PROBABILISTIC HARD-CONSTRAINED LEARNING VIA STRUCTURAL LATENT PARAMETERIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Many probabilistic predictors must satisfy exact structure in every stochastic realization, yet common hard-constraint approaches form predictions in ambient coordinates and subsequently complete, correct, or project them. We introduce `SLP-ProbHard`, a cross-family representation-centered framework for probabilistic hard-constrained learning when suitable explicit structural parameterizations are available. Its core modelling object is a Structural Feasible Latent Parameterization (*SFLP*), consisting of a structural latent law $Z \sim P_\theta^Z(\cdot \mid x)$ and an explicit feasible map $Y = h_\phi(x, Z)$, with $h_\phi(x, z) \in \mathcal{C}(x)$ for every $z \in \mathcal{Z}$. Together, the structural latent law and feasible map define the predictive model itself rather than a final feasibility wrapper. The latent law and map jointly determine predictive support and boundary probability; the chosen coordinates and map specify the stochastic representation and shape attainable dependence, calibration, expressiveness, and computation. The formulation does not prescribe a particular predictive backbone or latent distribution family; the present experiments instantiate it with Gaussian latent laws and fixed geometry-derived maps. We instantiate the formulation across representative affine equalities, convex structural sets (ordering, simplex, and nonnegative ordering), explicit nonlinear equalities/manifolds (circle and scale-shape), and three structural views of the same seven-basin hydrological FDC data. A verified official-source ProbHardE2E/DPPL affine comparison gives zero practical violations for both methods: `SLP-ProbHard` uses 8 rather than 11 stochastic coordinates and improves MSE/MAE, whereas DPPL attains better marginal CRPS and higher, closer-to-nominal coverage; the paired test does not detect an Energy Score difference across ten seeds. Real-world affine and nonlinear FDC representations reduce 13 to 7 and 14 ambient to 8 computational coordinates, respectively. Together, these results show that exact feasibility alone does not determine a predictive law and motivate direct structural generation when scientifically meaningful feasible coordinates are available.

1 INTRODUCTION

Probabilistic models are increasingly used where uncertainty matters but output structure is non-negotiable. In these settings feasibility must hold not only for a predictive mean but for every stochastic realization. Examples include conservation and aggregation identities, non-negativity, ordering, simplex constraints, explicit nonlinear equalities, composed constraints, and initial or boundary conditions.

A common strategy is to predict in an ambient space and then enforce feasibility through a soft penalty, post-processing rule, optimization layer, or differentiable projection/correction (Amos & Kolter, 2017; Agrawal et al., 2019; Donti et al., 2021). Recent methods broaden this route to convex, semidefinite, nonlinear equality, and nonlinear inequality constraints (Tordesillas et al., 2023; Grontas et al., 2025; Tang et al., 2026; Goertzen et al., 2026; Chu et al., 2026; Wang et al., 2026); ProbHardE2E is the closest dedicated probabilistic projection framework (Utkarsh et al., 2026). Projection is broadly useful, especially when a feasible set is implicit or difficult to parameterize.

`SLP-ProbHard` asks the complementary question: *when a scientifically meaningful generative representation is available, can the predictive distribution be defined directly in those feasible*

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.

coordinates? We use

$$Z \sim P_\theta^Z(\cdot \mid x), \qquad Y = h_\phi(x, Z), \qquad h_\phi(x, z) \in \mathcal{C}(x) \quad \forall z \in \mathcal{Z}. \tag{1}$$

so the output law is the pushforward $P_\theta^Y = h_{\phi,x} \# P_\theta^Z$. The structural map is therefore not merely a final feasibility wrapper: together with the latent law it specifies which feasible outputs can occur, which boundary events can receive probability, and how many stochastic coordinates are actually modeled.

Novelty scope. `SLP-ProbHard` introduces a cross-family representation-centered framework in which a structural latent law and an explicit feasible map jointly define the predictive law. Its methodological contribution is to formalize and test how this representation changes feasible support and boundary probability, stochastic-coordinate structure, uncertainty propagation, attainable dependence/covariance, expressiveness, calibration, and computation. Thus, two constructions can satisfy the same hard constraint while defining different predictive-law families. The individual feasibility constructions used here, null-space coordinates, nonnegative increments, normalization, trigonometric coordinates, and the associated pushforward identities, use established mathematical primitives. Important structural-probabilistic precedents also exist in hierarchical forecasting, general linear constraints, hard linear generative learning, and domain-specific stochastic ordering (Olivares et al., 2024; Girolimetto & Di Fonzo, 2024; Li et al., 2026; Bekele & D'Oria, 2026). The contribution of `SLP-ProbHard` is not these primitives individually, but their organization as a common probabilistic model-design methodology across representative affine, convex-structural, and explicit nonlinear constraint geometries.

Contributions. We make three connected contributions.

(1) Predictive-law abstraction. We introduce `SLP-ProbHard` and *SFLP* as its core modelling abstraction: a structural latent predictive law and an explicit feasible map jointly define a sample-wise hard-feasible pushforward law.

(2) Representation audit and consequences. We characterize and empirically validate how *SFLP* choice affects support and boundary probability, intrinsic versus computational stochastic dimension, uncertainty propagation, attainable dependence/covariance, calibration, parameterization bias, and computation across representative constraint geometries.

(3) Cross-family empirical validation. We evaluate the framework on affine equality, ordering, simplex, composed nonnegative ordering, explicit nonlinear manifold/scale–shape problems, and three representations of the same hydrological FDC data. A verified official-source ProbHardE2E/DPPL comparison on the compatible affine benchmark demonstrates shared practical hard feasibility but different stochastic representations and representation-dependent predictive trade-offs.

2 Related Work

Projection and correction. A major branch predicts in an ambient space and enforces feasibility through differentiable optimization, completion/correction, radial maps, or projection layers (Amos & Kolter, 2017; Agrawal et al., 2019; Donti et al., 2021; Liang et al., 2023; 2024; Schneider & Kuhn, 2026). ProbHardE2E (Utkarsh et al., 2026) is the closest dedicated probabilistic method in this branch: it constructs hard-constrained predictive laws through differentiable probabilistic projection.

Direct feasible parameterization. Another branch constructs outputs directly in feasible coordinates, including null-space and inequality parameterizations, convex feasible layers, and hard-constrained network architectures (Frerix et al., 2020; Tordesillas et al., 2023; Min et al., 2024; Nguyen & Donti, 2025). A domain-specific probabilistic precedent is SMT-HCSD (Bekele & D'Oria, 2026), which generates stochastically ordered temperature outputs using a nonnegative increment.

Structural and probabilistic constrained laws. Several works are particularly close to the probabilistic structural viewpoint. CLOVER (Olivares et al., 2024) constructs coherent probabilistic forecasts by construction through a differentiable structural factor representation for hierarchical forecasting. Girolimetto and Di Fonzo (Girolimetto & Di Fonzo, 2024) develop a structural-like formulation that separates free and constrained variables for general linearly constrained multiple time series and extend it to probabilistic forecast reconciliation. Li et al. (2026) develop probabilistically sound deep generative learning under hard linear equality constraints.

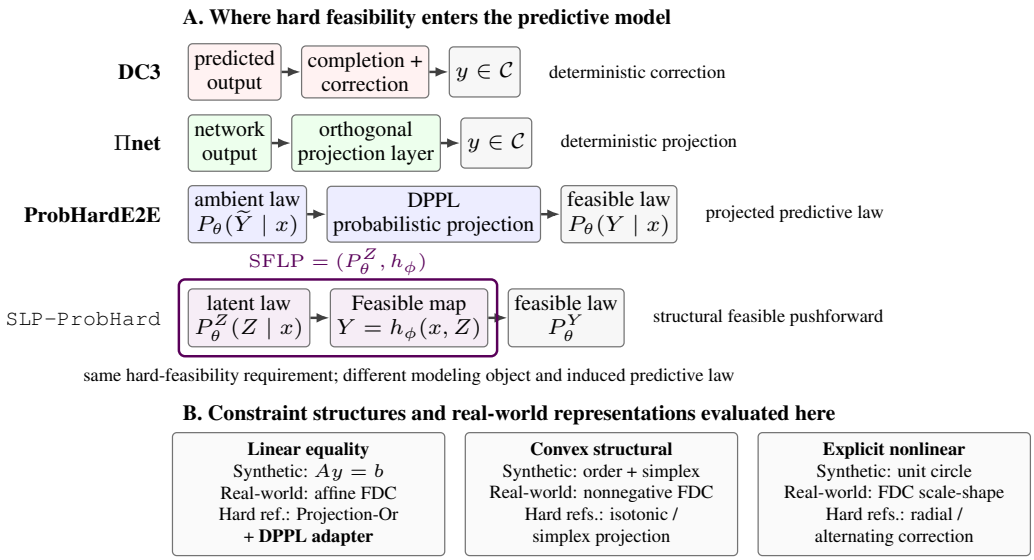

Figure 1: SLP-ProbHard **positioning and empirical scope. Panel A** contrasts representative locations at which hard feasibility enters learning: DC3 uses completion/correction (Donti et al., 2021), Πnet uses orthogonal projection layers (Grontas et al., 2025). At the probabilistic level, ProbHardE2E/DPPL and SLP-ProbHard/*SFLP* are parallel hard-constraint paradigms: DPPL constructs a feasible law by projecting an ambient predictive law (Utkarsh et al., 2026), whereas an *SFLP* defines the feasible law directly as the pushforward of its structural latent law. DC3 and Πnet are shown for conceptual positioning only; the direct official-source empirical comparison in this paper is with ProbHardE2E/DPPL on the compatible affine benchmark. **Panel B** summarizes the representative linear, convex-structural, and explicit nonlinear geometries evaluated synthetically and through three structural views of the same hydrological FDC data.

Position of SLP-ProbHard**.** These precedents address hierarchical forecasting (CLOVER), general linear time-series constraints (Girolimetto–Di Fonzo), hard linear generative learning (Li et al.), and stochastic temperature ordering (SMT-HCSD). The present study brings these structural-probabilistic perspectives into a common representation audit spanning affine, ordering, simplex, and explicit nonlinear geometries. SLP-ProbHard is organized around a complementary cross-family question: *when the feasible structural representation itself is chosen as the coordinate system of the predictive law, how does that choice affect support and boundary probability, stochastic-coordinate structure, uncertainty propagation, expressiveness/calibration, and computation?* Section 5 gives the compatible official-source ProbHardE2E/DPPL affine-adapter comparison, while Appendix A provides the broader mechanism-level comparison.

3 SLP-PROBHARD

3.1 STRUCTURAL FEASIBLE PREDICTIVE LAWS

Let $x \in \mathcal{X}$, $y \in \mathcal{Y} \subseteq \mathbb{R}^d$, let $\mathcal{Z}$ be the latent space, and let $\mathcal{C}(x)$ be the sample-level feasible set. Within SLP-ProbHard, an *SFLP* is the pair formed by a structural latent predictive law and a feasible structural map:

$$Z \sim P_\theta^Z(\cdot \mid x), \qquad Y = h_\phi(x, Z), \qquad h_\phi(x, z) \in \mathcal{C}(x) \quad \forall z \in \mathcal{Z}.$$

The induced predictive law is $P_\theta^Y = h_{\phi,x} \# P_\theta^Z$. Together, the latent law and map determine the family of feasible predictive distributions that can be represented. In all reported experiments, h_ϕ is fixed by known constraint geometry; learnable structural maps are not evaluated here.

When an output density is inconvenient, proper sample-based scores provide a generic training interface (Gneiting & Raftery, 2007; Jordan et al., 2019). Reparameterized sampling permits sam-

Tested geometry	Synthetic benchmark	Real-world representation	Primary SFLP	Hard comparator
Linear equality	hierarchy $Ay = b$	affine FDC coherence	$y_0 + Nz$; affine $(Q, \mathrm{diff}\ Q)$	orthogonal projection; **DPPL official-source adapter** (synthetic only)
Convex structural instances	weak order; simplex	nonnegative FDC ordering	cumulative / normalized positive-part	isotonic / simplex projection
Explicit nonlinear equality/manifold	unit circle	FDC scale-shape $\Delta_i = RP_i$	angle / composed scale-shape	radial projection / alternating feasible correction

Table 1: Empirical constraint coverage. These experiments evaluate `SLP-ProbHard` on representative structures from the listed geometry groups. The direct official-source DPPL comparison is restricted to the synthetic affine benchmark.

plewise backpropagation; interpreting this as a gradient of the expected score requires the regularity conditions discussed in Appendix E.4 (Kingma & Welling, 2014; Figurnov et al., 2018).

For affine equalities $Ay = b$, choose a feasible point y_0 and full null-space basis $N \in \mathbb{R}^{d\times q}$ with $q = d - m$, $m = \mathrm{rank}(A)$, $Ay_0 = b$, and $AN = 0$:

$$Y = y_0 + NZ, \qquad Z \in \mathbb{R}^q. \tag{2}$$

Here q is both the intrinsic and computational stochastic dimension. Other maps may be redundant, so we keep intrinsic degrees of freedom separate from computational latent dimension.

Algorithm 1: `SLP-ProbHard` **- design, train, and generate.**
A. SFLP design/audit:
(1) specify the sample-level feasible set $\mathcal{C}(x)$;
(2) choose structural coordinates and h_ϕ;
(3) audit the *SFLP*: verify $h_\phi(x, \mathcal{Z}) \subseteq \mathcal{C}(x)$ and record (i) map image and coverage/surjectivity limitations; (ii) boundary preimages and whether they have zero or positive probability under the selected latent law; (iii) intrinsic dimension q_{int} and computational latent dimension q_{comp}; (iv) sensitivity, redundancy, and induced dependence/covariance restrictions; (v) density/Jacobian tractability where applicable; and (vi) whether generation requires a separate optimization-based projection or correction stage.
B. Probabilistic training:
(4) predict latent parameters $\eta_\theta(x)$;
(5) draw reparameterized $z^{(s)} \sim P_\theta^Z(\cdot \mid x)$;
(6) construct $\widehat{y}^{(s)} = h_\phi(x, z^{(s)})$;
(7) compute the selected score or likelihood and its training gradient, with the estimator qualifications in Appendix E.4.
C. Inference:
(8) sample Z;
(9) apply h_ϕ and return feasible predictive samples without a separate correction stage.
Scope: If no useful explicit *SFLP* is available, `SLP-ProbHard` does not assume that structural parameterization is preferable to projection/correction.

This audit makes representation choice operational: feasible constructions can differ in coverage, boundary probability, stochastic coordinates, induced law families, and computation.

3.2 Representation-induced probabilistic consequences

The *SFLP* audit exposes probabilistic consequences beyond sample-level feasibility, using three standard mathematical facts. First, if $h_x(\mathcal{Z}) \subseteq \mathcal{C}(x)$, then the pushforward satisfies $P_\theta^Y(\mathcal{C}(x) \mid x) = 1$. Second, rank-nullity gives $q = d - m$ free coordinates for a consistent affine system. Third, for any measurable event B,

$$P^Y(B) = P^Z\big(h^{-1}(B)\big). \tag{3}$$

Two representation-level consequences. Equation (3) makes boundary probability representation-dependent: a scientifically meaningful boundary event can be excluded, reachable with zero probability, or receive positive predictive mass according to its latent preimage. A second axis is stochastic representation: we distinguish the intrinsic degrees of freedom q_{int} from the computational latent

dimension q_{comp}. They coincide for affine null-space coordinates and the circle angle map, need not decrease for ordering, and differ in the boundary-capable simplex and nonlinear FDC implementations. Dimensional economy is therefore map-dependent rather than a universal property of `SLP-ProbHard`.

The representation-level probabilistic analysis asks what a chosen map does *beyond* feasibility. Using standard truncated-Gaussian calculations (Tallis, 1961), for $R = [Z]_+$ and $Z \sim \mathcal{N}(\mu, \sigma^2)$, with $\sigma > 0$ and $\alpha = \mu/\sigma$,

$$\Pr(R = 0) = \Phi(-\alpha), \qquad \mathbb{E}[R] = \sigma\varphi(\alpha) + \mu\Phi(\alpha). \tag{4}$$

For the following product formulas, assume conditionally independent $Z_i \mid x \sim \mathcal{N}(\mu_i(x), \sigma_i^2(x))$ with $\sigma_i(x) > 0$; conditioning on x is suppressed below. The cumulative weak-order map gives $\Pr(Y_k = Y_{k-1} \mid x) = \Phi(-\mu_k/\sigma_k)$ for $k \geq 2$, while nonnegative cumulative ordering gives $\Pr(Y_k = 0 \mid x) = \prod_{j \leq k} \Phi(-\mu_j/\sigma_j)$. For the normalized positive-part simplex map, define the random active set $\mathcal{A} = \{i : Z_i > 0\}$. For any nonempty index set S,

$$\Pr(\mathcal{A} = S \mid x) = \prod_{i \in S} \Phi(\mu_i/\sigma_i) \prod_{j \notin S} \Phi(-\mu_j/\sigma_j). \tag{5}$$

Conditional on $\mathcal{A} = S \neq \varnothing$, the output lies in the relative interior of the face indexed by S. The all-inactive event is assigned to a vertex by the fallback in Appendix C.2, adding mass to vertex outcomes. For $Z \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = (\cos Z, \sin Z)$, $\mathbb{E}[\cos Z] = e^{-\sigma^2/2}\cos\mu$ and $\mathbb{E}[\sin Z] = e^{-\sigma^2/2}\sin\mu$. These formulas, their second moments, and covariance expressions are collected in Appendix E.3; here they expose representation-specific probability allocation. The same analysis records whether the stochastic coordinates are intrinsic (affine, circle), full-dimensional (ordering), or computationally redundant (normalized simplex and nonlinear FDC).

3.3 Structural-map design considerations

Structural-map choice introduces trade-offs beyond feasibility. We use three practical design considerations to characterize these choices; they are interpretive considerations rather than a universal optimization criterion. Coherence asks whether the map enforces the full stated structure jointly, sample by sample. Sensitivity describes how latent variation propagates through the map, including Jacobian structure, flat regions, saturation, and boundary atoms. Simplicity concerns explicitness, latent redundancy, and whether an additional correction or solve stage is necessary. For the positive-part map $h(z) = [z]_+$, $h'(z) = \mathbf{1}_{\{z>0\}}$ almost everywhere, so under $Z \sim \mathcal{N}(\mu, \sigma^2)$,

$$\mathbb{E}[h'(Z)] = \Pr(Z > 0) = \Phi(\mu/\sigma) = 1 - \Pr(h(Z) = 0)\,. \tag{6}$$

Thus the same latent parameters that control exact boundary mass also control the expected active-gradient probability for this map. This identity is a sensitivity diagnostic, not an optimization guarantee. These considerations explain why two equally feasible maps can still differ statistically or computationally.

3.4 Relationship to projection and ProbHardE2E

ProbHardE2E/DPPL constructs a feasible law by projecting an ambient predictive law, whereas an *SFLP* defines it directly through a structural latent law and feasible map. For affine $AY = b$, DPPL conditions an ambient diagonal Gaussian (Utkarsh et al., 2026); `SLP-ProbHard` models the $d - m$ free coordinates in Eq. (2). The DPPL conditioning and sampling equations are given in Appendix O.2.

The constructions can have identical feasible support yet admit different covariance families under the selected heads. Equality of support therefore does not imply equality of probability allocation. Projection remains useful for implicit or difficult feasible sets; structural generation is attractive when meaningful feasible coordinates exist or boundary behavior matters. Appendix F provides the detailed mechanism comparison.

