comparator within a dataset, AOBE has between 4.6% fewer and 4.9% more parameters and is the largest model on only three of six datasets. Its competitive predictive results are thus obtained under comparable parameter budgets, without a consistent model-size advantage.

Table 4: Realized trainable parameter counts in the main comparison. Counts are fixed within each dataset–method cell and shared across splits and final seeds.

Dataset	MLP	Pos-MLP	LMN	SMNN	SMM	PAS-MLP	AOBE
COMPAS	9,010	9,010	8,743	8,669	8,999	8,801	9,045
Heart Disease	9,028	9,028	8,759	8,695	8,998	9,088	9,119
BlogFeedback	26,651	26,651	26,113	27,001	26,715	26,651	26,399
Auto MPG	9,025	9,025	9,025	9,247	9,003	9,062	9,049
Wine Quality	8,889	8,889	9,158	9,344	9,001	9,149	8,916
Cleveland	9,014	9,014	8,747	8,673	8,998	8,805	9,047

Table 5 reports zero coordinate-wise and joint scalar MVRs for every constrained method on all six datasets and in every constituent run. MLP has positive rates in both audits, with lower joint than coordinate-wise rates on every dataset. Reporting both perturbation types makes this difference visible while confirming the shared forward-monotone behavior of the constrained implementations on the audit bank.

Table 5: Dataset-level coordinate-wise and joint scalar MVRs. Values are mean ± sample standard deviation across five seed-level averages, each formed over five splits.

(a) Coordinate-wise MVR ↓

Method	COMPAS	Heart Disease	Blog Feedback	Auto MPG	Wine Quality	Cleveland
MLP	0.2075 ± 0.0258	0.3534 ± 0.1404	0.3150 ± 0.0342	0.1769 ± 0.0259	0.3172 ± 0.0131	0.4458 ± 0.0554
Pos-MLP	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
LMN	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
SMNN	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
SMM	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
PAS-MLP	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
AOBE	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000

(b) Joint MVR ↓

Method	COMPAS	Heart Disease	Blog Feedback	Auto MPG	Wine Quality	Cleveland
MLP	0.0035 ± 0.0044	0.3192 ± 0.1882	0.0018 ± 0.0016	0.0104 ± 0.0078	0.0027 ± 0.0030	0.0516 ± 0.0214
Pos-MLP	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
LMN	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
SMNN	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
SMM	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
PAS-MLP	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000
AOBE	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000

Parameter-matched protocol. The study uses 16 predeclared optimizer configurations and paired bound assignments, activating only the bounds applicable to each embedding, with tuning seed 2026, a 500-epoch limit, and patience 60. The four embeddings use output dimension $d_M + 2d_U + 64$ and the same positive norm-controlled scalar readout, task head, loss, and pair banks within each dataset; their parameter counts differ from AOBE by at most 1.10%. Structural metrics are not used for hyperparameter selection. Table 6 reports the resulting predictive metrics.

Table 6: Primary predictive metrics in the parameter-matched representation comparison. Values are mean ± sample standard deviation across the five seed-level split averages.

Dataset and metric	Unconstrained	Partially monotonic	Norm-controlled monotonic	AOBE
COMPAS (AUROC ↑)	0.7221 ± 0.0243	$\mathbf{0.7355 \pm 0.0008}$	0.7344 ± 0.0009	0.7338 ± 0.0009
Heart Disease (AUROC ↑)	0.8963 ± 0.0118	0.7953 ± 0.0930	$\mathbf{0.9089 \pm 0.0064}$	0.8921 ± 0.0064
BlogFeedback (RMSE ↓)	26.166 ± 0.235	26.684 ± 0.329	26.193 ± 0.240	$\mathbf{26.081 \pm 0.333}$
Auto MPG (RMSE ↓)	$\mathbf{2.853 \pm 0.094}$	2.966 ± 0.137	2.965 ± 0.013	2.945 ± 0.035
Wine Quality (RPS ↓)	0.06573 ± 0.00039	0.06671 ± 0.00084	0.06575 ± 0.00028	$\mathbf{0.06569 \pm 0.00077}$
Cleveland (RPS ↓)	0.11356 ± 0.01358	0.13255 ± 0.00638	0.10346 ± 0.00483	$\mathbf{0.09438 \pm 0.00116}$

Table 6 shows that AOBE has the best predictive mean on BlogFeedback, Wine Quality, and Cleveland and the second-best on Auto MPG. Under the common dimension and capacity controls, Table 2 shows that AOBE has the lowest average predictive rank alongside zero audited OPVR and FCRs and positive margins.

