LOCALIZED SUBSPACE SOURCE ADAPTATION FOR FLOW MATCHING

Anonymous authors
Paper under double-blind review

ABSTRACT

Spatially varying local subspaces are not captured by the global statistics commonly used to adapt Flow Matching (FM) sources. We study which target information should be retained in the initial distribution without changing the probability path or FM objective. We introduce **Localized Subspace Source Adaptation (LSSA)**, a source-only adaptation that represents target regions with multiple anchors and orients Gaussian-mixture components using local PCA projectors. LSSA retains dominant regional directions without reproducing local covariance magnitudes, while a shared variance budget controls total scale and an isotropic residual preserves ambient variation. Across structured toy distributions, scientific data, and conditional embedding generation, LSSA consistently improves distributional fidelity. It also demonstrates competitive performance on MNIST compared with standard Gaussian and statistic-specific source baselines. These results show that explicitly modeling localized subspace structure is an effective source-design principle in the evaluated structured and representation spaces for FM.

1 INTRODUCTION

Generation with Flow Matching (FM) starts from a source distribution and follows a learned vector field toward the target data distribution (Lipman et al., 2023; Liu et al., 2023; Tong et al., 2024). While the FM commonly uses an isotropic Gaussian source, recent studies show that adapting this starting distribution can improve generation. SourceFM adjusts the source scale to bring its typical norm closer to that of the data. It also rejects samples pointing toward directions with little target data (Lee et al., 2026). Adapting Noise learns a separate quantile function for each coordinate to reproduce non-Gaussian marginal distributions (Chemseddine et al., 2026). RAFM adapts the distribution of distances from the origin and also modifies the transport path (Oubari & Mougeot, 2026). However, these statistics used in previous methods are pooled across spatial regions and do not explicitly describe how dominant directions of variation change with target-space location. This raises a more specific question: *Should an adaptive source also represent spatially varying local subspaces?*

Representing spatially varying subspaces requires information that global moments can omit. A two-dimensional example in Sec. 3.1 illustrates distributions with identical global means and covariance matrices but different dominant subspaces in the same regions. Thus, global moment matching alone cannot recover this regional assignment. Representing such variation requires information about both *where* the structure changes and *which directions* characterize each region. Figure 1 provides complementary empirical evidence. Statistic-specific sources capture their intended properties in three structured toys, while spatial mismatch remains. A source that additionally represents localized structure improves generation quality, reducing both SWD and MMD errors. Together, these observations motivate a source representation that explicitly encodes both the spatial localization of target structure and its region-specific dominant subspaces.

To encode these two aspects, we propose **Localized Subspace Source Adaptation (LSSA)**. LSSA represents spatially varying source structure with a finite Gaussian mixture. K-means anchors determine *where* components are placed, and nearest-neighbor PCA determines *which directions* dominate around each anchor. Remark 1 summarizes how anchor quantization and standard PCA perturbation theory (Yu et al., 2015) separate finite spatial resolution from local direction-estimation error; Table 7 examines anchor resolution empirically.

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.