4 Experiments

We evaluate frozen held-out synthetic and seven-basin FDC benchmarks against unconstrained, soft, post-hoc, and matched end-to-end hard baselines; Appendix G provides the

complete seed registries and protocols. The affine benchmark additionally includes ProbHardE2E/DPPL (Utkarsh et al., 2026) using the official released repository, frozen at commit `ff787718e79e5ac6b673807d40ce1cddb0fd17c2`, and a verified affine adapter; the existing `SLP-ProbHard` seed-20-29 results were reused unchanged rather than retrained.

Synthetic comparisons use matched backbones and training protocols, optimizing an equally weighted combination of marginal CRPS and Energy Score. The official-source DPPL (Utkarsh et al., 2026; Amazon Science, 2026) affine adapter preserves the released Gaussian conditioning equations and passes an equation-level equivalence audit. Appendix G gives the training and evaluation settings; Appendix O documents the adapter and verification.

The real-world source is a single frozen seasonal FDC dataset built from daily river discharge from the *GloFAS v5.0 hydrological reanalysis*, a LISFLOOD-based JRC dataset (Grimaldi et al., 2026; Van Der Knijff et al., 2010; Burek et al., 2013), for seven Emilia-Romagna basins. Basin-aggregated precipitation, minimum-temperature, and maximum-temperature descriptors are derived from ERACLITO61 (Antolini et al., 2016), with season, basin area, and basin identity completing the 22 predictors. The common 1996-2022 period yields 756 samples with a 532/112/112 temporal train/validation/test split. Following standard FDC/exceedance-probability representations and multi-output FDC learning (Burgan & Aksoy, 2018; Worland et al., 2019; Vogel & Fennessey, 1994; 1995), the ordered FDC target is

$$Q = [Q_{95}, Q_{90}, Q_{75}, Q_{50}, Q_{25}, Q_{10}, Q_{05}], \qquad 0 \le Q_{95} \le \cdots \le Q_{05}. \tag{7}$$

The three targets provide scientifically interpretable representations of the same underlying FDC. The seven levels describe discharge across exceedance regimes. Adjacent increments quantify changes between neighboring regimes, while the scale–shape representation separates the discharge range $R = Q_{05} - Q_{95}$ from its relative allocation across the curve. For $R > 0$, normalized increments $P_i = \Delta_i / R$ describe this allocation; Appendix G.7 specifies the flat-curve convention. The affine representation exposes increment coherence, whereas the nonlinear representation separates the low-flow baseline, discharge range, and relative increment allocation. Reusing the underlying FDCs and split allows the experiments to examine how these representation choices affect predictive laws.

The original nonnegative-ordering benchmark models $\log(1+Q)$. The affine and nonlinear extensions reuse the *same frozen predictors, raw FDCs, basin/season/year identities, and split* without rereading daily discharge or recomputing FDCs; they apply one train-only multiplicative RMS scale $s_Q = 43.7191$ to raw Q, preserving difference and scale-shape identities exactly. Cross-representation comparisons therefore concern the combined representation and preprocessing choices; they do not isolate a change of feasible map under identical target preprocessing. Appendix G gives the complete constructions, seed registries, baselines, metrics, and reproducibility details.

5 Results

5.1 Primary synthetic benchmarks

Table 2 reports matched end-to-end hard references, the direct official-source ProbHardE2E/DPPL affine-adapter comparison, and SLP. The affine comparison first highlights the stochastic representation: DPPL conditions an ambient Gaussian with 11 stochastic coordinates, whereas `SLP-ProbHard` learns an 8-dimensional Gaussian in affine free coordinates, a 27.27% reduction. Both have zero observed violations at the 10^{-5} threshold, but their primary diagonal heads induce different feasible covariance families. `SLP-ProbHard` lowers MSE by 1.30% ($p = 0.0072$) and MAE by 0.43% ($p = 0.0353$), with 2.48% narrower intervals ($p = 0.0341$). DPPL lowers marginal CRPS by 1.18% ($p = 3.76 \times 10^{-4}$) and gives higher coverage (0.809 versus 0.786), closer to the nominal 0.90 level. The paired test does not detect an Energy Score difference ($p = 0.391$). These results are consistent with representation-dependent probability allocation within the same feasible affine support. Projection-Or remains a transparent closed-form reference; richer `SLP-ProbHard` covariance later closes or reverses its diagonal-head score gap. In ordering and simplex, positive-part `SLP-ProbHard` significantly improves the reported CRPS/ES relative to the matched end-to-end projection.

On the unit circle, the angle map $h(z) = (\cos z, \sin z)$ achieves zero observed 10^{-5} violations using one stochastic coordinate versus two for radial projection. Paired tests detect no MSE, CRPS, or

Family	Method	MSE ↓	CRPS ↓	ES ↓	Cov.90	Stoch. dim.	Feasibility
Linear	Projection-Or	0.2965	0.2847	1.2088	0.785	11	$\mathrm{VR}_{10^{-5}} = 0$
Linear	DPPL (official-source adapter)	0.2997	0.2868	1.2190	0.809	11	$\mathrm{VR}_{10^{-5}} = 0$
Linear	`SLP-ProbHard`	0.2957	0.2901	1.2216	0.786	8	$\mathrm{VR}_{10^{-5}} = 0$
Order	Projection-isotonic	5.5677	0.7945	2.2440	0.784	5	exact order
Order	SLP-positive-part	5.0948	0.7820	2.1883	0.797	5	exact order
Simplex	Projection-simplex	0.0322	0.0890	0.2474	0.851	5	CE 1.51×10^{-8}
Simplex	SLP-positive-part	0.0320	0.0866	0.2431	0.846	5	CE 2.30×10^{-8}
Circle	Radial projection	0.3557	0.3277	0.5147	0.864	2	$\mathrm{VR}_{10^{-5}} = 0$
Circle	SLP-angle	0.3557	0.3272	0.5143	0.854	1	$\mathrm{VR}_{10^{-5}} = 0$

Table 2: Primary synthetic benchmarks (means across untouched seeds). "DPPL (official-source adapter)" denotes ProbHardE2E/DPPL using the official released source with the verified affine adapter; full source audit and paired statistics are in Appendix O.

Energy Score differences across ten seeds. Appendix H, under "Nonlinear circle primary results," gives the full comparison.

Boundary behavior illustrates Eq. (3). For ordering, SLP's exact-tie rate 0.208 is closer than end-to-end isotonic projection's 0.396 to the observed 0.255. For the $d = 5$ simplex, projection's zero rate 0.246 is closer to the observed 0.250 than SLP's 0.180. Thus boundary capability is not boundary calibration: the map determines whether an event can carry mass, while the latent law and training determine how much mass it receives.

5.2 Composed and real-world structural representations

A controlled synthetic nonnegative-ordering task generated targets by cumulative positive-part. Because the data-generating mechanism uses the same structure, this is mechanism validation for the boundary-preimage prediction rather than independent evidence that cumulative-positive-part `SLP-ProbHard` generally dominates alternative enforcement. Both nonnegative isotonic projection and `SLP-ProbHard` satisfy non-negativity and ordering exactly; predictive scores are close, with `SLP-ProbHard` numerically lower on CRPS and Energy Score.

Holding the underlying FDC data and split fixed, Table 3 tests how three scientifically meaningful representations affect feasibility, stochastic dimensionality, calibration, and predictive scores. First, the nonnegative-ordering map directly enforces Eq. (7); both hard methods eliminate negative discharge and crossings, with `SLP-ProbHard` descriptively lower CRPS/ES but projection better calibrated in 90% coverage. Second, the affine representation augments the seven FDC levels with six adjacent increments,

$$Y_{\text{aff}} = (Q_0, \ldots, Q_6, \Delta_0, \ldots, \Delta_5) \in \mathbb{R}^{13}, \qquad \Delta_i = Q_{i+1} - Q_i, \tag{8}$$

equivalently $Q_i - Q_{i+1} + \Delta_i = 0$. The six equalities have rank 6, so the feasible affine representation has 7 intrinsic coordinates. `SLP-ProbHard` uses $h(z) = (z, \operatorname{diff} z)$ with 7 stochastic coordinates versus 13 for ambient end-to-end affine projection, a 46.15% reduction. Both have zero affine violations at 10^{-5}; this branch enforces increment coherence only and does not additionally enforce nonnegativity or FDC ordering. The paired ten-seed tests did not detect differences in MSE, MAE, CRPS, or Energy Score; `SLP-ProbHard` intervals are significantly narrower (10.94 vs. 13.57) but coverage is lower (0.712 vs. 0.746), so the result is a calibration-sharpness trade-off rather than a score win.

Third, for $R > 0$, the nonlinear scale-shape representation defines

$$B = Q_0, \qquad \Delta_i = Q_{i+1} - Q_i, \qquad R = \sum_i \Delta_i, \qquad P_i = \Delta_i / R, \tag{9}$$

and models

$$Y_{\text{nl}} = (B, R, P, \Delta) \in \mathbb{R}^{14}, \qquad \Delta_i = RP_i, \quad P_i \geq 0, \quad \sum_i P_i = 1. \tag{10}$$

`SLP-ProbHard` uses $B = [z_B]_+$, $R = [z_R]_+$, normalized positive-part P, and $\Delta = RP$. For $d = 7$, the mathematical intrinsic dimension is 7 but this boundary-capable implementation uses

Real-world representation	Hard method	Raw-Q CRPS	Raw-Q ES	Cov.90	Width.90	Stoch. dim.	Hard VR
Nonneg. order	Projection-nonneg-isotonic	3.894	15.561	0.757	14.15	7	0
	SLP-nonneg-positive-part	3.710	14.603	0.722	22.26	7	0
Affine coherence	End-to-end affine projection	3.627	13.649	0.746	13.57	13	0
	SLP-affine	3.698	13.997	0.712	10.94	7	0
Nonlinear scale-shape	Post-hoc feasible correction	3.568	13.754	0.823	14.97	14	0
	End-to-end feasible correction	4.177	14.483	0.630	15.75	14	0
	SLP-scale-shape	4.047	14.313	0.649	15.67	8	0

Table 3: Three structural views of the same frozen seven-basin seasonal FDC dataset. "Hard VR" is the representation-specific 10^{-5} violation rate; the nonlinear hard methods also have zero negative reconstructed discharge and zero ordering violations. For the affine rows, `Hard VR` measures increment-coherence equalities only.

8 computational stochastic coordinates; the ambient correction model uses 14, giving a 42.86% reduction. Both coordinate reductions are relative to the corresponding augmented output representation. The original FDC has seven flow levels, and the nonlinear implementation retains one redundant computational coordinate. Both end-to-end hard methods have zero nonlinear violations, zero negative reconstructed discharge, and zero ordering violations. None of the paired Raw-Q MSE/MAE/CRPS/ES differences was statistically detected at the ten-seed level; `SLP-ProbHard` is numerically lower on CRPS (4.047 vs. 4.177) and Energy Score (14.313 vs. 14.483). The comparator is a five-iteration differentiable alternating feasible correction, *not* an exact joint Euclidean projection. Post-hoc feasible correction gives the lowest reported mean raw-Q CRPS among the nonlinear baselines (3.568 versus 4.047 for SLP) and higher 90% coverage (0.823 versus 0.649). Thus the nonlinear result supports feasible structural generation with fewer computational coordinates, while the strongest reported predictive scores favor post-hoc correction (Appendix H.7).

5.3 Covariance and computation

Richer `SLP-ProbHard` covariance heads separate probabilistic expressiveness from structural feasibility. On the linear benchmark, moving from diagonal to rank-4 or full covariance improves `SLP-ProbHard` CRPS from 0.2889 to about 0.2837 and Energy Score from 1.2171 to 1.2071 while keeping zero observed violations. This helps interpret the official-source DPPL affine-adapter result: covariance-weighted conditioning of an ambient diagonal Gaussian and a diagonal structural latent Gaussian induce different feasible covariance families, so identical support need not imply identical proper scores. This ablation varies covariance expressiveness within `SLP-ProbHard` while preserving structural feasibility; it is not a fully covariance-matched DPPL comparison.

Under matched CPU instrumentation, median inference latency is 1.65 ms for `SLP-ProbHard` versus 3.22 ms for the DPPL affine adapter. These implementation-specific timings and parameter counts are detailed in Appendix I and Appendix O.

5.4 Scaling, finite precision, and statistical interpretation

The scaling experiments test whether the primary observations survive beyond a single dimension. For affine equalities, `SLP-ProbHard` always uses $q = d - m$ stochastic coordinates across twelve pre-specified configurations, reducing stochastic dimension by 12.3-38.5%. Point accuracy remains close to Projection-Or; the latter usually retains a small CRPS/ES advantage under the diagonal `SLP-ProbHard` head. Re-evaluating the primary affine map in float64 collapses residuals from roughly 10^{-7} to 10^{-15}, confirming that base-case residuals are numerical realization effects rather than relaxed mathematical constraints.

Zero thresholded violations in the primary benchmark do not extend to every float32 scaling configuration: at $(d, m, q) = (97, 33, 64)$, $\mathrm{VR}_{10^{-5}}$ is 0.0193 for SLP and 0.3958 for Projection-Or (Appendix J.1). Ordering and simplex scaling show a different pattern. Positive-part `SLP-ProbHard` has lower CRPS and Energy Score than the matched end-to-end projection in every evaluated dimension $d \in \{3, 5, 10, 20, 40\}$.

Complete paired confidence intervals, Wilcoxon tests, source audits, scaling tables, and numerical diagnostics are given in the appendix.

6 Discussion and Limitations

Exact feasibility does not uniquely specify a predictive law. The experiments show how structural representation shapes boundary probability, stochastic coordinates, attainable covariance, and calibration-sharpness trade-offs. Coordinate savings depend on the selected map: affine and circle constructions exploit intrinsic coordinates, whereas normalized simplex and nonlinear scale-shape retain computational redundancy.

From constraint sets to structural representations. The three FDC views test physical ordering, increment coherence, and scale–shape coherence on one fixed dataset and split. Together with their distinct preprocessing, they show how scientifically meaningful representation choices affect probabilistic predictions. Their hydrological interpretation is developed in Appendices G and K.1.

Calibration. Hard feasibility does not ensure calibrated uncertainty. The reported 90% marginal coverage for SLP is 0.722, 0.712, and 0.649 across the three FDC representations. These values show substantial undercoverage; interval width must therefore be interpreted jointly with coverage, and narrower intervals alone do not establish better uncertainty quantification.

Explicit nonlinear-equality scope. The unit circle and the FDC scale-shape map demonstrate two explicit nonlinear equalities, but they do not imply a general `SLP-ProbHard` construction for arbitrary nonlinear constraints. The circle map is periodic and non-injective; the scale-shape map uses a boundary-capable simplex representation with one redundant computational coordinate. Projection/correction remains more general for implicit or difficult feasible sets.

Computation. Stochastic-coordinate savings are architectural and hardware independent. Wall-clock improvements depend on the workload and implementation; optimized affine conditioning is itself inexpensive. Appendix K.5 distinguishes these claims.

Limitations. `SLP-ProbHard` requires a useful explicit map, and non-surjective maps introduce representation bias. The primary positive-part boundary-mass implementation is nonsmooth and has a flat negative region. For maps with discontinuous fallbacks, samplewise differentiation alone does not establish unbiased gradients of expected scores. Nonlinear pushforwards may lack tractable densities or moments. The hydrological target is GloFAS v5.0 LISFLOOD-based reanalysis discharge rather than observed-gauge discharge (Grimaldi et al., 2026). The nonlinear FDC comparator is an alternating feasible correction rather than an exact joint Euclidean projection. The direct ProbHardE2E experiment is an official-source *affine-adapter* comparison, not a reproduction of the authors' PDE datasets or published wall-clock experiments; it establishes one compatible direct comparison, not superiority across ProbHardE2E's broader supported classes. Additional limitations and numerical diagnostics are in the appendix.

7 Conclusion

`SLP-ProbHard` introduces a cross-family representation-centered framework in which a structural latent law and a feasible map jointly define a hard-feasible predictive law. The latent law and map jointly determine predictive support and boundary probability; the chosen coordinates and map specify the stochastic representation and shape attainable dependence, calibration, expressiveness, and computation. Experiments span affine equalities, ordering, simplex, explicit nonlinear maps, and three representations of the same hydrological FDC data. The direct affine DPPL comparison demonstrates shared practical feasibility and representation-dependent predictive trade-offs. Structural generation is useful when suitable explicit coordinates exist; projection and correction remain useful for implicit or difficult feasible sets.

REFERENCES

Akshay Agrawal, Brandon Amos, Shane Barratt, Stephen Boyd, Steven Diamond, and J. Zico Kolter. Differentiable convex optimization layers. In *Advances in Neural Information Processing Systems*, 2019.

J. Aitchison. The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):139–177, 1982. doi: 10.1111/j.2517-6161.1982.tb01195.x.

J. Aitchison and S. M. Shen. Logistic-normal distributions: Some properties and uses. *Biometrika*, 67(2):261–272, 1980. doi: 10.1093/biomet/67.2.261.

Amazon Science. ProbHardE2E official source repository. GitHub repository, 2026. Frozen in this study at commit ff787718e79e5ac6b673807d40ce1cddb0fd17c2; https://github.com/amazon-science/probharde2e.

Brandon Amos and J. Zico Kolter. Optnet: Differentiable optimization as a layer in neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pp. 136–145. PMLR, 2017.

G. Antolini, L. Auteri, V. Pavan, F. Tomei, R. Tomozeiu, and V. Marletto. A daily high-resolution gridded climatic data set for emilia-romagna, italy, during 1961–2010. *International Journal of Climatology*, 36(4):1970–1986, 2016. doi: 10.1002/joc.4473.

Wondesen Teshome Bekele and Marco D'Oria. Smt-hcsd: Stochastic multi-target hard-constrained spatial deep learning framework and architecture for perfect prognosis cmip6 daily temperature downscaling. *Stochastic Environmental Research and Risk Assessment*, 40, 2026. doi: 10.1007/s00477-026-03303-1.

Souhaib Ben Taieb, James W. Taylor, and Rob J. Hyndman. Coherent probabilistic forecasts for hierarchical time series. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pp. 3348–3357. PMLR, 2017.

Stefano Berrone, Claudio Canuto, Moreno Pintore, and N. Sukumar. Enforcing dirichlet boundary conditions in physics-informed neural networks and variational physics-informed neural networks. *Heliyon*, 9(8):e18820, 2023. doi: 10.1016/j.heliyon.2023.e18820.

M. J. Best and N. Chakravarti. Active set algorithms for isotonic regression: A unifying framework. *Mathematical Programming*, 47:425–439, 1990. doi: 10.1007/BF01580873.

Mathieu Blondel, Olivier Teboul, Quentin Berthet, and Josip Djolonga. Fast differentiable sorting and ranking. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 950–959. PMLR, 2020.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Peter Andreas Burek, Johan van der Knijff, and Arie de Roo. LISFLOOD – distributed water balance and flood simulation model – revised user manual 2013. Technical Report EUR 26162, Publications Office of the European Union, Joint Research Centre, 2013.

Halil Ibrahim Burgan and Hafzullah Aksoy. Annual flow duration curve model for ungauged basins. *Hydrology Research*, 49(5):1684–1695, 2018. doi: 10.2166/nh.2018.109.