Table 7 reports the central spread of positive finite distance ratios. AOBE has the smallest local D_{90} on five datasets and global D_{90} on four; the unconstrained embedding is smaller on Auto MPG in both banks and on Wine Quality globally.

Figure 4 summarizes the cross-dataset medians. AOBE obtains local and global median D_{90} values of 6.88 and 6.93, respectively, below those of all three unanchored embeddings; the corresponding norm-controlled monotonic values are 63.13 and 48.84. This comparison shows that AOBE has a smaller central secant spread in both audited regimes. Because D_{90} is computed from positive finite ratios, Table 8 separately evaluates numerical collisions and near-collapse.

Direct collapse results. Across all audited families, full AOBE and its anchor channel have zero numerical collisions, zero near-collapse at both raw and both calibration-normalized thresholds, and

Table 7: Dataset-level local and global D_{90} on positive finite secant ratios in the parameter-matched representation comparison. Entries are means across five seed-level split averages; lower is better. Bold marks the smallest mean within each dataset and secant type.

Embedding	Secant type	COMPAS	Heart Disease	Blog Feedback	Auto MPG	Wine Quality	Cleveland
Unconstrained	Local	2.71×10^{3}	1.74×10^{3}	14.68	**17.20**	10.88	2.26×10^{4}
	Global	63.49	4.98×10^{4}	9.82	**6.18**	**4.40**	849.95
Partially monotonic	Local	48.60	1.13×10^{5}	25.30	182.98	41.64	1.21×10^{3}
	Global	60.11	4.42×10^{4}	16.76	19.47	17.55	610.12
Norm-controlled monotonic	Local	30.66	157.91	83.87	42.40	29.09	118.90
	Global	41.60	64.98	56.07	15.16	23.77	60.56
AOBE	Local	**9.81**	**3.97**	**1.59**	25.83	**9.27**	**4.49**
	Global	**14.00**	**2.99**	**1.46**	11.05	9.64	**4.22**

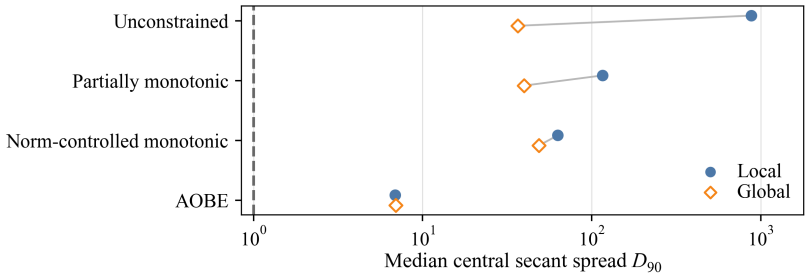


Figure 4: Median local and global D_{90} across the six dataset-level means in Table 7. Lower is better; the dashed line at $D_{90} = 1$ marks the minimum possible central spread.

zero lower-bound violations. Their minimum observed distance ratios divided by the corresponding $\underline{d}$ are 1.154 and 1.148, respectively, both exceeding one.

Table 8 reports the fixed-context and local audits. The nonlinear channel and the unanchored embeddings exhibit numerical collisions or near-collapse on these banks. Notably, the norm-controlled unanchored embedding has almost no exact collisions but still exhibits normalized near-collapse rates of 3.110% and 0.799%, respectively, illustrating that upper norm control alone does not prevent strong contraction. In contrast, the anchor and full AOBE have zero rates in both audits, even where the nonlinear channel collapses numerically. These results directly support the anchor's role in maintaining input separation in the complete embedding, complementing the central-spread information provided by D_{90}.

Table 8: Collision and normalized near-collapse rates (%), averaged over splits, seeds, and datasets in that order. Near-collapse uses $\widetilde{r} \leq 0.01$. Entries use 150 checkpoints; starred entries exclude one undefined normalized result.