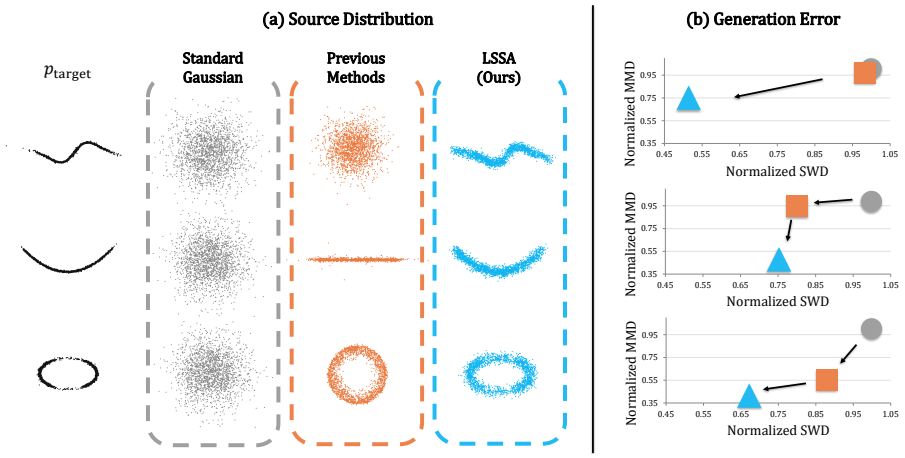

Figure 1: **Spatially varying local subspaces provide complementary source information.** (a) Target and *pre-generation source* samples from the Gaussian, the toy-specific prior method, and LSSA. The prior methods address norm, marginal, and radial statistics on Toys 1–3, respectively. LSSA also represents region-specific dominant directions and yields lower generation discrepancies in these comparisons. (b) Generation SWD and MMD (lower is better), normalized separately within each of three toys. See Appendix C for the toy protocols.

LSSA retains only each leading local subspace, not its eigenvalue magnitudes, so orientation is transferred while scale remains controlled by the source construction. The resulting Gaussian components share a variance budget matched to the expected squared norm of the standard Gaussian, and an isotropic residual preserves nonzero ambient variation outside the selected subspace. We test the locality–orientation decomposition and richer spectrum/scale variants in Sec. 4.3; Appendix A gives the representation and source-property calculations.

We evaluate LSSA on the scientific ProteinConformers dataset (Zhou et al., 2025), COCO (Lin et al., 2014) embeddings from CLIP (Radford et al., 2021) and HyCoCLIP (Pal et al., 2025), and MNIST image generation (LeCun et al., 1998). LSSA attains better SWD and MMD than the compared methods on the reported protein and embedding settings. On MNIST, it improves over the Gaussian in SWD (Rabin et al., 2012), MMD (Gretton et al., 2012), and FID (Heusel et al., 2017).

Our contributions are summarized as follows:

- We identify a limitation of existing source adaptation methods: global or statistic-specific source representations do not explicitly capture how dominant subspaces vary across target-space regions. We formulate this as a spatially varying subspace mismatch in source design.
- We introduce *Localized Subspace Source Adaptation (LSSA)*, a source-only adaptation that assigns region-specific subspaces to anchor-based Gaussian components. LSSA combines K-means anchors, nearest-neighbor PCA, and controlled residual variance to capture localized structure without reproducing full local covariance.
- We evaluate LSSA across structured toy distributions, protein conformers, CLIP/HyCoCLIP embedding generation, and MNIST in both pixel and VAE-latent spaces, showing consistent improvements in distributional fidelity across the tested domains.

2 Related Work

Source distribution design for Flow Matching. Although standard Flow Matching (FM) typically starts from an isotropic Gaussian, recent work treats the source as a modeling choice. SourceFM (Lee et al., 2026) combines norm-aligned training with directionally pruned sampling, emphasizing scale mismatch and broad directional coverage. Adapting Noise (Chemseddine et al., 2026) learns one-dimensional quantile processes to adapt coordinate-wise non-Gaussian marginals. RAFM (Oubari & Mougeot, 2026) matches the target radial law at the source and uses radial–angular transport, thereby modifying both source and path. These methods adapt condition-independent source properties but

do not explicitly index them by target-space location. Conditional-prior FM (Issachar et al., 2025), CAR-Flow (Chen et al., 2025), and CSFM (Kim et al., 2026) instead exploit external conditioning to construct, reparameterize, or learn condition-dependent sources. LSSA is orthogonal to this line: its source remains condition-independent and transfers spatially localized subspaces rather than conditioning information.

Geometry-aware generative modeling and estimation. Local geometry has also been incorporated through the transport process or geometry estimation. RFM (Chen & Lipman, 2024) defines FM paths and vector fields on non-Euclidean spaces, while CDC-FM (Bamberger et al., 2026b) injects spatially varying anisotropic noise reflecting local data geometry into the probability path. RMM (Bamberger et al., 2026a) is instead a learned estimator of local Riemannian geometry through the carré du champ operator. LSSA uses localized covariance subspaces to organize the initial source while leaving the standard FM objective and path unchanged; Appendix D compares an RMM-based projector and tests compatibility with RFM.