Hao Chen, Gonzalo E. Constante Flores, and Can Li. Physics-informed neural networks with hard linear equality constraints. *arXiv preprint arXiv:2402.07251*, 2024.

Ya-Chi Chu, Alkiviades Boukas, and Madeleine Udell. SnareNet: Flexible repair layers for neural networks with hard constraints. *arXiv preprint arXiv:2602.09317*, 2026.

Laurent Condat. Fast projection onto the simplex and the ℓ_1 ball. *Mathematical Programming*, 158 (1):575–585, 2016.

Marco Cuturi, Olivier Teboul, and Jean-Philippe Vert. Differentiable ranking and sorting using optimal transport. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2(4):303–314, 1989. doi: 10.1007/BF02551274.

Jan de Leeuw, Kurt Hornik, and Patrick Mair. Isotone optimization in R: Pool-adjacent-violators algorithm (PAVA) and active set methods. *Journal of Statistical Software*, 32(5):1–24, 2009. doi: 10.18637/jss.v032.i05.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using Real NVP. In *International Conference on Learning Representations*, 2017.

Priya L. Donti, David Rolnick, and J. Zico Kolter. Dc3: A learning method for optimization with hard constraints. In *International Conference on Learning Representations*, 2021.

John Duchi, Shai Shalev-Shwartz, Yoram Singer, and Tushar Chandra. Efficient projections onto the ℓ_1-ball for learning in high dimensions. In *Proceedings of the 25th International Conference on Machine Learning*, pp. 272–279, 2008.

Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. In *Advances in Neural Information Processing Systems*, volume 32, pp. 7509–7520, 2019.

Mikhail Figurnov, Shakir Mohamed, and Andriy Mnih. Implicit reparameterization gradients. In *Advances in Neural Information Processing Systems*, volume 31, pp. 441–452, 2018.

Thomas Frerix, Matthias Nießner, and Daniel Cremers. Homogeneous linear inequality constraints for neural network activations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 3229–3234, 2020.

Daniele Girolimetto and Tommaso Di Fonzo. Point and probabilistic forecast reconciliation for general linearly constrained multiple time series. *Statistical Methods & Applications*, 33:581–607, 2024. doi: 10.1007/s10260-023-00738-6.

Daniele Girolimetto, George Athanasopoulos, Tommaso Di Fonzo, and Rob J. Hyndman. Cross-temporal probabilistic forecast reconciliation: Methodological and practical issues. *International Journal of Forecasting*, 40(3):1134–1151, 2024. doi: 10.1016/j.ijforecast.2023.10.003.

Tilmann Gneiting and Matthias Katzfuss. Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1:125–151, 2014. doi: 10.1146/annurev-statistics-062713-085831.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

Tilmann Gneiting, Larissa I. Stanberry, Eric P. Grimit, Leonhard Held, and Nicholas A. Johnson. Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *TEST*, 17(2):211–235, 2008. doi: 10.1007/s11749-008-0114-x.

Andrea Goertzen, Kaveh Alim, and Navid Azizan. HardNet++: Nonlinear constraint enforcement in neural networks. *arXiv preprint arXiv:2604.19669*, 2026.

David Goldberg. What every computer scientist should know about floating-point arithmetic. *ACM Computing Surveys*, 23(1):5–48, 1991. doi: 10.1145/103162.103163.

Stefania Grimaldi, Peter Salamon, Carlo Russo, Juliana Disperati, Jesus Casado-Rodriguez, Mohamed Azhar, Cinzia Mazzetti, Nikolaos Mastrantonas, Francesca Moschini, Timo Schaffhauser, Arthur Ramos, Corentin Carton De Wiart, Ervin Zsoter, Goncalo Gomes, Rafael Garcia-Sanchez, Alejandro Serratosa-Marquez, Antonio Jimenez-Molina, Marina Gonzalez-Martin, and Christel Prudhomme. GloFAS v5.0 Hydrological Reanalysis. European Commission, Joint Research Centre [Dataset], 2026. PID: http://data.europa.eu/89h/83cd8a58-765a-4aa5-8e74-5dfc429c3706.

Panagiotis D. Grontas, Antonio Terpin, Efe C. Balta, Raffaello D'Andrea, and John Lygeros. Pinet: Optimizing hard-constrained neural networks with orthogonal projection layers. *arXiv preprint arXiv:2508.10480*, 2025.

Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002. Available from the author's Cornell University webpage.

Hans Hersbach. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570, 2000.

Nicholas J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM, Philadelphia, 2 edition, 2002. doi: 10.1137/1.9780898718027.

Michael Hintermüller and Jianfeng Ning. Constrained neural parameterization for optimization in function spaces. *arXiv preprint arXiv:2606.00855*, 2026.

Miranda J. S. Horne, Peter K. Jimack, Amirul Khan, and He Wang. Hard constraint projection in a physics informed neural network. *arXiv preprint arXiv:2601.06244*, 2026.

Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2): 251–257, 1991. doi: 10.1016/0893-6080(91)90009-T.

Rob J. Hyndman, Roman A. Ahmed, George Athanasopoulos, and Han Lin Shang. Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, 55(9): 2579–2589, 2011. doi: 10.1016/j.csda.2011.03.006.

Rob J. Hyndman, Alan J. Lee, and Earo Wang. Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics & Data Analysis*, 97:16–32, 2016.

Ashfaq Iftakher, Rahul Golder, Bimol Nath Roy, and M. M. Faruque Hasan. Physics-informed neural networks with hard nonlinear equality and inequality constraints. *arXiv preprint arXiv:2507.08124*, 2025.

Christian Igel. Smooth min-max monotonic networks. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 20908–20923. PMLR, 2024.

Martin Jankowiak and Fritz Obermeyer. Pathwise derivatives beyond the reparameterization trick. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pp. 2235–2244. PMLR, 2018.

Jooyoung Jeon, Anastasios Panagiotelis, and Fotios Petropoulos. Probabilistic forecast reconciliation with applications to wind power and electric load. *European Journal of Operational Research*, 279 (2):364–379, 2019. doi: 10.1016/j.ejor.2019.05.020.

Alexander Jordan, Fabian Krueger, and Sebastian Lerch. Evaluating probabilistic forecasts with scoringrules. *Journal of Statistical Software*, 90(12):1–37, 2019. doi: 10.18637/jss.v090.i12.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014.

Isaac E. Lagaris, Aristidis Likas, and Dimitrios I. Fotiadis. Artificial neural networks for solving ordinary and partial differential equations. *IEEE Transactions on Neural Networks*, 9(5):987–1000, 1998.

Moshe Leshno, Vladimir Ya. Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867, 1993.

Ruoyan Li, Dipti Ranjan Sahu, Guy Van den Broeck, and Zhe Zeng. Learning deep generative models under hard linear equality constraints. In *Proceedings of the 20th International Conference on Neurosymbolic Learning and Reasoning (NeSy)*, volume 284 of *Proceedings of Machine Learning Research*, pp. 1–37, 2026.

Enbin Liang, Minghua Chen, and Steven H. Low. Low complexity homeomorphic projection to ensure neural-network solution feasibility for constrained optimization. In *Proceedings of the 40th International Conference on Machine Learning*. PMLR, 2023.

Enbin Liang, Minghua Chen, and Steven H. Low. Homeomorphic projection to ensure neural-network solution feasibility for constrained optimization. *Journal of Machine Learning Research*, 25(329): 1–55, 2024.

Xingchao Liu, Xing Han, Na Zhang, and Qiang Liu. Certified monotonic neural networks. In *Advances in Neural Information Processing Systems*, volume 33, pp. 15427–15438, 2020.

Andre F. T. Martins and Ramon Fernandez Astudillo. From softmax to sparsemax: A sparse model of attention and multi-label classification. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48, pp. 1614–1623. PMLR, 2016.

Youssef Marzouk, Tarek Moselhy, Matthew Parno, and Alessio Spantini. An introduction to sampling via measure transport. *arXiv preprint arXiv:1602.05023*, 2016.

James E. Matheson and Robert L. Winkler. Scoring rules for continuous probability distributions. *Management Science*, 22(10):1087–1096, 1976. doi: 10.1287/mnsc.22.10.1087.

Yifei Min, Akash Sonar, and Navid Azizan. Hard-constrained neural networks with universal approximation guarantees. *arXiv preprint arXiv:2410.10807*, 2024.

Huy T. Nguyen and Priya L. Donti. Fsnet: Feasibility-seeking neural network for constrained optimization with guarantees. *arXiv preprint arXiv:2506.00362*, 2025.

Gary W. Oehlert. A note on the delta method. *The American Statistician*, 46(1):27–29, 1992. doi: 10.1080/00031305.1992.10475842.

Kin G. Olivares, Geoffrey Négiar, Ruijun Ma, Oinam Nganba Meetei, Mengfei Cao, and Michael W. Mahoney. Clover: Probabilistic forecasting with coherent learning objective reparameterization. *Transactions on Machine Learning Research*, 2024. URL `https://openreview.net/forum?id=q7YXEbFOAt`.

Anastasios Panagiotelis, Puwasala Gamakumara, George Athanasopoulos, and Rob J. Hyndman. Probabilistic forecast reconciliation: Properties, evaluation and score optimisation. *European Journal of Operational Research*, 306(2):693–706, 2023. doi: 10.1016/j.ejor.2022.07.040.

George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021.

Syama Sundar Rangapuram, Lucien D. Werner, Konstantinos Benidis, Pedro Mercado, Jan Gasthaus, and Tim Januschowski. End-to-end learning of coherent probabilistic forecasts for hierarchical time series. In *Proceedings of the 38th International Conference on Machine Learning*. PMLR, 2021.

Syama Sundar Rangapuram, Shubham Kapoor, Rajbir Singh Nirwan, Pedro Mercado, Tim Januschowski, Yuyang Wang, and Michael Bohlke-Schneider. Coherent probabilistic forecasting of temporal hierarchies. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206, pp. 9362–9376. PMLR, 2023.

Davor Runje and Sharath M. Shankaranarayana. Constrained monotonic neural networks. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 29338–29353. PMLR, 2023.

Michael Scheuerer and Thomas M. Hamill. Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, 143(4):1321–1334, 2015. doi: 10.1175/MWR-D-14-00269.1.

Paul J. Schneider and Daniel Kuhn. Soft-radial projection for constrained end-to-end learning. *arXiv preprint arXiv:2602.03461*, 2026.

Jayaram Sethuraman. A constructive definition of dirichlet priors. *Statistica Sinica*, 4(2):639–650, 1994.

N. Sukumar and Ankit Srivastava. Exact imposition of boundary conditions with distance functions in physics-informed deep neural networks. *Computer Methods in Applied Mechanics and Engineering*, 389:114333, 2022. doi: 10.1016/j.cma.2021.114333.

Gábor J. Székely and Maria L. Rizzo. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272, 2013. doi: 10.1016/j.jspi.2013.03.018.

G. M. Tallis. The moment generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, 23(1):223–229, 1961. doi: 10.1111/j.2517-6161.1961.tb00408.x.

Sunbochen Tang, Andrea Goertzen, and Navid Azizan. LMI-Net: Linear matrix inequality–constrained neural networks via differentiable projection layers. *arXiv preprint arXiv:2604.05374*, 2026.

Jesus Tordesillas, Jonathan P. How, and Marco Hutter. Rayen: Imposition of hard convex constraints on neural networks. *arXiv preprint arXiv:2307.08336*, 2023.

Utkarsh, Danielle C. Maddix, Ruilin Ma, Michael W. Mahoney, and Yuyang Wang. End-to-end probabilistic framework for learning with hard constraints. In *International Conference on Learning Representations*, 2026. arXiv:2506.07003.

Johan M. Van Der Knijff, Jutta Younis, and Ad De Roo. Lisflood: A gis-based distributed model for river basin scale water balance and flood simulation. *International Journal of Geographical Information Science*, 24(2):189–212, 2010. doi: 10.1080/13658810802549154.

Cédric Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009. doi: 10.1007/978-3-540-71050-9.

Richard M. Vogel and Neil M. Fennessey. Flow-duration curves. i: New interpretation and confidence intervals. *Journal of Water Resources Planning and Management*, 120(4):485–504, 1994.

Richard M. Vogel and Neil M. Fennessey. Flow duration curves ii: A review of applications in water resources planning. *Journal of the American Water Resources Association*, 31(6):1029–1039, 1995. doi: 10.1111/j.1752-1688.1995.tb03419.x.

Ziqian Wang, Chenxi Fang, and Zhen Zhang. DiffSlack: Learning under nonlinear inequality constraints via learnable slack variables. *arXiv preprint arXiv:2606.05247*, 2026.

Antoine Wehenkel and Gilles Louppe. Unconstrained monotonic neural networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Shanika L. Wickramasuriya, George Athanasopoulos, and Rob J. Hyndman. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, 114(526):804–819, 2019.

Scott C. Worland, Scott Steinschneider, William H. Asquith, Rodney Knight, and Michael E. Wieczorek. Prediction and inference of flow-duration curves using multioutput neural networks. *Water Resources Research*, 55(8):6850–6868, 2019. doi: 10.1029/2018WR024463.

Seungil You, David Ding, Kevin Canini, Jan Pfeifer, and Maya Gupta. Deep lattice networks and partial monotonic functions. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

A Extended Related Work

Work	Core mechanism	Prob. law?	Hard feasibility	Representation emphasis
OptNet / diff. convex (Amos & Kolter, 2017; Agrawal et al., 2019)	optimization layer	opt.-centered	supported problem	optimization geometry
DC3 (Donti et al., 2021)	completion + correction	det./opt.	correction-dependent	completion/correction
Frerix et al. (2020)	direct linear-inequality map	deterministic	yes, supported class	feasible geometry
RAYEN (Tordesillas et al., 2023)	direct convex feasible layer	deterministic	yes, convex class	convex parameterization
CLOVER (Olivares et al., 2024)	differentiable coherent structural generation	yes	yes, hierarchy	coherent probabilistic representation
Girolimetto-Di Fonzo (Girolimetto & Di Fonzo, 2024)	free/constrained structural-like representation	yes	yes, linear coherence	general linear structural coordinates
SMT-HCSD (Bekele & D'Oria, 2026)	stochastic nonnegative increment	yes	yes, ordering	domain-specific structural generation
ProbHardE2E (Utkarsh et al., 2026)	probabilistic projection	yes	yes, supported classes	projected-law formulation
Li et al. (2026)	hard linear constrained generative law	yes	yes, linear equality	probabilistic constrained linear law
Soft-Radial (Schneider & Kuhn, 2026)	radial feasible map	map-centered	interior feasible	interior support geometry
CNP (Hintermüller & Ning, 2026)	admissible neural coordinates	opt.-centered	yes, supported sets	function-space coordinates
`SLP-ProbHard`	structural latent pushforward	yes	yes, map image	support/boundary, stochastic coordinates, expressiveness/calibration, computation

Table 4: **Literature comparison by mechanism and probabilistic emphasis.** Structural probabilistic generation has important precedents in hierarchical, linear, and domain-specific settings. `SLP-ProbHard` uses structural representation as a cross-family probabilistic model-design object and jointly analyzes its consequences for support/boundary probability, stochastic coordinates, expressiveness/calibration, and computation. The direct ProbHardE2E comparison in this paper is restricted to the compatible synthetic affine benchmark.

A.1 Hard constraints in machine learning

Hard constraints define a feasible set $\mathcal{C}$ through equalities, inequalities, order relations, conservation laws, or boundary conditions. Soft penalties can be useful but leave a penalty-weight trade-off and cannot in general guarantee zero violation. Differentiable optimization layers such as OptNet and differentiable convex optimization layers bring exact or approximate optimization into end-to-end learning (Amos & Kolter, 2017; Agrawal et al., 2019). Classical convex projection geometry is summarized by Boyd & Vandenberghe (2004). For constraint families used in this paper, efficient Euclidean projection algorithms are well established for simplex/ℓ_1 geometry (Duchi et al., 2008; Condat, 2016), while isotonic methods provide classical projection machinery for order cones (Best & Chakravarti, 1990; de Leeuw et al., 2009). Differentiable relaxations and projections for sorting/ranking provide related trainable machinery (Cuturi et al., 2019; Blondel et al., 2020). More recent hard-constraint layers span operator-splitting projection, LMI projection, nonlinear repair, learnable slack-variable projection, and hard projection inside physics-informed networks (Grontas et al., 2025; Tang et al., 2026; Goertzen et al., 2026; Chu et al., 2026; Wang et al., 2026; Horne et al., 2026).

A.2 Probabilistic forecasting under constraints

In probabilistic prediction, hard feasibility should hold at the sample level, not only for a conditional mean. Proper scoring rules such as CRPS and Energy Score evaluate distributional quality through calibration and sharpness (Matheson & Winkler, 1976; Hersbach, 2000; Gneiting & Raftery, 2007; Gneiting et al., 2008; Gneiting & Katzfuss, 2014; Jordan et al., 2019); energy-distance ideas underlie multivariate distance-based evaluation (Székely & Rizzo, 2013), while variogram scores provide a dependence-sensitive alternative (Scheuerer & Hamill, 2015). Coherent forecasting developed from reconciliation of hierarchical point forecasts (Hyndman et al., 2011; 2016) to explicitly probabilistic reconciliation and end-to-end coherent prediction (Ben Taieb et al., 2017; Rangapuram et al., 2021; 2023; Jeon et al., 2019; Wickramasuriya et al., 2019; Panagiotelis et al., 2023; Girolimetto et al., 2024). ProbHardE2E (Utkarsh et al., 2026) provides a recent explicit probabilistic hard-constraint framework based on differentiable projection. `SLP-ProbHard` pursues the same high-level objective but makes the feasible representation generative rather than corrective.

A.3 FEASIBLE-BY-CONSTRUCTION PARAMETERIZATION

Many constraints admit direct parameterizations: nonnegative structural maps for non-negativity, cumulative nonnegative increments for ordering, null-space coordinates for linear equalities, normalized nonnegative weights for the simplex, and ansatz functions for initial or boundary conditions. Related monotonic-network constructions use lattice, integral, certified, constrained-weight, and smooth min-max mechanisms to encode monotonicity (You et al., 2017; Wehenkel & Louppe, 2019; Liu et al., 2020; Runje & Shankaranarayana, 2023; Igel, 2024). For compositional outputs, the simplex has a long statistical literature built around logistic-normal and compositional representations (Aitchison & Shen, 1980; Aitchison, 1982); sparse probability mappings such as sparsemax illustrate that exact zeros can be an intentional representation choice (Martins & Astudillo, 2016). Exact-condition ansatzes likewise have a long constructive lineage in neural differential-equation solvers (Lagaris et al., 1998), with later geometry-aware distance-function formulations and comparative PINN studies imposing Dirichlet conditions exactly (Sukumar & Srivastava, 2022; Berrone et al., 2023). These precedents motivate structural maps as established primitives; the present paper studies their role inside a probabilistic predictive-law methodology.