	Fixed context		Local	
Representation	Collision ↓	Near-collapse ↓	Collision ↓	Near-collapse ↓
Unconstrained	0.708	2.736*	1.453	3.068*
Partially monotonic	2.236	5.791	4.626	8.124
Norm-controlled monotonic	0.000	3.110	0.001	0.799
AOBE nonlinear channel	0.062	2.853	1.180	1.724
AOBE anchor channel	0.000	0.000	0.000	0.000
Full AOBE	0.000	0.000	0.000	0.000

Main-comparison AOBE audits. Table 9 reports dataset-level structural results for the 150 main-comparison AOBE checkpoints. All have zero OPVR, Context-FCR, and Mixed-FCR and positive normalized fifth-percentile margins; the run-wise context and mixed minima are 0.0083 and 0.0039, respectively. All checkpoints pass the embedding lower- and upper-bound checks and the scalar-score upper-bound check.

Table 9: Finite-sample structure audit of the main AOBE checkpoints. Values are mean ± sample standard deviation across five seed-level split averages.

Audit metric	COMPAS	Heart Disease	Blog Feedback	Auto MPG	Wine Quality	Cleveland
OPVR	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
Context-FCR	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
Mixed-FCR	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
Context margin	0.111 ± 0.010	0.212 ± 0.005	0.157 ± 0.007	0.0666 ± 0.004	0.153 ± 0.038	0.155 ± 0.006
Mixed margin	0.0857 ± 0.012	0.661 ± 0.016	0.0381 ± 0.003	0.151 ± 0.029	0.0294 ± 0.004	0.295 ± 0.005
Local D_{90}	11.2 ± 1.2	2.01 ± 0.10	1.42 ± 0.11	25.8 ± 3.4	9.01 ± 1.1	3.62 ± 0.09
Global D_{90}	14.6 ± 0.98	2.31 ± 0.06	1.43 ± 0.11	11.1 ± 0.94	9.33 ± 0.91	4.00 ± 0.18

C.4 Ablation and synthetic results

Real-data mechanism study. Full AOBE reuses its main-comparison split-specific configuration; reduced variants are selected from 16 candidates using predictive validation criteria under the same training budget. The variants change model capacity and are evaluated as mechanism studies. Table 10 reports their primary test performance. Full AOBE improves on no-norm-control on five tasks and anchor-only on four, while anchor-only performs better on Heart Disease and BlogFeedback.

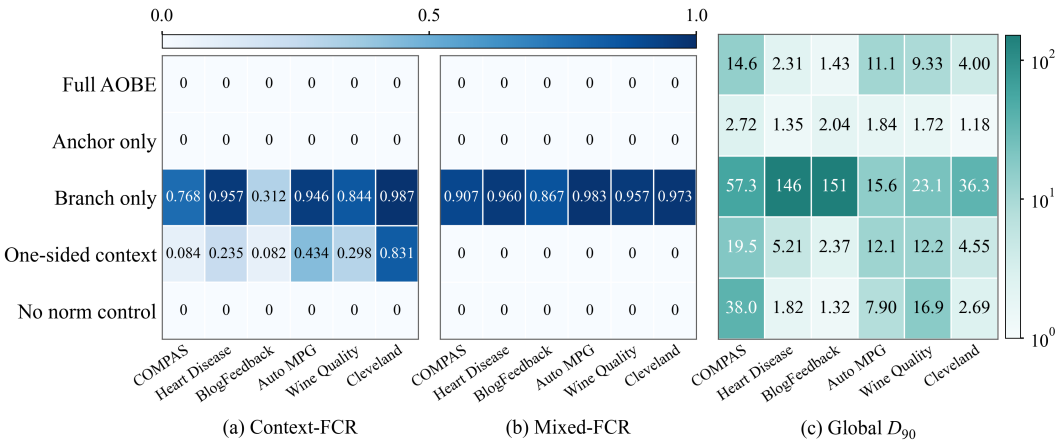
Table 10: Primary test performance of the real-data architectural ablations. Values are mean ± sample standard deviation across five seed-level split averages.