3 METHOD

3.1 PROBLEM SETUP AND SOURCE-DESIGN PRINCIPLE

Let p_1 be the target distribution on $\mathbb{R}^D$ and $\mathcal{X} = \{x_n\}_{n=1}^N$ its observed training samples, with $x_n \sim p_1$. Flow Matching (FM) learns a time-dependent vector field whose ODE transports samples from a source distribution q at $t = 0$ toward p_1 at $t = 1$. The conventional FM source is the isotropic Gaussian $q_{\text{ref}} = \mathcal{N}(0, I_D)$ (Lipman et al., 2023; Tong et al., 2024; Liu et al., 2023). We write $\|\cdot\|_2$ for the Euclidean vector norm or matrix spectral norm, and $\|\cdot\|_F$ for the Frobenius norm. Our goal is to change only q and test the effect of source design under a fixed FM learning protocol.

Our objective is not to approximate the target distribution as closely as possible. Prior work shows that increasingly target-like source distributions can even degrade FM generation due to source–target mode discrepancy and path entanglement (Lee et al., 2026). At the extreme, if $q = p_1$ exactly, no transport is required. However, sampling from such a source would already require a sampler for the target distribution itself. The source-design question is therefore *which target information is useful to transfer* for accurate generation. We focus on spatially varying local subspaces. A target distribution can be locally elongated along one direction in one region and along a different direction elsewhere. Global mean and covariance do not generally determine this location–direction association. A single global covariance also provides only one shared set of principal directions.

Local covariance and its subspace projector. For a neighborhood $\mathcal{N}_i$ containing k target samples, define

$$\mu_i = \frac{1}{k}\sum_{x\in\mathcal{N}_i} x, \qquad C_i = \frac{1}{k}\sum_{x\in\mathcal{N}_i}(x-\mu_i)(x-\mu_i)^\top. \tag{1}$$

The covariance C_i describes how nearby coordinates vary together. Let $u_{i,1}, \ldots, u_{i,r}$ be orthonormal eigenvectors associated with its r largest eigenvalues, and set

$$U_i = [u_{i,1}, \ldots, u_{i,r}], \qquad P_i = U_i U_i^\top. \tag{2}$$

The matrix P_i is the orthogonal projector onto the dominant local subspace: $P_i v$ retains the component of a vector v along these directions. It records which directions account for the largest local variation, without retaining their individual variance magnitudes. The rank r is a chosen hyperparameter, not an assumed intrinsic dimension. We specify the anchor neighborhoods in Sec. 3.3.

A two-dimensional example. Consider two distributions, each with equally weighted components centered at $(m, 0)$ and $(-m, 0)$. In p, the right component is elongated horizontally and the left component vertically, with regional covariances $C_+ = \mathrm{diag}(a, b)$ and $C_- = \mathrm{diag}(b, a)$ for $a > b > 0$. The distribution $\widetilde{p}$ swaps these orientations while keeping the centers and weights fixed. Using disjoint uniform ellipses makes the regional covariances exact (Figure 3). Both distributions have mean zero and global covariance $\mathrm{diag}(m^2 + (a+b)/2, (a+b)/2)$, yet their local covariances differ at the same spatial region. Let e_1, e_2 be the coordinate unit vectors, with $P_+ = e_1 e_1^\top$ and $P_- = e_2 e_2^\top$. Then

$$\|P_+ - P_-\|_F^2 = 2, \qquad \frac{1}{2}\sum_{s\in\{+,-\}} \|P - P_s\|_F^2 = 1 \tag{3}$$