A.4 CONSTRUCTIVE FEASIBILITY LINEAGE

`SLP-ProbHard` sits within a broader constructive-feasibility lineage. DC3 combines equality completion with inequality correction (Donti et al., 2021). Frerix et al. (2020) parameterize activations satisfying homogeneous linear inequalities. Homeomorphic Projection maps from a simple domain into a constrained feasible set (Liang et al., 2023; 2024). RAYEN enforces hard convex constraints without expensive orthogonal projection or inner gradient descent (Tordesillas et al., 2023). HardNet studies hard-constrained networks with universal approximation guarantees (Min et al., 2024), while FSNet integrates feasibility-seeking into constrained optimization learning (Nguyen & Donti, 2025). Soft-Radial Projection maps ambient points radially into the interior of a convex feasible set and is particularly relevant as a recent non-orthogonal feasibility transformation (Schneider & Kuhn, 2026). The rapidly developing projection/repair branch also includes ΠNet for differentiable orthogonal projection, LMI-Net for semidefinite constraints, HardNet++ and SnareNet for nonlinear feasibility, DiffSlack for nonlinear inequalities, and KKT-based hard-constraint PINNs (Grontas et al., 2025; Tang et al., 2026; Goertzen et al., 2026; Chu et al., 2026; Wang et al., 2026; Chen et al., 2024; Iftakher et al., 2025).

Most of these methods target deterministic feasible outputs or optimization solutions. `SLP-ProbHard` instead centers a full predictive distribution: stochastic latent samples are mapped into feasible outputs, and the resulting pushforward law is trained by likelihoods or proper scoring rules.

Adjacent probabilistic and feasible-parameterization work. Li et al. (2026) study probabilistically sound deep generative modeling under hard linear equalities, directly establishing that hard-constrained generation and probability modeling can be combined in the linear setting. Soft-Radial Projection (Schneider & Kuhn, 2026) constructs a differentiable radial reparameterization into the strict interior of convex feasible sets, while Constrained Neural Parameterization (Hintermüller & Ning, 2026) maps unconstrained neural coordinates into admissible function-space sets for optimization. These papers make the novelty boundary important: `SLP-ProbHard` does not claim that feasible-by-construction transforms are new. Its methodological contribution is the coupled probabilistic treatment of structural latent generation, map-induced support and boundary atoms, free versus redundant stochastic coordinates, conditional approximation/parameterization bias, and cross-family proper-score/computational consequences under a single hard-constrained predictive-law formalism.

Additional context. The constructive-feasibility landscape also includes differentiable sorting/ranking, isotonic regression, monotone networks, exact-condition ansatzes, semidefinite-output layers, nonlinear repair, KKT-based formulations, and structured probabilistic reconciliation. These threads motivate the broader taxonomy but differ in whether they target deterministic feasibility, constrained optimization, or a complete predictive distribution. Together, these adjacent lines motivate the paper's representation-centered view of probabilistic hard-constrained learning.

B Complete Method and Structural-Map Taxonomy

B.1 Structural latent distribution

A neural backbone f_θ predicts latent distribution parameters $\eta_\theta(x)$. Under a reparameterized Gaussian,

$$Z = \mu_\theta(x) + L_\theta(x)\epsilon, \qquad \epsilon \sim \mathcal{N}(0, I). \tag{11}$$

The framework is not restricted to Gaussian latents; Student-t, Gamma, LogNormal, zero-inflated, mixture, autoregressive, or flow-based families may be used when scientifically appropriate. Flexible invertible transports such as Real NVP and neural spline flows provide examples of richer latent-to-output distribution transforms (Dinh et al., 2017; Durkan et al., 2019; Papamakarios et al., 2021), while explicit and implicit pathwise-gradient methods broaden the trainable latent-family choices beyond simple location-scale Gaussians (Figurnov et al., 2018; Jankowiak & Obermeyer, 2018).

B.2 Structural coordinates and computational latent dimension

For $Ay = b$, choose a feasible point y_0 and a null-space basis $N \in \mathbb{R}^{d\times q}$ satisfying $Ay_0 = b$ and $AN = 0$. Then

$$Y = y_0 + NZ, \qquad Z \in \mathbb{R}^q, \tag{12}$$

uses exactly the free affine coordinates. An ambient projection model instead predicts d quantities before removing infeasible directions.

This dimensional advantage is not universal. A convenient nonlinear map can use a redundant latent representation. In the normalized positive-part boundary-mass implementation below, the simplex has intrinsic dimension $d-1$ but the computational map uses d latent coordinates. We therefore distinguish *intrinsic degrees of freedom* from *computational latent dimension*. Dimensional economy is a property of the chosen map, not of the `SLP-ProbHard` label by itself.

B.3 Feasibility map and pushforward law

The feasibility map h_ϕ encodes the known structure. For a measurable set $B \subseteq \mathcal{Y}$,

$$P_\theta^Y(B \mid x) = P_\theta^Z\big(h_{\phi,x}^{-1}(B) \mid x\big). \tag{13}$$

Thus geometric properties of preimages under h_ϕ directly determine where the constrained predictive distribution can place probability.

B.4 Training objectives

When the induced density is tractable, one may minimize negative log-likelihood. Otherwise, sample-based proper scores provide a generic objective,

$$\mathcal{L}_S(\theta, \phi) = \sum_{i=1}^{N} S\big(P_\theta^Y(\cdot \mid x_i), y_i\big), \tag{14}$$

where S may be CRPS, Energy Score, Variogram Score, or another proper score (Matheson & Winkler, 1976; Gneiting & Raftery, 2007; Gneiting et al., 2008; Jordan et al., 2019). Exact sample-level feasibility and analytic density tractability are separate properties.

B.5 Training and inference interface

Algorithm 1 in the main paper gives the complete generic loop. The same interface supports tractable transformed likelihoods or sample-based proper scores; the Structural Feasibility Map is applied during both training and inference, so feasibility does not depend on a separate post-training repair step.

C Structural Feasibility Map Taxonomy

The taxonomy below is organized by structural representation rather than by optimization constraint form. Affine maps encode free coordinates directly; nonnegative/ordered maps compose a Nonnegative Structural Map with cumulative structure; normalized/compositional maps combine nonnegative components with normalization; explicit nonlinear-equality/manifold maps parameterize known manifolds or graphs; composed maps enforce intersections of structural requirements; and functional maps embed initial or boundary conditions in the output form. This axis is complementary to projection taxonomies based on linear equality, nonlinear equality, or convex inequality classes.

Table 5: Representative Structural Feasibility Maps. Coverage, boundary reachability, and boundary probability are distinct properties.

Constraint	**Feasible set**	**Structural map**	**Qualification**
Non-negativity	$\mathbb{R}^d_+$	$h_i(z) = \psi_+(z_i)$	ψ_{bm}: boundary mass; ψ_{br}: reachable boundary; ψ_{int}: interior only.
Bounded interval	$[a, b]^d$	$a + (b-a)\sin^2(z_i)$ or $a + (b-a)\sigma(z_i)$	$\sin^2$ reaches endpoints; sigmoid excludes them.
Weak ordering	$y_1 \le \cdots \le y_d$	$y_1 = z_1$, $y_k = y_{k-1} + \psi_+(z_k)$	Nonnegative increments; boundary behavior depends on ψ_+.
Ordering implementations	same	$\psi_{\mathrm{bm}}(z) = [z]_+$ or $[z]_+^p$; $\psi_{\mathrm{br}}(z) = z^2$; $\psi_{\mathrm{int}}(z) = \mathrm{softplus}(z)$	Positive-part family gives tie mass; square is boundary-reachable; softplus is interior.
Linear equality	$Ay = b$	$h(z) = y_0 + Nz$, $AN = 0$, $Ay_0 = b$	Exact and surjective for a full null-space basis; latent dimension $d - m$.
Simplex	Δ^{d-1}	$a_i = \psi_+(z_i)$, $y_i = a_i / \sum_j a_j$ when active; feasible fallback if all inactive	Normalized nonnegative structural map; face mass depends on ψ_+.
Simplex implementations	Δ^{d-1}	normalized ψ_{bm}, normalized ψ_{br}, or exponential-normalization ψ_{int}	Boundary-mass, boundary-reachable, and interior simplex laws, respectively.
Unit circle	$y_1^2 + y_2^2 = 1$	$h(z) = (\cos z, \sin z)$	Surjective, periodic, non-injective; one structural coordinate.

Constraint	Feasible set	Structural map	Qualification
Scale-shape FDC	$B \geq 0, R \geq 0, P \in \Delta^{K-1}, \Delta = RP$	$B = \psi_{\mathrm{bm}}(z_B)$, $R = \psi_{\mathrm{bm}}(z_R)$, $P = \mathrm{Norm}_{\psi_{\mathrm{bm}}}(z_P)$, $\Delta = RP$	Composed nonlinear equality; computational latent dimension may exceed intrinsic DOF.
Initial condition	$y(t_0) = y_0$	$y(t) = y_0 + (t - t_0) g_\phi(z, t)$	Exact at t_0; path coverage depends on g_ϕ.
Boundary condition	endpoint constraints	endpoint interpolant plus vanishing residual factor	Exact endpoints; interior expressiveness is model-dependent.

C.1 Primary weak-order map

For

$$\mathcal{C}_{\mathrm{ord}} = \{y \in \mathbb{R}^d : y_1 \leq \cdots \leq y_d\}, \tag{15}$$

a generic ordered Structural Feasibility Map uses nonnegative increments

$$\delta_k = \psi_+(z_k), \tag{16}$$

$$y_1 = z_1, \tag{17}$$

$$y_k = y_{k-1} + \delta_k, \qquad k = 2, \ldots, d. \tag{18}$$

Every realization is weakly ordered for any $\psi_+ : \mathbb{R} \to \mathbb{R}_{\geq 0}$. The reported primary experiment chooses the boundary-mass implementation $\psi_{\mathrm{bm}}(z) = [z]_+ = \max(0, z)$, using the same positive-part implementation specified in the main experiment. This cumulative-increment construction is distinct from monotone-network constraints on the input-output function itself, but it shares the broader feasible-by-design motivation of monotonic architectures (You et al., 2017; Wehenkel & Louppe, 2019; Runje & Shankaranarayana, 2023; Igel, 2024). Since the entire half-line $z_k \leq 0$ maps to zero increment under this implementation, a continuous latent law can induce an atom at an exact tie. Other members of the same boundary-mass class include $\psi_{\mathrm{bm}}(z) = [z]_+^p$ for $p > 0$ and thresholded positive parts $[z - \tau]_+$.

C.2 Primary simplex map

A generic normalized Structural Feasibility Map first constructs $a_i = \psi_+(z_i)$ and $S(z) = \sum_j a_j$. The reported primary experiment again uses the boundary-mass positive-part implementation $\psi_{\mathrm{bm}}(z) = [z]_+$. For $S(z) > 0$ define

$$y_i = \frac{a_i}{S(z)}. \tag{19}$$

The all-inactive event $S(z) = 0$ has positive probability under common continuous latent laws. The implementation resolves it by assigning the sample to the simplex vertex indexed by $\arg\max_i z_i$. This preserves feasibility and assigns the all-inactive event to a vertex. The fallback is permutation-equivariant when the maximizing coordinate is unique, hence almost surely under an absolutely continuous latent law. A fixed measurable tie-breaking convention defines the remaining cases; pointwise permutation equivariance at ties is not asserted. The event is explicitly monitored; in the final $d = 5$ primary experiment its mean rate is 3.81×10^{-4}. This reported frequency alone does not characterize fallback use throughout optimization or establish negligible gradient bias.

The map is surjective onto the closed simplex: for any $y \in \Delta^{d-1}$, the latent choice $z = y$ is a valid preimage. Its computational latent dimension is d, even though the simplex has $d - 1$ intrinsic degrees

of freedom. This closed-boundary emphasis differs from logistic-normal/softmax-style interior parameterizations common in compositional modeling (Aitchison & Shen, 1980; Aitchison, 1982); constructive simplex distributions and sparse mappings provide additional context for representation-dependent boundary behavior (Sethuraman, 1994; Martins & Astudillo, 2016).

C.3 Explicit nonlinear-equality maps

For the unit circle

$$\mathcal{C}_{\mathrm{circ}} = \{y \in \mathbb{R}^2 : y_1^2 + y_2^2 = 1\}, \tag{20}$$

the structural angle map

$$h_{\mathrm{circ}}(z) = (\cos z, \sin z) \tag{21}$$

is surjective onto the circle and uses one structural coordinate. It is periodic and therefore non-injective; `SLP-ProbHard` does not require a unique latent preimage for every feasible output. The matched radial reference maps ambient $u \in \mathbb{R}^2$ to $u/\|u\|_2$ with a deterministic feasible fallback for the exact-zero vector.

The real-world scale-shape construction is a composed map. For a seven-level weakly increasing FDC, let $K = 6$, $B = Q_0$, $\Delta_i = Q_{i+1} - Q_i$, $R = \sum_i \Delta_i$, and $P \in \Delta^{K-1}$. The structural implementation uses

$$B = [z_B]_+, \qquad R = [z_R]_+, \qquad P_i = \frac{[z_{P,i}]_+}{\sum_j [z_{P,j}]_+}, \qquad \Delta_i = RP_i, \tag{22}$$

with the same feasible simplex fallback used by the normalized positive-part map. It therefore enforces nonnegative baseline/range, compositional shape, the bilinear equalities $\Delta_i = RP_i$, and nonnegative increments simultaneously. For $d = 7$, the mathematical structural object has 7 intrinsic degrees of freedom but the implementation uses 8 computational latent coordinates because $P \in \Delta^5$ is represented with six normalized nonnegative coordinates.

D Detailed Structural-Map Remarks

D.1 Ordering and monotone paths

The ordering and monotone-path constructions use the same cumulative Nonnegative Structural Map principle. The constraint only requires $\psi_+(z) \geq 0$; probabilistic boundary behavior depends on which implementation is selected. Boundary-mass choices include the positive-part family $[z]_+^p$, $p > 0$, and thresholded positive parts $[z - \tau]_+$, because a latent region of positive measure maps exactly to zero. The reported primary experiment uses $p = 1$. Boundary-reachable choices such as z^2 can reach a tie but assign it zero probability under an absolutely continuous latent law. Interior choices such as softplus give strictly positive increments in real arithmetic. Floating-point underflow or rounding during cumulative addition can nevertheless produce exact numerical ties.

D.2 Simplex maps

The simplex Structural Feasibility Map can be constructed by normalizing nonnegative structural components, $a_i = \psi_+(z_i)$ and $y_i = a_i / \sum_j a_j$, together with a feasible rule for the all-inactive event when the chosen map permits it. The reported primary boundary-mass implementation uses $\psi_{\mathrm{bm}}(z) = [z]_+$ and an argmax-vertex fallback. Normalized positive-part powers remain boundary-mass alternatives. Normalized squares are boundary-reachable and are surjective onto the closed simplex (aside from the singular all-zero latent vector handled numerically), but individual coordinate zeros require a measure-zero latent event under continuous Gaussian laws. Exponential normalization (softmax/logistic-normal) is a classical interior simplex representation (Aitchison & Shen, 1980; Aitchison, 1982); sparsemax provides another mapping capable of exact zero coordinates (Martins & Astudillo, 2016). Thus "simplex map" is the structural family, while positive-part, square, softmax, and related transforms are implementations with different induced face probabilities.

D.3 Initial and boundary conditions

An initial condition $y(t_0) = y_0$ can be enforced by

$$y(t) = y_0 + (t - t_0)g_\phi(z, t). \tag{23}$$

Two endpoint conditions can be embedded in an interpolating ansatz plus a residual term multiplied by $(t - t_0)(t - t_1)$. These remain part of the general taxonomy but are not empirically benchmarked in this study.

E Complete Theory and Proofs

Scope of this theoretical appendix. This appendix collects established identities and approximation results used to analyze *SFLP*s. Rank-nullity, standard universal approximation, basic pushforward identities, truncated-Gaussian moments, and standard circular-distribution facts are used as foundations rather than claimed as standalone novelties. The `SLP-ProbHard` contribution is their specialization and organization to characterize representation coverage, boundary probability, stochastic coordinates, covariance/expressiveness, and parameterization bias within a structural probabilistic predictive law.

E.1 Pushforward and sample-level feasibility

Proposition 1 (Feasibility under pushforward)**.** *Let $\mathcal{C}(x)$ be measurable, let $h_{\phi,x} : \mathcal{Z} \to \mathcal{Y}$ be measurable with $h_{\phi,x}(\mathcal{Z}) \subseteq \mathcal{C}(x)$, and let $P_\theta^Z(\cdot \mid x)$ be any probability measure. If $P_\theta^Y = h_{\phi,x} \# P_\theta^Z$, then*

$$P_\theta^Y(\mathcal{C}(x) \mid x) = 1. \tag{24}$$

Proof. For $B = \mathcal{Y} \setminus \mathcal{C}(x)$, the image condition implies $h_{\phi,x}^{-1}(B) = \varnothing$. By the definition of pushforward, $P_\theta^Y(B \mid x) = 0$. □

Corollary 1 (Sample-level feasibility)**.** *Every realization $z^{(s)} \in \mathcal{Z}$ produces $h_\phi(x, z^{(s)}) \in \mathcal{C}(x)$.*

Proof. This is immediate from the image condition and holds independently for every sample. □

E.2 Affine moment propagation

Proposition 2 (Affine moment propagation)**.** *If $h(z) = Bz + c$, then*

$$\mathbb{E}[Y \mid x] = B\mathbb{E}[Z \mid x] + c, \tag{25}$$

$$\mathrm{Cov}(Y \mid x) = B\,\mathrm{Cov}(Z \mid x)B^\top. \tag{26}$$

The result provides exact first- and second-moment propagation for affine equalities and additive identities.

E.3 Distributional propagation through structural maps

Let

$$Z \mid x \sim P_\theta^Z(\cdot \mid x), \qquad Y = h(Z), \tag{27}$$

with latent mean $\mu_Z = \mathbb{E}[Z \mid x]$ and covariance $\Sigma_Z = \mathrm{Cov}(Z \mid x)$. The feasible predictive distribution is exactly the pushforward $P_\theta^Y = h \# P_\theta^Z$. Whenever the moments exist,

$$\mu_Y = \mathbb{E}[h(Z) \mid x], \tag{28}$$

$$\Sigma_Y = \mathbb{E}\left[(h(Z) - \mu_Y)(h(Z) - \mu_Y)^\top \mid x\right]. \tag{29}$$

These identities are exact for every structural map; what changes across constraint families is whether they admit closed-form evaluation. For affine $h(z) = Bz + c$, the preceding proposition gives $\mu_Y = B\mu_Z + c$ and $\Sigma_Y = B\Sigma_Z B^\top$. For nonlinear maps, `SLP-ProbHard` need not approximate the full predictive law as Gaussian: samples from P_θ^Z can be transformed through h directly. This is especially important when a closed structural map induces discrete probability mass on feasible boundaries.

Positive-part boundary-mass implementation. Consider the primary boundary-mass implementation $R = [Z]_+ = \max(0, Z)$ with

$$Z \sim \mathcal{N}(\mu, \sigma^2), \qquad \sigma > 0, \qquad \alpha = \mu/\sigma, \tag{30}$$

and let φ and Φ denote the standard-normal density and distribution function. The pushforward has an atom at the feasible boundary,

$$\Pr(R = 0) = \Phi(-\alpha), \tag{31}$$

and a continuous component on $(0, \infty)$. Its first two raw moments are

$$m(\mu, \sigma) \equiv \mathbb{E}[R] = \sigma\varphi(\alpha) + \mu\Phi(\alpha), \tag{32}$$

$$\mathbb{E}[R^2] = (\mu^2 + \sigma^2)\Phi(\alpha) + \mu\sigma\varphi(\alpha), \tag{33}$$

so

$$v(\mu, \sigma) \equiv \mathrm{Var}(R) = \mathbb{E}[R^2] - m(\mu, \sigma)^2. \tag{34}$$

These expressions follow by writing $Z = \mu + \sigma U$, $U \sim \mathcal{N}(0, 1)$, and evaluating the first two Gaussian tail integrals over $U > -\alpha$. Thus the latent location and scale determine both ordinary predictive moments and the probability assigned exactly to the feasible boundary.