Dataset / metric	Full AOBE	Anchor only	Branch only	One-sided context	No norm control
COMPAS AUROC ↑	0.7324 ± 0.0009	0.7271 ± 0.0002	$\mathbf{0.7343 \pm 0.0008}$	0.7321 ± 0.0007	0.7341 ± 0.0009
Heart Disease AUROC ↑	0.8993 ± 0.0066	$\mathbf{0.9060 \pm 0.0054}$	0.8732 ± 0.0448	0.9038 ± 0.0037	0.8964 ± 0.0154
BlogFeedback RMSE ↓	26.256 ± 0.161	$\mathbf{25.391 \pm 0.012}$	26.457 ± 0.119	25.881 ± 0.138	26.936 ± 0.268
Auto MPG RMSE ↓	$\mathbf{2.945 \pm 0.035}$	3.454 ± 0.018	2.977 ± 0.019	2.983 ± 0.013	2.966 ± 0.148
Wine Quality RPS ↓	$\mathbf{0.0651 \pm 0.0007}$	0.0667 ± 0.0001	0.0652 ± 0.0007	0.0655 ± 0.0006	0.0663 ± 0.0007
Cleveland RPS ↓	0.0957 ± 0.0010	0.0990 ± 0.0000	0.0953 ± 0.0010	$\mathbf{0.0936 \pm 0.0014}$	0.0963 ± 0.0023

Figure 5 identifies the anchor's contribution to order reflection. Full AOBE and anchor-only have zero Context-FCR and Mixed-FCR on all six datasets. Branch-only has Context-FCRs of 0.312–0.987, Mixed-FCRs of 0.867–0.983, and larger global D_{90} than full AOBE on every dataset. Replacing the signed context pair with a one-sided anchor yields Context-FCRs of 0.082–0.831 while retaining zero Mixed-FCR, identifying the signed pair's specific role in cross-context incomparability. Removing norm control retains zero audited FCRs and the analytic order-isomorphism and lower-bound guarantees, but removes the parameter-uniform upper bound. Its global D_{90} is lower than full AOBE on four datasets and higher on two, showing a task-dependent effect on empirical distance scaling.

Norm control bounds expansion uniformly over the input domain and admissible parameters, whereas D_{90} summarizes the relative spread of sampled distance ratios. Together, these ablations distinguish the positive constrained-coordinate anchor's role in reflecting the constrained-coordinate order, the signed context pair's role in cross-context incomparability, and branch norm control's role in parameter-uniform upper Lipschitz control.

Controlled synthetic study. Inputs are sampled uniformly from $[-1, 1]^d$. Responses combine nonlinear monotone bases with positive context-dependent slopes, a nonlinear context term, and Gaussian noise. The 12 settings cross $(d_M, d_U) \in \{(2, 2), (4, 4), (8, 8)\}$, noise standard deviations $\{0, 0.2\}$, and interaction strengths $\{0.5, 1.5\}$. Each uses $2{,}048/512/1{,}024$ training/validation/test samples, fixed across training seeds 42–46. All five variants use a predeclared configuration, at most 400 epochs, and validation-loss early stopping. We report $\mathrm{NRMSE} = \mathrm{RMSE} \,/\, \mathrm{sd}_0(y_{\mathrm{test}})$, where sd_0 denotes population standard deviation.



(a) Context-FCR (b) Mixed-FCR (c) Global D_{90}

Figure 5: Real-data representation-level mechanism audits: (a) Context-FCR, (b) Mixed-FCR, and (c) global D_{90}. Each cell is the mean across five seed-level averages, each formed over five splits. The FCR panels share a linear $[0, 1]$ scale; the D_{90} panel uses a logarithmic scale.

Table 11: Controlled synthetic component analysis. Order reflection and uniform bi-Lipschitzness refer to guarantees for all admissible embedding parameter values.

Variant	Nonlinear branch	Order reflection	Uniform bi-Lipschitz	NRMSE ↓	OPVR ↓	Context-FCR ↓	Mixed-FCR ↓	Global D_{90} ↓
Full AOBE	Yes	Yes	Yes	0.447	**0.000**	**0.000**	**0.000**	1.323
Anchor only	No	Yes	Yes	0.569	**0.000**	**0.000**	**0.000**	**1.257**
Branch only	Yes	No	No	0.541	**0.000**	1.000	1.000	25.151
One-sided context	Yes	No	Yes	0.459	**0.000**	0.398	**0.000**	1.409
No norm control	Yes	Yes	No	**0.383**	**0.000**	**0.000**	**0.000**	4.453

Table 11 separates predictive contribution from structural guarantees. Adding the nonlinear branch to anchor-only reduces NRMSE from 0.569 to 0.447, while retaining zero FCRs and uniform bi-Lipschitzness; global D_{90} increases only from 1.257 to 1.323. The learned branch thus improves prediction within the anchored construction with a modest increase in empirical secant spread.