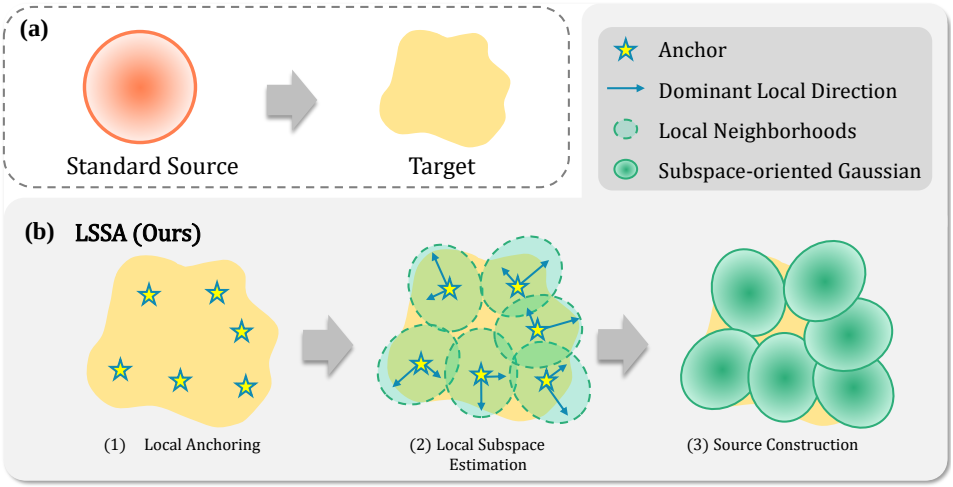

Figure 2: **Overview of LSSA.** (a) The conventional isotropic Gaussian source and the target distribution. (b) LSSA consists of (1) local anchoring to locate target regions, (2) local subspace estimation to identify their dominant directions, and (3) source construction to form subspace-oriented Gaussian components with a shared variance budget.

for any single rank-one orthogonal projector P in $\mathbb{R}^2$. The first quantity makes the difference between the two regional directions explicit; the second shows that one shared direction cannot represent both regions exactly. Assigning the correct projector separately in each region instead gives zero error. Thus, identical global moments can conceal different location–direction associations. Appendix A.2 gives the construction and calculations.

LSSA represents these associations using local anchors and covariance projectors, then combines them in a Gaussian-mixture source. Figure 2 summarizes the three steps, and Algorithm 1 in Appendix B provides the complete procedure.

3.2 Local Anchoring

A single Gaussian component cannot attach different covariance orientations to different target regions. We therefore use a finite mixture as a tractable representation of spatially varying source structure. The mixture requires a set of locations that summarize the target support. Possible choices include random prototypes, grid-based centers, or data-adaptive anchors. We use K-means anchors (Lloyd, 1982) to reduce empirical target-to-anchor squared distances. The quantization view of K-means is established in approximation analysis (Canas et al., 2012). Here, the corresponding population distance controls the error of representing a smoothly varying subspace field. This does not claim that K-means is the only valid anchor rule. We evaluate the effect of anchor resolution by varying K in the experiments.

We fit K anchors to $\mathcal{X}$ before FM training and keep them fixed thereafter. For $i \in \{1, \dots, K\}$, let $a_i \in \mathbb{R}^D$ be the i-th anchor. Its empirical Voronoi cell is

$$\mathcal{V}_i = \{x_n : i = \arg\min_{1 \le j \le K} \|x_n - a_j\|_2\}. \tag{4}$$

Ties are resolved by a fixed rule. We set the mixture weight to the empirical region frequency,

$$\pi_i = \frac{|\mathcal{V}_i|}{N}, \qquad \sum_{i=1}^{K} \pi_i = 1. \tag{5}$$

Thus, π_i determines how often a region is sampled. It does not determine the covariance scale of that component.

To describe the retained information, fix a population neighborhood rule. Let $P^\star(x)$ be its dominant rank-r covariance projector for $1 \le r < D$. The field is defined on a domain containing the target

support and all anchors. Assume $\|P^\star(x) - P^\star(y)\|_F \le L\|x - y\|_2$ on this domain. Let $A_K(x)$ denote the nearest-anchor index. The ideal anchor-based approximation is $P_K^\star(x) = P^\star(a_{A_K(x)})$. Applying the Lipschitz condition at the nearest anchor directly gives

$$\mathbb{E}\|P^\star(X) - P_K^\star(X)\|_F^2 \le L^2\Delta_K, \qquad \Delta_K = \mathbb{E}\|X - a_{A_K(X)}\|_2^2. \tag{6}$$

Expectations use a fresh draw $X \sim p_1$ conditional on the fitted representation. This population localization term is summarized together with local direction-estimation error in Remark 1 below.