Weak ordering with independent Gaussian coordinates. For the primary ordering map,

$$Y_1 = Z_1, \tag{35}$$

$$Y_k = Z_1 + \sum_{j=2}^{k} [Z_j]_+, \qquad k = 2, \ldots, d, \tag{36}$$

assume the primary diagonal-Gaussian latent model $Z_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$ independently. Define $m_j = m(\mu_j, \sigma_j)$ and $v_j = v(\mu_j, \sigma_j)$ using Eqs. (32)-(34). Then

$$\boxed{\mathbb{E}[Y_k]} = \mu_1 + \sum_{j=2}^{k} m_j, \tag{37}$$

$$\boxed{\mathrm{Cov}(Y_k, Y_\ell)} = \sigma_1^2 + \sum_{j=2}^{\min(k,\ell)} v_j. \tag{38}$$

Moreover, the adjacent equality event has exact probability

$$\boxed{\Pr(Y_k = Y_{k-1})} = \Phi\left(-\frac{\mu_k}{\sigma_k}\right), \qquad k \geq 2. \tag{39}$$

The moment identities follow from linearity and independence: two ordered outputs share Z_1 and exactly the increment terms up to $\min(k, \ell)$. Hence the network-predicted latent location and scale parameters directly determine the predictive probability of an exact tie.

Nonnegative ordering. For the composed map

$$Y_k = \sum_{j=1}^{k} [Z_j]_+, \qquad k = 1, \ldots, d, \tag{40}$$

the same independent diagonal-Gaussian assumption gives

$$\boxed{\mathbb{E}[Y_k]} = \sum_{j=1}^{k} m_j, \tag{41}$$

$$\boxed{\mathrm{Cov}(Y_k, Y_\ell)} = \sum_{j=1}^{\min(k,\ell)} v_j. \tag{42}$$

The adjacent-tie probability remains

$$\boxed{\Pr(Y_k = Y_{k-1})} = \Phi\left(-\frac{\mu_k}{\sigma_k}\right), \qquad k \geq 2, \tag{43}$$

while independence yields the exact zero-state probability

$$\boxed{\Pr(Y_k = 0)} = \prod_{j=1}^{k} \Phi\left(-\frac{\mu_j}{\sigma_j}\right). \tag{44}$$

Thus the same structural coordinates control both ordering ties and exact zero-valued states.

Boundary-reachable square map. For $R = Z^2$ with $Z \sim \mathcal{N}(\mu, \sigma^2)$,

$$\mathbb{E}[R] = \mu^2 + \sigma^2, \tag{45}$$

$$\mathrm{Var}(R) = 2\sigma^4 + 4\mu^2\sigma^2. \tag{46}$$

Nevertheless, $\Pr(R = 0) = 0$ under a continuous Gaussian latent law. Boundary reachability and positive boundary probability therefore remain distinct even when output moments are analytically available.

Normalized boundary-mass simplex map. For

$$A_i = [Z_i]_+, \qquad Y_i = \frac{A_i}{\sum_j A_j} \tag{47}$$

whenever at least one coordinate is active, define the random active set $\mathcal{A} = \{i : Z_i > 0\}$. Under conditionally independent Gaussian coordinates with $\sigma_i(x) > 0$, for any fixed nonempty index set S,

$$\Pr(\mathcal{A} = S \mid x) = \prod_{i \in S} \Phi\left(\frac{\mu_i}{\sigma_i}\right) \prod_{j \notin S} \Phi\left(-\frac{\mu_j}{\sigma_j}\right), \qquad S \neq \varnothing. \tag{48}$$

Here and below the dependence of μ_i, σ_i on x is suppressed. Conditional on $\mathcal{A} = S \neq \varnothing$, the predictive sample lies in the relative interior of the simplex face indexed by S. These are latent active-set probabilities. For a singleton $S = \{i\}$, the output vertex probability additionally includes mass routed to e_i from the all-inactive event:

$$\Pr(Y = e_i \mid x) = \Pr(\mathcal{A} = \{i\} \mid x) + \Pr(\mathcal{A} = \varnothing,\ I(Z) = i \mid x),$$

where $I(Z)$ is the fallback's maximizing index, including its tie-breaking convention. The all-inactive event has probability

$$p_\varnothing(x) = \prod_{j=1}^{d} \Phi\left(-\frac{\mu_j}{\sigma_j}\right), \tag{49}$$

and is handled by the feasible argmax-vertex fallback in Appendix C.2. Thus the structural push-forward is a mixture over simplex faces, with the additional fallback contribution included at the vertices. The mean can be decomposed as

$$\mathbb{E}[Y \mid x] = \sum_{S \neq \varnothing} \Pr(\mathcal{A} = S \mid x)\mathbb{E}[Y \mid \mathcal{A} = S, x] + p_\varnothing(x)\mathbb{E}[Y \mid \mathcal{A} = \varnothing, x], \tag{50}$$

with covariance obtained analogously by the law of total covariance. The conditional ratio moments generally do not reduce to simple elementary closed forms and are evaluated by Monte Carlo or numerical integration.

Circular structural map. If $Z \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = (\cos Z, \sin Z)$, then the first two moments are exact:

$$\mathbb{E}[\cos Z] = e^{-\sigma^2/2}\cos\mu, \qquad \mathbb{E}[\sin Z] = e^{-\sigma^2/2}\sin\mu, \tag{51}$$

$$\mathbb{E}[\cos^2 Z] = \tfrac{1}{2}\left(1 + e^{-2\sigma^2}\cos 2\mu\right), \qquad \mathbb{E}[\sin^2 Z] = \tfrac{1}{2}\left(1 - e^{-2\sigma^2}\cos 2\mu\right), \tag{52}$$

$$\mathbb{E}[\sin Z \cos Z] = \tfrac{1}{2}e^{-2\sigma^2}\sin 2\mu. \tag{53}$$

Thus the circle benchmark supplies a nonlinear pushforward whose moments remain analytically tractable even though the induced Cartesian law is non-Gaussian.

Smooth nonlinear maps. For a differentiable structural map h with latent mean μ_Z and covariance Σ_Z, the first-order delta approximation (Oehlert, 1992) is

$$\mathbb{E}[Y] \approx h(\mu_Z), \tag{54}$$

$$\mathrm{Cov}(Y) \approx J_h(\mu_Z)\Sigma_Z J_h(\mu_Z)^\top. \tag{55}$$

This approximation is optional rather than required for feasibility. Exact pushforward sampling remains available whenever moments or densities are analytically inconvenient. In particular, a Gaussian moment approximation should not replace the full pushforward law when boundary atoms or face-supported probability mass are scientifically relevant.

Structural map	Predictive-law structure	Moment propagation	Boundary characterization
Affine equality	affine pushforward	exact mean/covariance	feasible affine manifold
Positive-part boundary-mass	atom at zero + continuous positive component	exact for Gaussian latent	$\Pr(Y = 0) = \Phi(-\mu/\sigma)$
Weak ordering	cumulative mixed law	exact for independent Gaussian coordinates	exact adjacent-tie probabilities
Nonnegative ordering	cumulative mixed law	exact for independent Gaussian coordinates	exact zero and tie probabilities
Square	squared-Gaussian transform	exact Gaussian moments	boundary reachable, zero mass
Boundary-mass normalized simplex	mixture over simplex faces	conditional/numerical moments	analytic active-set probabilities
Smooth generic map	nonlinear pushforward	delta approximation or Monte Carlo	map-dependent

Table 6: Distributional propagation under representative structural maps. Analytic tractability is separate from hard feasibility: every listed map is defined through its pushforward law, while closed-form summaries are used when available.

E.4 Samplewise differentiation and expected-score gradients

Proposition 3 (Samplewise chain rule)**.** *Fix x and θ_0, and let h be a fixed structural map. Assume that $\theta \mapsto z_\theta(x,\epsilon)$ is differentiable at θ_0 almost surely and that h is differentiable at $z_{\theta_0}(x,\epsilon)$ almost surely. Then $\theta \mapsto h(z_\theta(x,\epsilon))$ is differentiable at θ_0 almost surely, with*

$$D_\theta h(z_\theta(x,\epsilon))\big|_{\theta_0} = J_h(z_{\theta_0}(x,\epsilon)) D_\theta z_\theta(x,\epsilon)\big|_{\theta_0}.$$

Proof. Apply the ordinary chain rule on the probability-one event where both derivatives exist. □

This samplewise statement does not by itself justify differentiating an expected score. Let $L(\theta) = \mathbb{E}_\epsilon[\ell(\theta,\epsilon)]$, where ℓ may include a complete predictive ensemble and its sample-based score. A sufficient condition for differentiation at θ_0 is that $\ell(\theta_0,\epsilon)$ is integrable, that $\ell(\cdot,\epsilon)$ is differentiable at θ_0 almost surely, and that it is locally Lipschitz in a neighborhood of θ_0 with an integrable random Lipschitz bound. Dominated convergence applied to difference quotients then gives

$$\nabla_\theta L(\theta_0) = \mathbb{E}_\epsilon[\nabla_\theta \ell(\theta_0,\epsilon)].$$

The gradient of an expected finite-ensemble score is a separate question from whether that score estimator is unbiased for a population scoring rule. The positive-part map is continuous and globally Lipschitz, despite its kink at zero. The normalized positive-part simplex map, however, contains an argmax-vertex fallback on the all-inactive region and is discontinuous at some boundaries between fallback choices. Almost-everywhere samplewise differentiation therefore does not automatically establish unbiased expected-score gradients for the complete map. The corresponding estimator requires separate justification or an explicitly reported approximation. Reparameterization methods for the latent sampler (Figurnov et al., 2018; Jankowiak & Obermeyer, 2018) do not by themselves resolve discontinuities in the structural map. A small reported fallback frequency does not establish negligible gradient bias.

E.5 Affine rank-nullity

Proposition 4 (Free-coordinate dimension)**.** *Let $A \in \mathbb{R}^{m\times d}$ have rank $m < d$ and assume $Ay = b$ is consistent. Then $\mathcal{C} = \{y : Ay = b\}$ has dimension $q = d - m$. If $N \in \mathbb{R}^{d\times q}$ has linearly independent columns spanning $\ker(A)$, every feasible point has a unique representation $y = y_0 + Nz$.*

Proof. The rank-nullity theorem gives $\dim \ker(A) = d - m$. For any feasible y, $A(y - y_0) = 0$, so $y - y_0 \in \ker(A)$ and equals Nz; linear independence makes z unique. □

E.6 Constrained universal approximation

Theorem 1 (Approximation consequence under a continuous structural lifting)**.** *Let $\mathcal{X} \subset \mathbb{R}^p$ be compact and $h : \mathcal{X} \times \mathbb{R}^q \to \mathbb{R}^d$ continuous with $h(x, z) \in \mathcal{C}(x)$. Let $F^\star : \mathcal{X} \to \mathbb{R}^d$ be continuous and feasible. Assume there exists a continuous lifting $z^\star : \mathcal{X} \to \mathbb{R}^q$ with $F^\star(x) = h(x, z^\star(x))$, and that the latent network family is a universal approximator of continuous maps into $\mathbb{R}^q$. Then $h(x, f_\theta(x))$ can uniformly approximate $F^\star$ while remaining feasible for all x.*

If h is uniformly L_h-Lipschitz in z,

$$\sup_x \|h(x, f_\theta(x)) - F^\star(x)\| \le L_h \sup_x \|f_\theta(x) - z^\star(x)\|. \tag{56}$$

The unconstrained approximation ingredient follows from standard universal-approximation results (Cybenko, 1989; Hornik, 1991; Leshno et al., 1993). This theorem is a direct structural consequence of standard universal approximation once a continuous feasible lifting exists; its role here is to characterize when a chosen representation preserves approximation capacity. The substantive assumption is therefore the existence of the lifting, a property of the chosen representation.

Corollary 2 (Approximation for weak-order and simplex targets)**.** *Let X be compact and let the latent network family uniformly approximate continuous maps into the required Euclidean latent space. Every continuous weakly ordered target can be uniformly approximated through the primary weak-order map. Every continuous simplex-valued target can also be uniformly approximated through the normalized positive-part map with its feasible fallback.*

Proof. For a weakly ordered vector y, the continuous lifting

$$s_{\mathrm{ord}}(y) = (y_1, y_2 - y_1, \ldots, y_d - y_{d-1})$$

satisfies $h_{\mathrm{ord}}(s_{\mathrm{ord}}(y)) = y$. The weak-order map is continuous, so Theorem 1 applies.

For a simplex-valued target $F^\star$, use the continuous lifting $z^\star(x) = F^\star(x)$. Choose a latent approximator satisfying

$$\sup_{x \in X} \|f_\theta(x) - F^\star(x)\|_\infty < \delta < \frac{1}{2d}.$$

Write $a(x) = [f_\theta(x)]_+$ and $S(x) = \sum_i a_i(x)$. Since $F_i^\star(x) \ge 0$ and $\sum_i F_i^\star(x) = 1$,

$$S(x) \ge 1 - d\delta > \frac{1}{2}.$$

The fallback is therefore never used. Moreover,

$$\begin{aligned} \|h_\Delta(f_\theta(x)) - F^\star(x)\|_1 &\le \|a(x)/S(x) - a(x)\|_1 + \|a(x) - F^\star(x)\|_1 \\ &= |1 - S(x)| + \|a(x) - F^\star(x)\|_1 \\ &\le 2\|a(x) - F^\star(x)\|_1 \\ &\le 2d\delta. \end{aligned}$$

Taking δ arbitrarily small proves uniform approximation. This local argument does not require global continuity of the fallback map. □

Remark 1 (Circle topology)**.** *The circle map $h_{\mathrm{circ}}(z) = (\cos z, \sin z)$ is continuous and surjective, but it admits no global continuous section $S^1 \to \mathbb{R}$; this is the standard covering-space/topological obstruction associated with the angle map (Hatcher, 2002): a globally continuous choice of angle would contradict the topology of the circle. A measurable branch-cut section exists, so surjectivity still supports the measurable distributional-approximation result below, while Theorem 1 does not apply globally through a continuous angle lifting. This separates map surjectivity from existence of a continuous inverse/section.*

E.7 Distributional approximation and parameterization bias

Theorem 2 (Classical pushforward approximation consequence under a measurable section)**.** *Let $h : \mathcal{Z} \to \mathcal{C}$ be measurable and surjective between standard Borel spaces, and suppose a measurable section $s : \mathcal{C} \to \mathcal{Z}$ exists with $h(s(y)) = y$. If a latent family is dense in probability measures on $\mathcal{Z}$ under total variation, then its pushforward family is dense in probability measures on $\mathcal{C}$ under total variation.*

Proof. Let ν be a probability measure on $\mathcal{C}$ and define $P = s_{\#}\nu$. Since $h \circ s$ is the identity on $\mathcal{C}$, $h_{\#}P = \nu$. By the assumed total-variation density, choose latent laws P_n with $\|P_n - P\|_{\mathrm{TV}} \to 0$. Then

$$\|h_{\#}P_n - \nu\|_{\mathrm{TV}} = \sup_B |P_n(h^{-1}(B)) - P(h^{-1}(B))| \le \|P_n - P\|_{\mathrm{TV}} \to 0,$$

where the supremum is over measurable subsets of $\mathcal{C}$. □

This result is conditional on the stated density assumption and is not a universal-expressiveness guarantee for the Gaussian heads used in the experiments. In particular, absolutely continuous latent families cannot approximate every atomic probability measure on Euclidean latent space in total variation.

This standard pushforward consequence is included to characterize distributional expressiveness of a chosen representation, not as a new measure-theoretic result.

Proposition 5 (Wasserstein stability)**.** *If h is L_h-Lipschitz, then $W_p(h \# P, h \# Q) \le L_h W_p(P, Q)$ whenever the moments exist, using the standard Wasserstein/optimal-transport metric framework (Villani, 2009). More generally, transport-map formulations provide a useful probability-theoretic lens for deterministic pushforwards of latent laws (Marzouk et al., 2016).*

Proposition 6 (Parameterization bias)**.** *Let $h : \mathcal{Z} \to \mathcal{C}$ be measurable, and assume $H = h(\mathcal{Z})$ is measurable. Every pushforward law $P^Y = h_{\#}P^Z$ satisfies $P^Y(H) = 1$. Consequently, any target law ν with $\nu(\mathcal{C} \setminus H) > 0$ cannot be represented exactly by this map.*

Proof. Because $h(z) \in H$ for every $z \in \mathcal{Z}$, $h^{-1}(H) = \mathcal{Z}$, and hence $P^Y(H) = 1$. □

Reachability, topological support, and event probability are distinct. A point outside $h(\mathcal{Z})$ may belong to the topological support as a limit point. For example, softplus never outputs zero in real arithmetic, but zero belongs to the topological support of its pushforward under a nondegenerate Gaussian latent law.

E.8 Classical identity used as a reachability and boundary-mass taxonomy

Proposition 7 (Boundary-preimage identity)**.** *Let $B \subseteq \mathcal{C}$ be measurable and let $P^Y = h \# P^Z$. Then*

$$P^Y(B) = P^Z\left(h^{-1}(B)\right). \tag{57}$$

Consequently: (i) if $h^{-1}(B) = \varnothing$, the boundary event is excluded; (ii) if $h^{-1}(B)$ is nonempty but has zero latent probability, the boundary is reachable but receives no predictive mass; and (iii) if the preimage has positive latent probability, the predictive law has positive boundary mass.

Proof. This is the defining identity of the pushforward measure applied to the measurable set B. □

The following image and boundary-mass statements concern real arithmetic. Floating-point underflow and rounding can change exact numerical zero or tie diagnostics.

This elementary identity becomes a useful design taxonomy when applied to constrained probabilistic maps. For an ordering increment Δ and boundary event $B = \{\Delta = 0\}$:

- $\Delta = \mathrm{softplus}(Z)$ has empty finite preimage and excludes the boundary;
- $\Delta = Z^2$ has preimage $\{0\}$, which has probability zero under an absolutely continuous latent law;

- $\Delta = [Z]_+$ has preimage $(-\infty, 0]$, which has positive probability whenever the latent law assigns positive mass there. For $Z \sim \mathcal{N}(\mu, \sigma^2)$, this mass is $\Phi(-\mu/\sigma)$.

For the normalized positive-part boundary-mass simplex, the face event $\{Y_i = 0\}$ includes latent states with $Z_i \leq 0$ and at least one other active coordinate, and therefore can have positive probability under a continuous latent law. By contrast, normalized squares require the measure-zero event $Z_i = 0$ for a coordinate to be exactly zero, and softmax excludes zero entirely. Feasibility, boundary reachability, and boundary probability are therefore distinct modeling properties.

E.9 Finite precision

All feasibility theorems above are real-arithmetic statements. Matrix products, summation, null-space reconstruction, and projection can introduce small residuals in float32; this is the standard distinction between exact real-arithmetic identities and their floating-point evaluation (Goldberg, 1991; Higham, 2002). A numerical threshold is therefore an evaluation convention, not a relaxation of the mathematical constraint. Continuous residuals are the primary numerical evidence; thresholded violation rates are secondary diagnostics.