All five variants have zero audited OPVR, but branch-only makes every tested context and mixed incomparable pair comparable, showing that forward order preservation does not prevent false comparability. The one-sided context variant retains uniform bi-Lipschitzness but loses cross-context incomparability, separating distance control from conditional-order fidelity. Removing norm control yields the lowest NRMSE but substantially increases global D_{90} and removes the parameter-uniform upper bound. Together, these results show the complementary roles of learned nonlinear features, the signed anchor, and norm control, alongside an empirical accuracy–geometry trade-off.

C.5 Frozen-channel contribution and score stability

Frozen-channel readouts. Readouts on the frozen encoders share the common task head and inherit the full configuration's learning rate, weight decay, score-norm radius, and training budget, without channel-specific tuning or parameter matching. Test evaluation follows fitting and validation checkpoint selection. Two-sided 95% Student-t intervals are computed from five seed-level means of paired split differences, conditional on the fixed splits and without multiple-comparison adjustment. This frozen-encoder analysis complements the separately retrained architectural ablations in Appendix C.4.

Table 12 shows that the full view has the best predictive mean on all six datasets, with full and nonlinear tied at the displayed precision on Wine Quality. The channel contributions differ across tasks. On Auto MPG, adding the nonlinear features reduces RMSE from 5.812 for the affine anchor score to 2.971 for full; substantial improvements over anchor also occur on both ordinal datasets. Conversely, full improves on the nonlinear-only view on COMPAS, Heart Disease, and BlogFeed-

back, indicating predictive utility of the anchor features in addition to their structural role. Figure 3 quantifies these task-dependent contributions through paired comparisons.

Table 12: Readouts refitted on frozen full-AOBE encoders. Values are mean ± sample standard deviation over five seed-level split averages.

Dataset	Metric	Anchor channels	Nonlinear channels	Full channels
COMPAS	AUROC ↑	0.6864 ± 0.0029	0.7179 ± 0.0012	$\mathbf{0.7327 \pm 0.0009}$
Heart Disease	AUROC ↑	0.9009 ± 0.0035	0.8865 ± 0.0186	$\mathbf{0.9032 \pm 0.0084}$
BlogFeedback	RMSE ↓	26.390 ± 0.084	28.259 ± 0.467	$\mathbf{26.179 \pm 0.316}$
Auto MPG	RMSE ↓	5.812 ± 0.017	3.040 ± 0.048	$\mathbf{2.971 \pm 0.040}$
Wine Quality	RPS ↓	0.0814 ± 0.0001	$\mathbf{0.0651 \pm 0.0009}$	$\mathbf{0.0651 \pm 0.0009}$
Cleveland	RPS ↓	0.1222 ± 0.0002	0.0931 ± 0.0004	$\mathbf{0.0930 \pm 0.0008}$

Score and task-output stability. Across the 150 main-AOBE checkpoints and four perturbation families, all 600 audit units pass the score and task-output upper-bound checks. Table 13 reports task-output sensitivity ratios between AOBE and separately validation-selected no-norm-control variants, using the split–seed aggregation defined in Appendix C.2.

AOBE has lower task-output sensitivity in $16/24$ dataset–family comparisons, while the separate score audit shows lower sensitivity in $17/24$. Task-output sensitivity is lower in all four families on COMPAS and in three families on Heart Disease, Auto MPG, Wine Quality, and Cleveland, whereas BlogFeedback has higher sensitivity in all four families. Thus, AOBE combines certified Lipschitz control of task-output changes with lower observed sensitivity in most comparisons. The upper certificate constrains each model individually and does not order the realized sensitivities of separately fitted models.

Table 13: Task-output sensitivity ratios (AOBE/no-norm-control). Sensitivity is the split–seed average of per-run 95th percentiles; values below one indicate lower observed sensitivity for AOBE.

Dataset	Context only	Fixed-context constrained	Joint local	Joint global
COMPAS	0.938	0.749	0.775	0.937
Heart Disease	0.735	1.768	0.711	0.876
BlogFeedback	1.632	1.312	1.126	1.081
Auto MPG	0.898	0.992	0.998	1.009
Wine Quality	0.645	1.003	0.779	0.987
Cleveland	0.948	0.843	0.839	1.036