For population-optimal anchors, Δ_K is non-increasing in K and tends to zero on compact target support. These are standard quantization properties (Canas et al., 2012). The bound explains why finer spatial resolution can represent smoothly varying directions more accurately. It does not guarantee monotonic improvement for empirical K-means fits or generation metrics. We assess generation through the single-anchor and anchor-resolution analyses.

3.3 Local Subspace Estimation

Anchors identify where the target is represented. Local covariance estimates supply the directions of variation in those regions. Around each anchor, the target can be stretched more strongly along some directions than others. We estimate these dominant directions from a local neighborhood. For each a_i, let $\mathcal{N}_i$ be its k nearest target samples in Euclidean distance. The neighborhoods may overlap and are distinct from the Voronoi cells used for mixture weights. This separation keeps regional sampling frequency and local direction estimation as different design choices.

We compute C_i and P_i using Eqs. (1)–(2), following neighborhood PCA (Zhang & Zha, 2004). LSSA transfers the dominant subspace without its eigenvalue magnitudes. The information ablation tests this choice against alternatives that also transfer the local spectrum or a density-dependent scale.

Reliable direction estimates are needed to retain this information. For the following analysis, take $1 \le r < D$. Write $C_i^\star$ for the population neighborhood covariance and $P_i^\star = P^\star(a_i)$ for its top-r projector. Let $\varepsilon_i = \|C_i - C_i^\star\|_2$ and $\delta_i = \lambda_r(C_i^\star) - \lambda_{r+1}(C_i^\star) > 0$. The Davis–Kahan bound controls projector error through ε_i/δ_i (Yu et al., 2015, Theorem 2). The following remark records how this standard estimation result combines with the localization bound.

Remark 1 (Localized subspace representation error)**.** *Under the Lipschitz assumption in Sec. 3.2, the population approximation satisfies Eq. (6). For fixed fitted anchors and estimated rank-r projectors, let $\widehat{P}_K(x) = P_{A_K(x)}$ and $\rho_i = \Pr[A_K(X) = i]$. Then*

$$\mathbb{E}\|P^\star(X) - \widehat{P}_K(X)\|_F^2 \le 2L^2\Delta_K + 2\sum_i \rho_i \|P_i - P_i^\star\|_F^2. \tag{7}$$

If $\delta_i > 0$ at every anchor, the cited perturbation bound gives

$$\mathbb{E}\|P^\star(X) - \widehat{P}_K(X)\|_F^2 \le 2L^2\Delta_K + 16r\sum_i \rho_i \left(\frac{\varepsilon_i}{\delta_i}\right)^2. \tag{8}$$

The first term measures spatial approximation error. The second measures local direction-estimation error. Both terms must vanish for this bound to establish consistency of the estimated field. Increasing K alone does not control the second term. Fixed-k neighborhoods do not in general provide covariance consistency as the dataset grows. Appendix A.3 gives the short derivation and states the population-neighborhood conditions. Importantly, this remark does not predict FM generation error. Rather, it characterizes when a finite anchor representation can faithfully retain a spatially varying subspace field. The analysis concerns the subspaces assigned to anchors and does not assert that overlapping source components reproduce the target's local covariance field. Together, local anchoring and local PCA therefore provide a finite, piecewise description of where dominant directions occur; we next encode this description in the source while controlling its overall scale.