F Extended Comparison with Projection-Based Methods

F.1 Mechanism-centered comparison

Method	Primary mechanism	Probabilistic predictive law	Boundary behavior	Main qualification
DC3 (Donti et al., 2021)	Equality completion plus iterative inequality correction	Not a dedicated probabilistic framework	Completion or correction dependent	Requires correction machinery.
Frerix et al. (2020)	Direct parameterization for homogeneous linear inequalities	Deterministic focus	Represented feasible boundary available	Specialized constraint class.
ProbHard-E2E (Utkarsh et al., 2026)	Differentiable probabilistic projection	Yes	Determined by projected law	Broad supported constraint classes; projection machinery required.
Soft-Radial (Schneider & Kuhn, 2026)	Radial map into convex-set interior	Not central	Interior-oriented	Requires convex geometry and an interior anchor.
`SLP-ProbHard`	Stochastic latent law plus structural feasibility map	Yes	Map-dependent: excluded, reachable-only, or positive mass	Requires a suitable structural map.

Table 7: Representative mechanisms. Several approaches provide hard feasibility; `SLP-ProbHard` is distinguished by making a constraint-specific structural map part of a probabilistic pushforward model.

F.2 Detailed comparison with ProbHardE2E

ProbHardE2E and `SLP-ProbHard` target complementary regimes. Projection is preferable when a feasible set is implicit, complex, or easier to project onto than to parameterize. `SLP-ProbHard` is attractive when an explicit generative representation is available and scientifically acceptable. The compatible affine experiment below gives a direct empirical comparison using the frozen official source and a verified generic A, b adapter; it is intentionally narrower than ProbHardE2E's full supported constraint scope.

Table 8: Mechanistic comparison with probabilistic projection.

Criterion	**Projection / ProbHardE2E family**	`SLP-ProbHard`	**Implication**
Core mechanism	Unconstrained probabilistic representation followed by projection/correction	$Y = h(Z)$ with $h(\mathcal{Z}) \subseteq \mathcal{C}$	Different constructions of the feasible law.
Organizing taxonomy	Linear/nonlinear equalities and convex inequalities under projection	Affine, nonnegative/ordered, normalized/compositional, composed, and functional structural families	Different axes of generality.
Constraint generality	Broad for supported equality and convex-inequality classes	Selective and map-dependent	Projection is generally broader.
Linear equality coordinates	Ambient prediction before projection	$q = d - m$ free coordinates	Exact intrinsic reduction in this family.
Nonlinear coordinates	Ambient or method-specific	May be intrinsic or redundant	Dimensional benefit is not universal.
Projection metric	Selected geometry may matter	No projection metric for explicit maps	Removes metric choice.
Solver / correction stage	Required by the chosen projection/correction mechanism	None when map is explicit	Strongest for maps that replace nontrivial solves.
Boundary probability	Induced by projection geometry and base law	Induced by structural-map preimages and latent law	Both can create boundary mass; mechanisms differ.
Affine moments	Can be exact	Exact	Tie in tractability for affine maps.
Main risk	Projection cost/conditioning or solver complexity	Parameterization bias and map-induced optimization effects	Complementary failure modes.

G Experimental Protocol and Reproducibility

G.1 Frozen benchmark families and held-out seeds

The reported evidence uses development-independent seed ranges after map and soft-penalty choices were frozen:

- linear primary: seeds 20-29; linear scaling: 30-34; covariance ablation: 40-44;
- ordering primary: seeds 60-69; ordering scaling: 70-74;
- real-world weak-order Benchmark A and composed nonnegative-ordering Benchmark B: seeds 80-89, intentionally matched across A/B;
- simplex primary: seeds 100-109; simplex scaling: 110-114;
- synthetic composed nonnegative-ordering Benchmark B: seeds 115-119;
- nonlinear circle primary: seeds 130-139;

- real-world affine FDC coherence: seeds 150-159;
- real-world nonlinear scale-shape coherence: seeds 170-179.

All reported training experiments use float32 on CPU. Primary probabilistic evaluations use 100 predictive samples. Smoke/development seeds are disjoint from the reported primary seeds.

Frozen item	Value used in the primary synthetic benchmarks
Train / validation / test size	3000 / 750 / 750
Backbone	MLP, hidden width 128, depth 3, dropout 0
Batch size	128
Maximum epochs / early stopping	60 / patience 20, minimum improvement 10^{-6}
Learning rate / weight decay	10^{-3} / 10^{-5}
Gradient clipping	5.0 where exposed by the frozen runner
Monte Carlo samples	12 training, 32 validation, 100 evaluation
Probabilistic objective	0.5 marginal CRPS + 0.5 Energy Score
Precision / device	float32 / CPU
Primary network dimensions	input 12; outputs 11 (linear), 5 (ordering/simplex), 2 (circle)

Table 9: Primary synthetic training configuration. Competing methods share the same backbone/training protocol within each benchmark; only the constraint mechanism and corresponding output/latent head differ where required.

G.2 SYNTHETIC-DGP TRANSPARENCY AND FAIRNESS

The synthetic tasks are diagnostic benchmarks rather than claims of universal DGP dominance. Table 10 makes structural alignment explicit. Alignment can favor an `SLP-ProbHard` representation, so the scientific question for each task is stated separately from raw score comparison.

Benchmark	Frozen target-generation structure	Same structure as primary `SLP-ProbHard` map?	Potential `SLP-ProbHard` alignment	Intended scientific question
Affine	Feasible latent regression is mapped through a fixed affine null-space representation $y = y_0 + Nz$.	Yes	Yes: coordinates match the feasible DGP	Can a probabilistic model learn only the $q = d - m$ free directions while matching ambient projection?
Weak ordering	A base response is combined with cumulative nonnegative positive-part increments, producing exact-tie mass.	Yes	Yes: boundary mechanism is aligned	Does a boundary-mass ordering representation change tie probability and proper scores relative to isotonic projection?
Simplex	Nonnegative components with an exact-zero mechanism are normalized to the simplex, producing face-supported observations.	Yes, at the structural level	Yes: boundary-capable normalization is aligned	How do direct boundary-capable normalization and Euclidean simplex projection differ in face probability and predictive scoring?
Circle	A conditional angle is drawn from a von Mises law and mapped as $(\cos\Theta, \sin\Theta)$. Learned `SLP-ProbHard` uses a Gaussian angle; projection uses an ambient bivariate Gaussian.	Geometry yes; latent family no	Partial	Does one structural angle provide competitive hard-feasible prediction relative to ambient radial projection under a non-Gaussian angular DGP?
Composed nonneg. order	Targets are generated directly by cumulative positive-part from the first coordinate onward.	Yes, exactly	Yes, deliberately strong	Mechanism validation: do the predicted zero/tie boundary effects follow the preimage analysis when the structural DGP is known?

Table 10: Synthetic-DGP transparency. The table separates structural alignment from the question each benchmark is designed to test. The composed benchmark is deliberately favorable and is interpreted only as mechanism validation.

G.3 BASELINES

Linear equality compares unconstrained prediction, validation-selected soft penalty, post-hoc orthogonal projection, end-to-end orthogonal projection (Projection-Or), covariance-weighted/oblique projection, and `SLP-ProbHard`. Ordering compares unconstrained output, soft crossing penalty, post-hoc sorting, post-hoc isotonic correction, end-to-end isotonic projection, and positive-part SLP. Simplex compares unconstrained output, soft constraint loss, post-hoc simplex projection, end-to-end Euclidean simplex projection, and normalized positive-part SLP. Circle compares unconstrained

bivariate Gaussian output, soft radius penalty, post-hoc radial projection, end-to-end radial projection, and one-coordinate circular SLP. The real-world affine extension uses unconstrained/soft augmented outputs, post-hoc affine projection, end-to-end affine projection, and structural affine SLP. The real-world nonlinear extension uses unconstrained/soft (B, R, P, Δ) outputs, post-hoc alternating feasible correction, end-to-end alternating feasible correction, and structural scale-shape SLP.

G.4 Shared hydrological FDC source and preprocessing

The hydrological experiments reuse one frozen dataset created before the affine and nonlinear extensions. Daily river discharge comes from the *GloFAS v5.0 hydrological reanalysis* of the European Commission Joint Research Centre (Grimaldi et al., 2026); GloFAS uses the LISFLOOD hydrological modelling system (Van Der Knijff et al., 2010; Burek et al., 2013). Seven Emilia-Romagna basins are used: Crostolo, Enza, Panaro, Parma, Secchia, Taro, and Trebbia. Basin-aggregated precipitation, minimum-temperature, and maximum-temperature descriptors are derived from ERACLITO61 (Antolini et al., 2016), together with season indicators, basin area, and basin identity. The common 1996-2022 period yields 756 seasonal basin-season-year samples. The temporal split is 532 training samples (1996-2014), 112 validation samples (2015-2018), and 112 held-out test samples (2019-2022), with 22 standardized predictor features.

Using the conventional exceedance-probability interpretation of flow-duration curves (Burgan & Aksoy, 2018; Vogel & Fennessey, 1994; 1995) and the multi-output FDC precedent of Worland et al. (2019), the seven FDC components are

$$Q = [Q_{95}, Q_{90}, Q_{75}, Q_{50}, Q_{25}, Q_{10}, Q_{05}], \tag{58}$$

where Q_{95} is the ordinary 5th percentile and Q_{05} the ordinary 95th percentile of the seasonal daily discharge distribution. Thus physically admissible FDCs satisfy $0 \le Q_{95} \le \cdots \le Q_{05}$. The original ordering experiment uses $\log(1 + Q)$ targets and transforms samples back by $\exp(y) - 1$ without clipping.

The affine and nonlinear extensions do *not* rerun the daily-discharge, predictor-NetCDF, basin-mask, FDC-quantile, or split construction. They read the frozen raw Q arrays and the already-standardized x arrays. A single positive multiplicative scale,

$$s_Q = \sqrt{\mathrm{mean}(Q_{\mathrm{train}}^2)} = 43.7190796, \tag{59}$$

is fitted on training raw FDC values only. No centering is used. The scaled target $Q^\star = Q/s_Q$ preserves both adjacent-difference and scale-shape equalities exactly. The frozen source SHA-256 is `18cd85cac0405b740866c088a61566c93e4ec6d020bbf881a9f33ee154afff47`.

G.5 Real-world nonnegative ordering

The composed map

$$Q_k = \sum_{j=1}^{k} [z_j]_+ \tag{60}$$

guarantees $0 \le Q_1 \le \cdots \le Q_d$. The matched hard reference is the frozen nonnegative isotonic projection. This representation directly tests physical order/non-negativity and boundary mass in the FDC itself.

G.6 Real-world affine FDC coherence

For scaled $Q^\star \in \mathbb{R}^7$, define six adjacent increments

$$\Delta_i = Q^\star_{i+1} - Q^\star_i, \qquad i = 0, \ldots, 5, \tag{61}$$

and augment the target as

$$Y_{\mathrm{aff}} = (Q^\star_0, \ldots, Q^\star_6, \Delta_0, \ldots, \Delta_5) \in \mathbb{R}^{13}. \tag{62}$$

The six affine constraints are

$$r_i(Y) = Q^\star_i - Q^\star_{i+1} + \Delta_i = 0. \tag{63}$$

If $A \in \mathbb{R}^{6\times 13}$ collects these rows, $\mathrm{rank}(A) = 6$, so the affine feasible set has intrinsic dimension $13 - 6 = 7$. The structural map takes seven stochastic coordinates,

$$h_{\mathrm{aff}}(z) = (z, \mathrm{diff}(z)), \tag{64}$$

and therefore satisfies every equality algebraically. The end-to-end affine reference predicts a 13-dimensional diagonal Gaussian and applies the exact Euclidean projection onto $AY = 0$. Hence the hard comparison is 13 ambient stochastic coordinates versus 7 structural coordinates, a 46.1538% reduction.

This benchmark intentionally isolates *affine coherence only*. Neither the affine projection nor h_{aff} additionally enforces non-negativity or FDC ordering; negative-value and crossing rates are therefore reported only as secondary diagnostics and must not be described as violations of the affine constraint. The validation-selected soft weight is 0.01; primary seeds are 150-159. Training uses the same 128-width/depth-3 MLP, batch 128, 60-epoch cap, 10^{-3} learning rate, 10^{-5} weight decay, 12/32 Monte Carlo samples for train/validation, and 100 test samples.

Primary predictive reporting is in reconstructed raw-Q space, $Q = s_Q Q^\star$. Augmented $(Q^\star, \Delta^\star)$ scores are secondary. Constraint metrics use r_i and report mean absolute residual, mean squared residual, maximum absolute residual, and vector violation rate $\mathbf{1}[\max_i |r_i| > 10^{-5}]$.

G.7 Real-world nonlinear scale-shape coherence

For the same scaled FDC, define

$$B = Q_0^\star, \qquad \Delta_i = Q_{i+1}^\star - Q_i^\star, \qquad R = \sum_{i=0}^{5} \Delta_i = Q_6^\star - Q_0^\star. \tag{65}$$

For $R > 0$, define the compositional shape

$$P_i = \frac{\Delta_i}{R}, \qquad P_i \geq 0, \qquad \sum_i P_i = 1. \tag{66}$$

The augmented target is

$$Y_{\mathrm{nl}} = (B, R, P_0, \ldots, P_5, \Delta_0, \ldots, \Delta_5) \in \mathbb{R}^{14}, \tag{67}$$

with nonlinear bilinear equalities

$$\Delta_i = RP_i, \qquad i = 0, \ldots, 5. \tag{68}$$

No train/validation/test FDC is flat at the frozen scaled threshold $R \leq 10^{-8}$ (flat counts 0/0/0). The protocol nevertheless freezes the convention $P = e_1$ for an exactly flat FDC, under which $\Delta = RP = 0$ remains feasible.

The structural map uses

$$B = [z_B]_+, \qquad R = [z_R]_+, \tag{69}$$

$$P = \mathrm{NormPosPart}(z_P), \qquad \Delta = RP. \tag{70}$$

It therefore enforces $B \geq 0$, $R \geq 0$, $P \geq 0$, $\sum_i P_i = 1$, $\Delta = RP$, and consequently nonnegative increments and a nonnegative weakly ordered reconstructed FDC. For $d = 7$, the mathematical intrinsic DOF is 7. The boundary-capable normalized positive-part shape uses six computational coordinates for the five-dimensional simplex, so `SLP-ProbHard` uses $1 + 1 + 6 = 8$ computational stochastic coordinates. The ambient correction model predicts all 14 coordinates; the reduction is 42.8571%. This is reported as a *computational* stochastic-coordinate reduction, not as an 8-dimensional intrinsic manifold claim.

The end-to-end hard comparator is a five-iteration differentiable alternating feasible correction. Given ambient predictions $(B_0, R_0, P_0, \Delta_0)$, it projects B_0 to $\mathbb{R}_+$ and alternates

$$R \leftarrow \left[\frac{R_0 + P^\top \Delta_0}{1 + \|P\|_2^2}\right]_+, \tag{71}$$

$$P \leftarrow \Pi_\Delta\left(\frac{P_0 + R\Delta_0}{1 + R^2}\right), \tag{72}$$

then sets $\Delta = RP$. These are exact conditional minimizers for the two blocks of a feasible-correction objective, but the finite alternating procedure is *not claimed to be the exact joint Euclidean projection*. The validation-selected soft weight is 0.01; primary seeds are 170-179.

Primary nonlinear feasibility metrics are

$$\mathrm{CE}_{\mathrm{bilinear}} = \mathrm{mean}_{i,s} \, |\Delta_i - RP_i|, \tag{73}$$

$$\mathrm{CE}_{\mathrm{simplex}} = \mathrm{mean}_s \left| \sum_i P_i - 1 \right|, \tag{74}$$

together with maxima, range-conservation diagnostics, a joint 10^{-5} nonlinear violation rate, and reconstructed raw-Q negative/order violation rates. Raw-Q probabilistic metrics are primary.

G.8 Predictive and constraint metrics

MSE and MAE are computed from the predictive mean. Marginal CRPS is averaged across observations and coordinates, and multivariate Energy Score across observations. Central 90% marginal interval coverage and width are reported. Circle additionally reports circular/geodesic MAE and the equality residual $y_1^2 + y_2^2 - 1$. Affine FDC uses the six residuals $Q_i - Q_{i+1} + \Delta_i$; nonlinear FDC uses bilinear, simplex, range-conservation, nonnegativity, and reconstructed-order diagnostics. Boundary frequencies remain distributional diagnostics rather than feasibility metrics.

G.9 Covariance ablation

The covariance ablation uses the frozen linear benchmark and seeds 40-44. It compares the standard Projection-Or reference with `SLP-ProbHard` using diagonal, rank-4 low-rank-plus-diagonal, and full covariance latent heads. The purpose is to test whether the diagonal latent law limits `SLP-ProbHard` probabilistic expressiveness; it is not a matched full-covariance projection study.

G.10 Controlled computational profiling

Controlled profiling performs no training. Under a fixed CPU environment (8 threads, float32), it measures (i) the feasibility operator alone and (ii) a matched MLP + diagonal Gaussian sampling + feasibility operator pipeline. Each configuration uses 100 predictive samples, five warm-up calls, and 20 measured repetitions. Application-sized batches use $B = 112$, with selected $B = 1024$ throughput measurements. The five-family rerun includes affine equality, ordering, simplex, the nonlinear circle, and the real-world nonnegative-ordering geometry. The real-world affine/nonlinear numbers reported elsewhere are direct frozen-run implementation timings rather than additions to this controlled profiler.

H Additional Experimental Results

For FDC comparisons, seed-based intervals and tests describe run variability conditional on the fixed dataset and temporal split. They do not quantify sampling uncertainty across new years, basins, or hydrological datasets.

H.1 Float64 affine diagnostic

Method	CEabs FP32	CEabs FP64	CEmax FP32	CEmax FP64	VR FP32	VR FP64
Projection-Or	4.22×10^{-7}	6.89×10^{-16}	4.29×10^{-6}	4.44×10^{-15}	0.0000	0.0000
Projection-Ob	5.68×10^{-7}	1.28×10^{-15}	2.29×10^{-5}	1.89×10^{-13}	0.0017	0.0000
`SLP-ProbHard`	2.69×10^{-7}	4.97×10^{-16}	2.86×10^{-6}	5.33×10^{-15}	0.0000	0.0000

Table 11: Seed-20 diagnostic. Trained models are unchanged; only feasibility-map / residual computation is repeated in float64.

The mean residuals collapse from approximately 10^{-7} in float32 to 10^{-15} in float64 for all three hard-constrained affine methods. Projection-Ob's seed-20 float32 violation rate of 0.0017 also disappears in float64. This supports the interpretation that the small base-case residuals are numerical realization effects rather than a relaxation of the mathematical constraints. The diagnostic is not sufficient by itself to attribute every high-dimensional scaling exceedance to the same source, so the manuscript keeps continuous residuals primary.

H.2 Linear scaling

Across the twelve pre-specified (d, m, q) configurations, `SLP-ProbHard` always uses $q = d - m$ stochastic coordinates, giving reductions from 12.3% to 38.5%. Point accuracy remains close. CRPS and Energy Score usually favor Projection-Or slightly, but the gap generally narrows at higher codimension. At larger configurations, float32 10^{-5} threshold exceedances appear for both affine methods; `SLP-ProbHard` is consistently lower in the largest cases. For $(d, m, q) = (97, 33, 64)$, Projection-Or has $\mathrm{VR}_{10^{-5}} = 0.3958$ versus 0.0193 for SLP, while mean absolute residuals remain at 1.99×10^{-6} and 1.23×10^{-6}, respectively. The complete table is in Appendix J.

H.3 Ordering scaling

From $d = 3$ to $d = 40$, positive-part `SLP-ProbHard` has lower CRPS and Energy Score than end-to-end isotonic projection in every evaluated dimension. The CRPS gain grows from 0.55% at $d = 3$ to 25.96% at $d = 40$. More strikingly, isotonic projection's tolerance-based tie rate rises from 0.300 at $d = 3$ to 0.891 at $d = 40$, while the observed rate stays near 0.25. Positive-part `SLP-ProbHard` remains close to that target from $d = 5$ onward; averaged across the five scaling dimensions, its absolute tie-rate error is 0.030 versus 0.283 for end-to-end isotonic projection. Appendix J gives the complete scaling table.

H.4 Synthetic composed nonnegative ordering: mechanism validation

A controlled $d = 5$ benchmark generated targets by cumulative positive-part, so both non-negativity and ordering are structurally active. Because the target generator shares the cumulative-positive-part mechanism with the `SLP-ProbHard` map, this is a deliberately favorable controlled test of the boundary-preimage prediction rather than an independent test of generalization across unknown data-generating mechanisms. Across frozen seeds 115-119, both nonnegative isotonic projection and nonnegative cumulative positive-part `SLP-ProbHard` achieve zero feasibility violations. Predictive scores are close, with `SLP-ProbHard` slightly lower CRPS and Energy Score.

Method	MSE	MAE	CRPS	ES	$P(y_1 = 0)$	Pair-tie rate
Projection-nonneg-isotonic	0.48397 ± 0.00507	0.49376 ± 0.00177	0.35222 ± 0.00122	0.93289 ± 0.00328	0.5803	0.5440
SLP-nonnegative-positive-part	0.48433 ± 0.00751	0.49437 ± 0.00413	0.35027 ± 0.00315	0.92047 ± 0.00740	0.5802	0.4374
Target	–	–	–	–	0.5850	0.4798

Table 12: Synthetic composed constraint benchmark, seeds 115-119. Both hard methods have zero negative values and zero ordering violations. `SLP-ProbHard` is closer on adjacent-tie frequency, whereas projection is closer on the anchor-zero frequency.

The paired Energy Score difference favors `SLP-ProbHard` in all five seeds (paired t $p = 0.019$; Wilcoxon $p = 0.0625$), while the paired test did not detect a CRPS difference at this five-seed level ($p = 0.163$). With only five seeds, we report effect sizes, confidence intervals, and paired tests rather than an observed post-hoc power calculation. The absolute pair-tie calibration error is 0.0423 for `SLP-ProbHard` versus 0.0643 for projection and favors `SLP-ProbHard` in all five seeds ($p = 0.00368$). Conversely, projection has the smaller anchor-zero error. This controlled benchmark therefore supports the boundary-preimage mechanism without implying universal boundary-calibration superiority.

H.5 NONLINEAR CIRCLE PRIMARY RESULTS

Method	MSE	MAE	CRPS	ES	Cov.90	Width.90	CEabs	$\mathrm{VR}_{10^{-5}}$
Unconstrained	0.3622	0.4857	0.3422	0.5391	0.835	1.593	0.623	1.0000
Soft penalty	0.3613	0.4850	0.3438	0.5417	0.810	1.484	0.584	1.0000
Post-hoc radial	0.3649	0.4827	0.3344	0.5253	0.807	1.496	4.80×10^{-8}	0
End-to-end radial	0.3557	0.4893	0.3277	0.5147	0.864	1.704	4.81×10^{-8}	0
SLP-angle	0.3557	0.4836	0.3272	0.5143	0.854	1.611	1.70×10^{-8}	0

Table 13: Unit-circle primary results, seeds 130-139. Residual differences between the two hard maps are numerical roundoff; both have zero 10^{-5} violations. `SLP-ProbHard` uses one stochastic coordinate versus two for radial projection.

For end-to-end radial projection versus SLP, paired tests detected no differences in MSE ($p = 0.988$), CRPS ($p = 0.750$), Energy Score ($p = 0.888$), or circular MAE ($p = 0.633$) at the ten-seed level. `SLP-ProbHard` MAE is 1.17% lower ($p = 0.0279$). The correct primary conclusion is therefore practical hard feasibility with a 50% stochastic-coordinate reduction and no detected differences in those principal scores, not universal score superiority.

H.6 REAL-WORLD AFFINE FDC RESULTS

Method	Raw MSE	Raw MAE	Raw CRPS	Raw ES	Cov.90	Width.90	CEabs	$\mathrm{VR}_{10^{-5}}$
Unconstrained	104.47	4.738	3.689	13.882	0.766	13.86	0.1270	1.000
Soft affine	104.37	4.731	3.686	13.882	0.763	13.62	0.1250	1.000
Post-hoc affine projection	103.72	4.685	3.717	14.099	0.681	9.90	3.51×10^{-8}	0
End-to-end affine projection	106.40	4.707	3.627	13.649	0.746	13.57	5.85×10^{-8}	0
SLP-affine	105.14	4.649	3.698	13.997	0.712	10.94	0	0

Table 14: Real-world affine FDC coherence, seeds 150-159. Raw-Q metrics are primary; the affine branch enforces only $Q_i - Q_{i+1} + \Delta_i = 0$, not nonnegativity/order.

Metric	Projection	`SLP-ProbHard`	Mean diff.	95% CI	paired t p	Wilcoxon p
Raw MSE	106.398	105.136	-1.262	[-10.616, 8.092]	0.767	0.770
Raw MAE	4.707	4.649	-0.058	[-0.194, 0.078]	0.361	0.432
Raw CRPS	3.627	3.698	+0.072	[-0.032, 0.176]	0.153	0.193
Raw ES	13.649	13.997	+0.348	[-0.051, 0.747]	0.080	0.105
Coverage.90	0.746	0.712	-0.034	[-0.061, -0.0068]	0.0197	0.0195
Width.90	13.567	10.937	-2.630	[-3.075, -2.186]	3.00×10^{-7}	0.00195

Table 15: Paired real-world affine comparison: `SLP-ProbHard` minus end-to-end affine projection. Paired tests did not detect differences in the principal accuracy/proper-score metrics at ten seeds; interval width and coverage show a significant sharpness-calibration trade-off.

The affine hard methods both have zero 10^{-5} violations. `SLP-ProbHard` uses 7 rather than 13 stochastic coordinates (46.15% reduction) and 37,774 versus 39,322 parameters. In the frozen implementation, median inference latency is 1.148 ms for `SLP-ProbHard` and 1.623 ms for end-to-end projection; this approximately 29% reduction is implementation-specific. The soft penalty selected at validation weight 0.01 retains $\mathrm{VR}_{10^{-5}} = 1$, illustrating that the penalty baseline does not hard-enforce affine coherence.

H.7 REAL-WORLD NONLINEAR SCALE-SHAPE RESULTS

Method	Raw MSE	Raw MAE	Raw CRPS	Raw ES	Cov.90	Width.90	Bilinear CE	Hard VR
Unconstrained	103.39	4.892	3.850	14.292	0.804	14.19	0.09695	1.000
Soft nonlinear	103.93	4.971	3.921	14.488	0.805	14.62	0.09597	1.000
Post-hoc feasible correction	102.33	4.840	3.568	13.754	0.823	14.97	0	0
End-to-end feasible correction	104.92	5.151	4.177	14.483	0.630	15.75	0	0
SLP-scale-shape	106.17	5.175	4.047	14.313	0.649	15.67	0	0

Table 16: Real-world nonlinear scale-shape FDC coherence, seeds 170-179. All three hard methods also have zero negative reconstructed discharge and zero ordering violations.

Metric	Correction	`SLP-ProbHard`	Mean diff.	95% CI	paired t p	Wilcoxon p
Raw MSE	104.916	106.166	+1.250	[-3.778, 6.279]	0.588	0.432
Raw MAE	5.151	5.175	+0.024	[-0.175, 0.223]	0.790	0.432
Raw CRPS	4.177	4.047	-0.130	[-0.285, 0.025]	0.090	0.232
Raw ES	14.483	14.313	-0.171	[-0.609, 0.268]	0.402	0.557
Coverage.90	0.630	0.649	+0.020	[-0.066, 0.105]	0.619	0.375
Width.90	15.747	15.669	-0.078	[-1.909, 1.753]	0.925	1.000

Table 17: Paired real-world nonlinear comparison: `SLP-ProbHard` minus the five-iteration end-to-end alternating feasible correction. Paired tests did not detect differences in the principal predictive metrics at ten seeds.

The post-hoc comparator is also feasible and has lower reported mean raw-Q CRPS than both end-to-end alternatives. Table 17 compares SLP with the end-to-end correction model only; it does not provide a paired significance test against the post-hoc model. These results do not identify the cause of the performance difference.

Both primary end-to-end hard methods have zero bilinear violations at 10^{-5}, simplex residuals near float32 roundoff, zero negative reconstructed discharge, and zero FDC ordering violations. `SLP-ProbHard` uses 8 computational stochastic coordinates versus 14 ambient coordinates for the correction model (42.86% reduction), with one redundant computational coordinate relative to the 7 intrinsic DOF. Parameter counts are 38,032 versus 39,580. Frozen median inference latency is 1.478 ms for `SLP-ProbHard` versus 7.369 ms for the five-iteration end-to-end correction; the roughly 80% reduction is specific to these implementations and iteration count. The validation-selected soft nonlinear baseline has bilinear CE about 0.096 and hard violation rate 1.0.

H.8 COMPLETE PAIRED STATISTICS

Family	Metric	Reference mean	`SLP-ProbHard` mean	Rel. `SLP-ProbHard` diff.	95% CI of paired diff.	paired t p	Wilcoxon p
Linear equality	mse	0.29646	0.29574	-0.25%	[-0.003834, 0.00238]	0.609	0.77
Linear equality	mae	0.39286	0.39295	+0.02%	[-0.001766, 0.001944]	0.916	0.77
Linear equality	crps	0.28467	0.29012	+1.91%	[0.003935, 0.00696]	1.91e-05	0.00195
Linear equality	energy score	1.20878	1.22158	+1.06%	[0.005968, 0.01964]	0.00218	0.00391
Weak ordering	mse	5.56771	5.09482	-8.49%	[-1.049, 0.1034]	0.0964	0.084
Weak ordering	mae	1.09838	1.08457	-1.26%	[-0.02347, -0.004143]	0.0103	0.0195
Weak ordering	crps	0.79450	0.78204	-1.57%	[-0.01804, -0.006897]	0.000678	0.00195
Weak ordering	energy score	2.24397	2.18828	-2.48%	[-0.07136, -0.04002]	2.13e-05	0.00195
Simplex	mse	0.03222	0.03196	-0.84%	[-0.0004537, -7.085e-05]	0.0127	0.00977
Simplex	mae	0.11799	0.11833	+0.28%	[-0.0002802, 0.000968]	0.244	0.232
Simplex	crps	0.08899	0.08662	-2.67%	[-0.002826, -0.00191]	9.62e-07	0.00195
Simplex	energy score	0.24740	0.24306	-1.76%	[-0.005441, -0.003236]	9.33e-06	0.00195

Table 18: Paired untouched-seed comparisons. Relative difference is $100(\text{SLP} - \text{reference})/\text{reference}$; negative values favor `SLP-ProbHard` for error/score metrics. p-values are unadjusted and are reported as diagnostics rather than as a multiple-comparison discovery exercise.

H.9 LATENT COVARIANCE EXPRESSIVENESS

Table 19 tests whether the diagonal latent Gaussian is responsible for the linear `SLP-ProbHard` probabilistic-score gap. All variants remain feasible at the 10^{-5} threshold. Moving from diagonal `SLP-ProbHard` to low-rank or full covariance leaves MSE/MAE essentially unchanged but improves CRPS from 0.28887 to about 0.2837 and Energy Score from 1.21714 to about 1.20710. The low-rank and full variants are numerically very close in predictive score on this small $q = 8$ benchmark. Because the rank-4 implementation has slightly more parameters than the full Cholesky head at this dimension, no parameter-efficiency claim is made for low-rank covariance here.

Method	MSE	CRPS	ES	Cov.90	Parameters	$\mathrm{VR}_{10^{-5}}$
Projection-Or	0.29741	0.28544	1.21124	0.7825	37,526	0
`SLP-ProbHard` diagonal	0.29543	0.28887	1.21714	0.7954	36,752	0
`SLP-ProbHard` low-rank	0.29591	0.28364	1.20710	0.8016	40,880	0
`SLP-ProbHard` full	0.29587	0.28376	1.20710	0.7969	40,364	0

Table 19: Covariance expressiveness ablation, seeds 40-44. The experiment compares covariance families within `SLP-ProbHard` against the standard Projection-Or reference; it is not a matched full-covariance projection study.

I CONTROLLED COMPUTATIONAL PROFILING

I.1 CONTROLLED COMPUTATIONAL PROFILING

Table 20 reports the circle-inclusive controlled rerun for the application-sized $B = 112$, $S = 100$ end-to-end workload. Affine equality shows a 1.30× `SLP-ProbHard` advantage. Simplex savings rise from 1.06× at $d = 5$ to 2.12× at $d = 40$. On the nonlinear circle, `SLP-ProbHard` is 1.22× faster end to end while sampling one structural coordinate rather than two ambient coordinates. Ordering remains dramatically faster with cumulative positive-part than the project's reference PAVA/isotonic implementation; the real-world nonnegative-ordering pair shows the same implementation pattern. These ratios are evidence about this controlled implementation and workload, not universal solver speedups.

Family	d	Reference ms	`SLP-ProbHard` ms	Ref./ `SLP-ProbHard`	Latency reduction
Linear equality	11	1.784	1.367	1.30×	23.4%
Ordering	5	248.802	0.958	259.7×	99.61%
Ordering	40	504.658	3.872	130.3×	99.23%
Simplex	5	1.471	1.382	1.06×	6.0%
Simplex	40	7.492	3.540	2.12×	52.8%
Circle	2	0.616	0.503	1.22×	18.3%
Real-world nonneg. order	7	272.462	0.949	287.1×	99.65%

Table 20: Controlled end-to-end CPU profiling: matched MLP + diagonal Gaussian sampling + feasibility operation, batch 112 and 100 predictive samples, 8 threads, float32, 5 warm-ups and 20 measured repeats. Ordering/PAVA ratios are implementation-level.

For the circle, isolated operator latency is 0.149 ms for radial normalization and 0.100 ms for $(\cos z, \sin z)$ at $B = 112$, a 33.0% reduction. At $B = 1024$, end-to-end circle latency is 3.111 ms versus 2.359 ms (1.32×). Both circle operators satisfy the equality to numerical precision. For affine equality, the $B = 1024$ end-to-end pipeline is 13.564 ms versus 10.153 ms (1.34×), illustrating that lower stochastic dimension can matter even when both feasibility operators are simple tensor operations.

J Complete Scaling Tables

J.1 Linear equality scaling

In Table 21, Δ is $100(\mathrm{SLP} - \mathrm{Projection})/\mathrm{Projection}$; negative predictive-score values favor SLP. The scaling tables report predictive, feasibility, boundary, and representation-dimensional behavior only. Runtime comparisons are reported separately under the controlled profiling protocol in Appendix I.

d	m	q	Dim.red.	ΔMSE	ΔCRPS	ΔES	VR Proj.	VR `SLP-ProbHard`	CEabs Proj.	CEabs `SLP-ProbHard`
10	2	8	20.0%	+0.46%	+3.56%	+2.00%	0.0000	0.0000	2.81×10^{-7}	3.55×10^{-7}
11	3	8	27.3%	+0.71%	+2.06%	+1.22%	0.0000	0.0000	4.13×10^{-7}	2.63×10^{-7}
13	5	8	38.5%	-0.56%	+0.58%	+0.36%	0.0000	0.0000	3.78×10^{-7}	2.19×10^{-7}
19	3	16	15.8%	+0.03%	+2.45%	+1.11%	0.0000	0.0000	5.53×10^{-7}	6.11×10^{-7}
21	5	16	23.8%	-0.57%	+1.08%	+0.38%	0.0000	0.0000	4.19×10^{-7}	4.00×10^{-7}
25	9	16	36.0%	-0.19%	+0.46%	+0.36%	0.0000	0.0000	4.78×10^{-7}	3.91×10^{-7}
37	5	32	13.5%	+0.95%	+1.51%	+1.03%	0.0000	0.0000	9.42×10^{-7}	7.75×10^{-7}
41	9	32	22.0%	+0.02%	+0.50%	+0.43%	0.0021	0.0000	9.67×10^{-7}	7.95×10^{-7}
49	17	32	34.7%	-0.37%	+0.16%	+0.27%	0.0007	0.0000	6.73×10^{-7}	5.34×10^{-7}
73	9	64	12.3%	-0.38%	+0.60%	+0.10%	0.0577	0.0015	1.71×10^{-6}	1.60×10^{-6}
81	17	64	21.0%	-0.16%	+0.04%	+0.16%	0.1672	0.0000	1.47×10^{-6}	9.36×10^{-7}
97	33	64	34.0%	-0.10%	+0.06%	+0.05%	0.3958	0.0193	1.99×10^{-6}	1.23×10^{-6}

Table 21: Twelve held-out linear scaling configurations, seeds 30-34.

J.2 Ordering scaling

Positive gain values in Table 22 mean lower/better `SLP-ProbHard` error or score. Boundary-error columns are absolute differences from the observed tolerance-based tie rate.

d	MSE gain	CRPS gain	ES gain	Obs.tie	`SLP-ProbHard` tie	Proj.tie	`SLP-ProbHard` tie err.	Proj tie err.
3	+5.25%	+0.55%	+1.14%	0.261	0.181	0.300	0.079	0.040
5	+17.78%	+2.35%	+3.05%	0.251	0.235	0.401	0.017	0.149
10	+11.65%	+3.91%	+4.79%	0.255	0.234	0.504	0.021	0.249
20	+10.41%	+3.85%	+4.48%	0.254	0.232	0.591	0.022	0.337
40	+17.75%	+25.96%	+23.50%	0.253	0.241	0.891	0.012	0.638

Table 22: Ordering scaling, seeds 70-74.

J.3 Simplex scaling

d	MSE gain	CRPS gain	ES gain	Obs.zero	`SLP-ProbHard` zero	Proj.zero	`SLP-ProbHard` zero err.	Proj zero err.
3	+0.85%	+2.80%	+2.50%	0.245	0.164	0.195	0.081	0.050
5	+1.51%	+2.80%	+1.87%	0.250	0.164	0.233	0.086	0.017
10	+0.82%	+3.38%	+1.54%	0.250	0.253	0.325	0.003	0.075
20	+2.53%	+4.47%	+2.16%	0.252	0.338	0.404	0.086	0.152
40	+1.99%	+4.76%	+1.82%	0.252	0.372	0.453	0.120	0.202

Table 23: Simplex scaling, seeds 110-114. Positive gains mean lower `SLP-ProbHard` score.

K Extended Discussion and Limitations

K.1 The feasible representation is part of the probabilistic model

The central empirical conclusion is that an ***SFLP*** inside `SLP-ProbHard` is not merely a constraint-enforcement wrapper. Because $P^Y(B) = P^Z(h^{-1}(B))$, changing h changes support geometry and probability allocation. The weak-order and nonnegative-order maps illustrate this distinction: weak-order cumulative positive-part supports adjacent ties but not a positive anchor-zero atom under a continuous anchor, whereas nonnegative cumulative positive-part gives positive-probability preimages to both. Both constructions are feasible for the constraints they encode, yet they define different predictive-law families.