3.4 Source Construction

We combine the anchors and their projectors in a Gaussian mixture:

$$q_{\mathrm{LSSA}}(x) = \sum_{i=1}^{K} \pi_i \mathcal{N}(x; m_i, \Sigma_i), \tag{9}$$

with

$$m_i = c(a_i - \bar{a}), \qquad \bar{a} = \sum_i \pi_i a_i, \tag{10}$$

$$\Sigma_i = R^2 \left[(1-\beta)\frac{I_D}{D} + \beta \frac{P_i}{r} \right]. \tag{11}$$

This low-rank-plus-isotropic form follows probabilistic PCA mixtures (Tipping & Bishop, 1999). The fraction β emphasizes the selected subspace, while the isotropic residual retains variance in all ambient directions. Each component has the same trace R^2. We fix $\beta = 0.5$ and calibrate the shared scales c, R to match the Gaussian reference second moment $T_{\text{ref}} = D$. Appendix A.4 gives the calibration and verifies these properties. Thus, LSSA transfers regional frequency and orientation without copying region-specific covariance magnitudes.

3.5 Flow Matching with LSSA

LSSA changes only the source distribution. Let $\gamma \in \Pi(q_{\text{LSSA}}, p_1)$ denote the chosen source–target coupling. We sample $(x_0, x_1) \sim \gamma$ and $t \sim \mathcal{U}[0,1]$. Standard linear FM then uses

$$x_t = (1-t)x_0 + tx_1, \qquad \mathcal{L}_{\text{FM}} = \mathbb{E}_{(x_0,x_1)\sim\gamma,\, t}\left[\|v_\theta(x_t,t) - (x_1 - x_0)\|_2^2 \right], \tag{12}$$

where $v_\theta : \mathbb{R}^D \times [0,1] \to \mathbb{R}^D$ is the learned time-dependent vector field. For conditional generation, the vector field additionally receives a condition y, i.e., $v_\theta(x_t, t; y)$, while the source distribution itself remains condition-independent. For COCO embedding generation, we use classifier-free guidance (CFG) through the conditional vector field. Importantly, LSSA adapts only the unconditional source marginal $q(x_0)$; conditioning information y is provided only to the vector field. The conditional path construction and FM objective are otherwise unchanged. Algorithm 1 summarizes source preparation, sampling, and FM training.

4 Experiments

4.1 Evaluation Protocol

Datasets and tasks. We evaluate source adaptation in three domains. For scientific data, we evaluate distribution generation on ProteinConformers (Zhou et al., 2025) and report aggregate metrics across its evaluated targets. For representation-space generation, we use image and text embeddings of MS COCO (Lin et al., 2014) obtained with CLIP (Radford et al., 2021) and HyCoCLIP (Pal et al., 2025), evaluating image-to-text (I2T) and text-to-image (T2I) directions separately. For image generation, we train on MNIST (LeCun et al., 1998). Preprocessing, data splits, and representation details are described in Appendix D.

Baselines and comparison protocol. We compare Localized Subspace Source Adaptation (LSSA) with the standard isotropic Gaussian (Standard) (Lipman et al., 2023), SourceFM (Lee et al., 2026), Adapting Noise (Chemseddine et al., 2026), and RAFM (Oubari & Mougeot, 2026). Within each dataset and generation direction, methods share the representation-appropriate backbone. For the main method-level comparison, we follow the recommended method-specific source, coupling, and path configurations of each baseline where applicable, including RAFM's radial path. Thus, the main tables compare LSSA against the complete baseline methods rather than restricting all competitors to source-only modifications. LSSA itself modifies only the source distribution. To isolate the contribution of source design, the LSSA ablations keep the backbone, path, coupling, optimizer, and training budget fixed across variants. All vector-field models are optimized with AdamW in our experiments. LSSA estimates its anchors and local subspaces using training data only. Implementation details are provided in Appendix D.

Backbones. We use a time-conditioned MLP for ProteinConformers, a conditional MLP with classifier-free guidance (CFG) for COCO embedding generation, and spatial U-Nets for MNIST pixel and VAE-latent generation. Within each comparison, the backbone is shared across source choices; architecture settings are given in Appendix D.

Metrics. We primarily measure distributional agreement using sliced Wasserstein distance (SWD) (Rabin et al., 2012) and maximum mean discrepancy (MMD) (Gretton et al., 2012); lower

Table 1: **ProteinConformers generation.** LSSA achieves the lowest aggregate SWD and MMD among the compared sources. Best and second-best results are shown in **bold** and underlined.

Method	SWD ↓	MMD ↓
Standard	0.3577	0.0079
SourceFM	0.3296	0.0052
AdaptingNoise	0.3239	0.0035
RAFM	0.3037	0.0040
LSSA (Ours)	**0.2374**	**0.0021**

Table 2: **Generation in CLIP embedding space.** LSSA leads both distributional metrics and all retrieval metrics except T2I R@10.

	CLIP I2T					CLIP T2I				
	SWD ↓	MMD ↓	R@1 ↑	R@5 ↑	R@10 ↑	SWD ↓	MMD ↓	R@1 ↑	R@5 ↑	R@10 ↑
Standard	0.0053	0.0061	8.32	22.66	33.10	0.0061	0.0108	3.34	11.70	19.08
SourceFM	0.0037	0.0033	9.01	23.55	34.06	0.0029	0.0031	4.37	14.45	**22.62**
AdaptingNoise	0.0041	0.0036	8.69	23.72	34.13	0.0023	0.0016	4.12	13.63	21.23
RAFM	0.0044	0.0045	9.96	25.67	36.36	0.0055	0.0092	3.80	13.07	21.21
LSSA (Ours)	**0.0023**	**0.0014**	**10.18**	**26.10**	**36.68**	**0.0017**	**0.0008**	**4.58**	**14.50**	22.51

Table 3: **Generation in HyCoCLIP embedding space.** LSSA achieves the best distributional and retrieval results in both directions.

	HyCoCLIP I2T					HyCoCLIP T2I				
	SWD ↓	MMD ↓	R@1 ↑	R@5 ↑	R@10 ↑	SWD ↓	MMD ↓	R@1 ↑	R@5 ↑	R@10 ↑
Standard	0.0043	0.0122	8.22	23.12	32.92	0.0068	0.0127	5.07	17.28	27.12
SourceFM	0.0021	0.0037	9.62	25.16	35.16	0.0033	0.0039	7.54	22.97	34.08
AdaptingNoise	0.0020	0.0035	9.06	24.34	34.64	0.0030	0.0033	8.02	23.58	34.51
RAFM	0.0026	0.0081	8.54	24.00	34.68	0.0031	0.0046	7.41	22.80	33.85
LSSA (Ours)	**0.0010**	**0.0018**	**10.02**	**25.60**	**36.22**	**0.0021**	**0.0017**	**8.20**	**23.75**	**34.93**

values indicate closer agreement under their specified estimators. Image generation is additionally evaluated using Fréchet Inception Distance (FID) on generated images (Heusel et al., 2017). For images, SWD and MMD are computed on consistently preprocessed pixel vectors and VAE latents, FID is computed at the image level. For COCO embeddings, we additionally report retrieval recall at ranks 1, 5, and 10 (R@1/5/10; higher is better) for each image-to-text (I2T) or text-to-image (T2I) direction.

4.2 Main Results

Scientific distribution generation. We evaluate whether localized source structure improves the generation of protein conformational distributions. Table 1 shows that LSSA achieves lower aggregate SWD and MMD than the standard Gaussian, SourceFM, Adapting Noise, and RAFM. These results support the value of transferring regional subspaces for distributional fidelity on ProteinConformers.

Embedding-space generation. We test whether source adaptation improves both distributional agreement and cross-modal retrieval across different embedding representations. Tables 2 and 3 show that LSSA achieves the lowest SWD and MMD in both directions for CLIP and HyCoCLIP. It also leads all reported retrieval metrics except CLIP T2I R@10, where SourceFM performs slightly better. These results indicate that localized source structure benefits both embedding fidelity and retrieval utility across the tested representations.