Hydrological interpretation. The hydrological experiments are deliberately constructed from one frozen dataset. Nonnegative ordering asks whether every FDC sample lies in the physical order cone; affine coherence asks whether explicitly represented increments agree with adjacent FDC differences; nonlinear scale-shape asks whether increment magnitude and compositional shape satisfy $\Delta = RP$. The underlying data and split are held fixed, with the representation-specific target transformations described in Section 4. This makes the new real-world evidence a test of the central claim that feasible representation is part of probabilistic model design.

K.2 Dimensionality is family- and map-dependent

The affine equality case provides the clean intrinsic-coordinate result: $q = d - m$ is both the mathematical dimension of the feasible affine set and the `SLP-ProbHard` stochastic dimension. The primary synthetic experiment reduces 11 stochastic coordinates to 8, the circle reduces two ambient coordinates to one structural angle, and the real-world affine FDC representation reduces 13 ambient stochastic coordinates to 7. This experiment is not expected to establish universal score superiority: under diagonal covariance Projection-Or scores better, while richer `SLP-ProbHard` covariance closes or reverses that gap without changing feasibility or the q-coordinate representation. By contrast, the current weak-order map uses d latents for a d-dimensional cone, normalized positive-part boundary-mass simplex uses d computational latents for a $d-1$ dimensional simplex, and the nonlinear real-world scale-shape map uses 8 computational latents for a 7-DOF structural object. The safe general claim is that `SLP-ProbHard` *can exploit intrinsic coordinates when the map permits*, not that every `SLP-ProbHard` map is lower-dimensional.

K.3 Boundary capability is not boundary calibration

Boundary capability does not imply boundary calibration. In the primary $d = 5$ simplex experiment, projection's zero rate is closer to the observed rate, whereas SLP is closer on the reported primary ordering tie rate. The map image determines reachability, while the latent probabilities of event preimages determine boundary mass.

K.4 Covariance expressiveness and structural feasibility are separate design axes

The covariance ablation shows that richer latent dependence can improve probabilistic scores without altering the structural feasibility guarantee. This supports a modular interpretation of `SLP-ProbHard`: the structural map specifies feasible images and coordinates, while the latent law allocates probability through its preimages; together they constitute the ***SFLP***. The diagonal Gaussian is therefore a modeling choice rather than a defining limitation of `SLP-ProbHard`. A matched full-covariance projection comparison remains outside this particular ablation.

K.5 Computation depends on the constraint mechanism and implementation

Controlled profiling replaces the earlier training-log timing observations. The affine case shows that structural generation need not always make an isolated operator faster, although lower stochastic dimension can still reduce end-to-end workload. Simplex advantages grow with dimension but are moderated by shared neural inference. Ordering shows very large gains under the project's PAVA implementation because cumulative positive-part is a direct tensor operation while isotonic projection processes constrained rows. The correct claim is therefore family- and implementation-dependent computational advantage, not a universal complexity theorem for `SLP-ProbHard` versus projection.

Architectural and timing claims. The primary computational statement is architectural: $11 \to 8$ stochastic coordinates is representation-level and hardware independent. Wall-clock timing is secondary evidence. Under the present matched CPU implementation `SLP-ProbHard` has lower latency, but optimized affine conditioning is itself inexpensive and other families depend strongly on implementation. Accordingly, all runtime statements are qualified by the exact workload and reference code.

K.6 Relationship to ProbHardE2E/DPPL

The completed affine experiment confirms a complementary rather than uniformly ordered relationship. ProbHardE2E/DPPL conditions an ambient Gaussian, while affine `SLP-ProbHard` learns the $q = d - m$ structural Gaussian whose pushforward is feasible by construction. On the matched ten-seed benchmark, both are hard feasible at 10^{-5}; `SLP-ProbHard` has lower MSE/MAE and a 27.27% smaller stochastic representation, whereas DPPL has lower marginal CRPS and better 90% coverage, and the paired test did not detect an Energy Score difference. This comparison is direct but deliberately scoped: it uses the frozen official source through a verified affine adapter and does not reproduce the authors' PDE application or establish dominance on other constraint classes.

L Limitations

1. **Explicit map requirement.** Some feasible sets are easier to project onto than to parameterize, and projection is preferable in those cases.
2. **Map-dependent coverage and calibration.** A non-surjective map introduces parameterization bias, while even a boundary-capable map does not guarantee correct boundary frequencies. The present experiments do not resolve the source of the observed undercoverage or establish nominal calibration.
3. **Map-dependent dimensionality.** Only some structural maps achieve intrinsic computational coordinates; normalized positive-part simplex is intentionally redundant by one coordinate.
4. **Latent distribution remains a modeling choice.** Richer covariance improves `SLP-ProbHard` probabilistic scores in the linear ablation, but the present study does not exhaust non-Gaussian or flow-based latent families.
5. **Nonsmooth and discontinuous maps introduce optimization effects.** The positive-part map is continuous but nonsmooth and has a flat negative region. The normalized positive-part map additionally uses a discontinuous argmax-vertex fallback on the all-inactive region. The latter requires a separate analysis of expected-score gradient estimation.
6. **Exact feasibility and analytic tractability differ.** Nonlinear pushforward densities or moments may be unavailable even when every sample is feasible.
7. **Nonlinear-equality coverage is explicit rather than general.** The unit circle and scale-shape FDC supply clean nonlinear equalities, but no arbitrary nonlinear-equality parameterization or higher-dimensional manifold scaling is claimed.
8. **Controlled timing is implementation-specific.** In particular, the very large ordering gaps reflect the project's CPU PAVA/isotonic implementation, and the nonlinear FDC comparison uses a five-iteration alternating correction; these should not be read as universal speedups over every optimized implementation.
9. **The hydrological target is model-based reanalysis discharge.** The seven-basin benchmark uses GloFAS v5.0 LISFLOOD-based reanalysis discharge rather than observed-gauge discharge (Grimaldi et al., 2026); it validates real-world constraint structure and distributional behavior rather than observational hydrological skill.
10. **The direct ProbHardE2E comparison is affine-adapter scoped.** The official repository and released Gaussian conditioning equations are frozen and verified, but the released PDE wrapper does not expose arbitrary A, b. We therefore substitute the benchmark hierarchy matrix while retaining the released equations; this is not a reproduction of the original PDE datasets or published runtime.
11. **Finite precision remains visible.** Affine real-arithmetic guarantees can leave small float32 residuals; the primary float64 diagnostic supports this interpretation for the base case but does not establish the cause of every scaling exceedance.

M Reproducibility and Reporting Conventions

All frozen training-result directories were generated on the HPC CPU partition using the validated Python 3.11 / PyTorch CPU environment. The environment check used the exact Conda interpreter,

disabled user-site packages, and forced the environment's C++ runtime to avoid system-library contamination. Final jobs used eight allocated CPU threads. Result bundles preserve per-seed CSVs, configuration JSON files, histories, aggregate summaries, paper tables, checkpoints where enabled, and the float64 diagnostic.

Continuous feasibility residuals remain primary. Thresholded violation rates at 10^{-5} are finite-precision diagnostics. Ordering tie and simplex zero rates are distributional boundary diagnostics and are not conflated with violation metrics. The real-world affine and nonlinear extensions reuse the same frozen source artifact and preserve its SHA-256 in their run tables; raw-Q metrics are primary for hydrological interpretation. The official-source DPPL affine-adapter comparison preserves the upstream Git SHA and source hashes, verifies the generic adapter against a direct transcription of the released equations, reuses the frozen `SLP-ProbHard` rows without retraining, and profiles only after the predictive primary is frozen. Paired p-values are unadjusted and interpreted alongside effect sizes.

N Scope of Empirical Claims

The evidence supports `SLP-ProbHard` for constraint classes admitting suitable explicit *SFLP*s: direct feasible generation, exact affine free coordinates, representation-dependent boundary probability, one-coordinate circular generation, real-world affine and nonlinear coherence, composed nonnegative/ordered discharge generation, covariance-head modularity, and family-dependent computational differences under the stated implementations. The official-source ProbHardE2E evidence is a direct ten-seed comparison on one compatible affine benchmark; it supports the reported representation trade-off there but does not establish superiority across ProbHardE2E's nonlinear or application-specific constraint classes. Generality remains map dependent, and arbitrary nonlinear equalities still require suitable explicit maps or other enforcement mechanisms.

O Official ProbHardE2E/DPPL Affine Comparison Audit

O.1 Frozen source provenance and adapter scope

The direct comparator uses the official `amazon-science/probharde2e` repository (Utkarsh et al., 2026; Amazon Science, 2026) frozen at Git commit

`ff787718e79e5ac6b673807d40ce1cddb0fd17c2`

(commit date 2026-03-07, branch `main`). The worktree was clean. SHA-256 hashes were recorded for `probconserv.py`, `ProbHardE2E_Linear.ipynb`, `README.md`, and `LICENSE`; the run also checked for the released hard-linear equation fragments and the linear-notebook calls to the probabilistic conservation layer.

The released wrapper (Utkarsh et al., 2026; Amazon Science, 2026) constructs its PDE-specific conservation matrix g internally and does not expose the arbitrary affine interface needed for the hierarchy benchmark. The adapter therefore makes one problem-interface substitution,

$$g_{\mathrm{PDE}} \longrightarrow A_{\mathrm{benchmark}},$$

while retaining the released Gaussian conditioning equations. It is consequently described throughout as *ProbHardE2E/DPPL (official-source affine adapter)*, not as an untouched call to the original PDE wrapper.

For archival reproducibility, the affine-adapter source distributed in the frozen comparison patch is included with the v24 source bundle as `frozen_dppl_evidence/dppl_official_adapter.py`. Its SHA-256 is

`442cdbbe3dbd58ef0a8abbe1824baf09108a72e0ab258dfffc8438feb6f10087`,

matching the patch manifest. A machine-readable provenance record accompanies the source. This archive documents the adapter implementation used for the comparison workflow; it does not alter or rerun the frozen predictive experiment.

O.2 AFFINE DPPL EQUATIONS AND COHERENT SAMPLING

For an ambient diagonal Gaussian $\widetilde{Y} \mid x \sim \mathcal{N}(\mu, \Sigma)$ with affine equality $AY = b$, the verified adapter retains the released ProbHardE2E/DPPL hard-linear conditioning equations (Utkarsh et al., 2026; Amazon Science, 2026):

$$S = A\Sigma A^\top, \qquad K = \Sigma A^\top S^{-1}, \tag{75}$$

$$\mu_c = \mu - K(A\mu - b), \tag{76}$$

$$\Sigma_c = \Sigma - KA\Sigma. \tag{77}$$

Because the common objective and evaluation use sample-based CRPS and Energy Score, independent sampling from only the projected marginal variances would discard the induced cross-coordinate covariance. Samples are therefore generated by the equivalent reparameterization

$$Y = \widetilde{Y} - K(A\widetilde{Y} - b), \qquad \widetilde{Y} \sim \mathcal{N}(\mu, \Sigma), \tag{78}$$

which preserves the singular conditioned joint Gaussian in exact arithmetic. The constraint solve is performed in float64 and returned to the model dtype, matching the released source's double-precision constraint path.

O.3 EQUATION-LEVEL EQUIVALENCE AND FINITE-PRECISION AUDIT

Before primary evaluation, the generic affine adapter was compared with a direct transcription of the released hard-linear equations using the same benchmark A, b. Table 24 reports the frozen check.

Audit quantity	Result
Adapter vs. direct-reference mean, max abs. diff.	0
Adapter vs. direct-reference covariance, max abs. diff.	0
Conditioned-mean constraint max abs. residual	1.19×10^{-7}
$A\Sigma_c$ max abs. residual	4.84×10^{-8}
Float64 sample constraint max abs. residual	2.66×10^{-15}
Float32 sample $\mathrm{CE}_{\mathrm{abs}}$	4.60×10^{-8}
Float32 sample $\mathrm{CE}_{\max}$	5.36×10^{-7}
Float32 $\mathrm{VR}_{10^{-5}}$	0

Table 24: Frozen equivalence and feasibility audit for the official-source affine adapter.

O.4 PRE-SPECIFIED PROTOCOL AND EXECUTION AUDIT

Before inspecting official-source DPPL affine-adapter test results, the direct-comparator protocol froze the existing affine task, seed registry 20-29, common metrics, float32 evaluation with 100 predictive samples, and the rule that existing `SLP-ProbHard` results would not be retrained or retuned after comparator inspection. The executed benchmark uses 12 inputs, 11 outputs, rank-3 affine constraints, and intrinsic dimension 8, with $3000/750/750$ train/validation/test samples; both methods use the same width-128, depth-3 backbone, batch size 128, 60-epoch cap, learning rate 10^{-3}, weight decay 10^{-5}, gradient clip 5, 12/32 train/validation Monte Carlo samples, and 0.5 CRPS + 0.5 Energy Score objective.

The execution audit confirms all ten primary seeds, DPPL stochastic dimension 11 versus `SLP-ProbHard` dimension 8, and no `SLP-ProbHard` retraining. The frozen `SLP-ProbHard` result file retained the same SHA-256 before and after the DPPL job. Seed-specific train/validation/test tensor hashes were stored for every DPPL seed, and profiling was run only after the predictive primary had been frozen.

O.5 FULL DIRECT-COMPARISON RESULTS

Table 25 gives the complete primary means and standard deviations. Both methods have zero 10^{-5} violations. DPPL's smaller float32 residuals and SLP's larger ones are both far below the diagnostic threshold and are not interpreted as a substantive feasibility difference.

Metric	DPPL source adapter	SLP-ProbHard
MSE	0.29968 ± 0.00797	$\mathbf{0.29574 \pm 0.00614}$
MAE	0.39466 ± 0.00479	$\mathbf{0.39295 \pm 0.00360}$
Marginal CRPS	$\mathbf{0.28676 \pm 0.00367}$	0.29012 ± 0.00277
Energy Score	1.21896 ± 0.01550	1.22158 ± 0.01228
Coverage 90	$\mathbf{0.8093 \pm 0.0096}$	0.7856 ± 0.0078
Interval width 90	1.2967 ± 0.0358	$\mathbf{1.2637 \pm 0.0232}$
CE_{abs}	6.55×10^{-8}	2.60×10^{-7}
CE_{max}	9.54×10^{-7}	2.59×10^{-6}
$VR_{10^{-5}}$	0	0
Stochastic coordinates	11	**8**
Trainable parameters	37,526	**36,752**
Primary median inference latency (ms)	3.222 ± 0.136	$\mathbf{1.647 \pm 0.191}$

Table 25: Official-source ProbHardE2E/DPPL affine-adapter comparison versus the already-frozen SLP-ProbHard primary, seeds 20-29. Bold marks lower error/score, coverage closer to the nominal 0.90 level, smaller coordinate/parameter counts, and lower measured latency. Interval width is a sharpness diagnostic and must be interpreted jointly with coverage; latency is implementation-specific.

O.6 Paired statistical interpretation

Table 26 reports SLP-ProbHard minus DPPL paired differences. MSE and MAE favor SLP; marginal CRPS and coverage favor DPPL; the paired test did not detect an Energy Score difference. The interval result is a sharpness-coverage trade-off rather than an unconditional improvement.

Metric	DPPL mean	SLP-ProbHard mean	SLP−DPPL	95% CI	paired t p	Wilcoxon p
MSE	0.29968	0.29574	-0.00395	$[-0.00653, -0.00137]$	0.0072	0.0098
MAE	0.39466	0.39295	-0.00171	$[-0.00328, -0.00015]$	0.0353	0.0645
CRPS	0.28676	0.29012	$+0.00336$	$[0.00198, 0.00474]$	3.76×10^{-4}	0.0020
Energy Score	1.21896	1.22158	$+0.00263$	$[-0.00397, 0.00922]$	0.3913	0.5566
Coverage 90	0.8093	0.7856	-0.02367	$[-0.03246, -0.01488]$	1.81×10^{-4}	0.0039
Width 90	1.2967	1.2637	-0.03302	$[-0.06293, -0.00310]$	0.0341	0.0371
CE_{abs}	6.55×10^{-8}	2.60×10^{-7}	$+1.94 \times 10^{-7}$	$[1.86, 2.03] \times 10^{-7}$	1.93×10^{-12}	0.0020
CE_{max}	9.54×10^{-7}	2.59×10^{-6}	$+1.63 \times 10^{-6}$	$[1.42, 1.85] \times 10^{-6}$	3.16×10^{-8}	0.0020

Table 26: Paired direct-comparator statistics across the same ten seeds. p-values are descriptive and unadjusted. Residual differences are numerical-scale effects: both methods have $VR_{10^{-5}} = 0$.

O.7 Stochastic dimension, covariance family, and computation

The direct comparison makes the representation difference explicit. DPPL starts from an 11-coordinate ambient diagonal Gaussian and conditions it onto the rank-3 affine set. SLP-ProbHard instead parameterizes the 8 free coordinates directly. Hence the structural representation removes

$$100\left(1 - \frac{8}{11}\right) = 27.27\%$$

of stochastic coordinates before feasibility is considered. These two models need not induce the same covariance family on the feasible subspace: DPPL's covariance weighting can exploit ambient marginal variances, whereas diagonal affine SLP-ProbHard induces covariance through $N\Sigma_Z N^\top$. This distinction is consistent with the observed DPPL marginal-CRPS advantage and the separate finding that richer SLP-ProbHard covariance heads improve CRPS/ES without changing feasibility.

The primary run's matched instrumentation gives median inference latency 3.222 ± 0.136 ms for DPPL versus 1.647 ± 0.191 ms for SLP. A separate DPPL-only frozen profile, performed after the predictive primary and without retraining, uses batch 128, 100 samples, eight CPU threads, 20

warmups, and 200 repeats; its per-seed median averages 3.115 ± 0.116 ms. These values describe the present adapter/`SLP-ProbHard` implementations and hardware only; they are not reproductions of the runtime reported by the original ProbHardE2E paper.

O.8 What the direct comparison does and does not establish

The result closes the compatible affine-comparator gap: `SLP-ProbHard` remains practically hard feasible against the verified official-source ProbHardE2E affine adapter while using fewer stochastic coordinates; predictive advantages are metric dependent. It does *not* show that `SLP-ProbHard` dominates ProbHardE2E across all metrics or constraint classes. DPPL has the better marginal CRPS and higher, closer-to-nominal coverage here; `SLP-ProbHard` has better MSE/MAE and a smaller/sharper representation; the paired test did not detect an Energy Score difference. Nor is the experiment a reproduction of ProbHardE2E's original PDE application. The supported conclusion is a representation-dependent trade-off on a shared affine benchmark.

P AI Use and Ethics Statements

AI use. Generative AI tools were used as research-assistance tools for brainstorming, language editing, code drafting and debugging, experimental workflow scripting, and organization of literature and manuscript structure. All AI-assisted outputs were critically reviewed by the author(s); numerical claims were traced to saved result files and bibliographic additions were checked against publication or publisher sources. The author(s) take responsibility for the final methodology, experiments, analysis, citations, text, and all other content.

Ethics. The study uses synthetic benchmarks and hydrological model/reanalysis products and does not involve human subjects or personal data. The main ethical considerations are research integrity, transparent reporting of limitations, and avoiding overstatement of implementation-specific computational comparisons.