Image generation. We assess whether the same source design improves MNIST generation in both pixel and VAE latent spaces. In Table 4, LSSA achieves the lowest SWD and MMD in both spaces and the lowest FID for latent-space generation; for pixel-space generation, its FID is second to Adapting Noise. This comparison supports the applicability of LSSA to image generation, with consistent distributional gains and competitive image quality.

Table 4: **MNIST generation in pixel and latent spaces.** LSSA leads SWD and MMD in both spaces and latent-space FID, with second-best pixel-space FID.

	Pixel-level			Latent-level		
	SWD ↓	MMD ↓	FID ↓	SWD ↓	MMD ↓	FID ↓
Standard	0.0274	0.0049	7.99	0.2740	0.0205	9.75
SourceFM	0.0273	0.0030	4.90	<u>0.2623</u>	<u>0.0180</u>	<u>9.35</u>
AdaptingNoise	<u>0.0269</u>	<u>0.0027</u>	**3.29**	0.2627	0.0188	9.72
RAFM	0.0331	0.0040	9.25	0.2648	0.0187	9.43
LSSA (Ours)	**0.0249**	**0.0025**	<u>3.49</u>	**0.2481**	**0.0153**	**9.05**

Table 5: **Localization and orientation ablation.** Local-Subspace achieves the lowest SWD and MMD across all evaluated domains. All variants preserve the reference source second moment; $K^* > 1$ denotes the fixed dataset-specific anchor count, and subspace variants use the fixed LSSA setting $\beta = 0.5$.

			ProteinConformers		CLIP (I2T)		HyCoCLIP (I2T)	
Variant	K	β	SWD↓	MMD↓	SWD↓	MMD↓	SWD↓	MMD↓
Global-Iso	1	0	0.3577	0.0079	0.0053	0.0061	0.0043	0.0122
Global-Subspace	1	0.5	0.3250	0.0044	0.0041	0.0037	0.0016	<u>0.0021</u>
Local-Iso	K^*	0	<u>0.3007</u>	<u>0.0034</u>	<u>0.0026</u>	<u>0.0036</u>	<u>0.0015</u>	0.0022
Local-Subspace (Ours)	K^*	0.5	**0.2374**	**0.0021**	**0.0023**	**0.0014**	**0.0010**	**0.0018**

Table 6: **Local information ablation.** Projector-only Local-Subspace outperforms Local-Iso and the spectrum- and density-augmented variants. Anchors, mixture weights, and the overall source second moment are matched. Spectrum transfers normalized eigenvalue ratios; Density adapts component scales.

	ProteinConformers		CLIP (I2T)		HyCoCLIP (I2T)	
Variant	SWD↓	MMD↓	SWD↓	MMD↓	SWD↓	MMD↓
Local-Iso	<u>0.3007</u>	0.0034	0.0026	0.0036	0.0015	<u>0.0022</u>
Local-Subspace (Ours)	**0.2374**	**0.0021**	**0.0023**	**0.0014**	**0.0010**	**0.0018**
Subspace + Spectrum	0.3110	<u>0.0029</u>	0.0029	0.0023	<u>0.0012</u>	0.0023
Subspace + Density	0.3069	0.0033	<u>0.0025</u>	<u>0.0017</u>	0.0013	0.0027

4.3 Ablation Studies

Contributions of localization and subspace orientation. To test the location–direction design in Sec. 3, we independently vary localization through K and orientation through β in Eq. (11). The factorial design compares Global-Iso ($K = 1, \beta = 0$), Global-Subspace ($K = 1, \beta = 0.5$), Local-Iso ($K = K^*, \beta = 0$), and Local-Subspace ($K = K^*, \beta = 0.5$; LSSA), where $K^* > 1$ is the fixed LSSA anchor count. For $K = 1$, we use a zero component mean and trace T_{ref}, with global PCA providing the subspace projector; the local variants share anchors and mixture weights. All variants preserve the reference source second moment. Table 5 shows that Local-Subspace outperforms both global variants and Local-Iso on every reported metric. Although the relative benefit of localization alone varies across domains, combining regional locations with their subspace orientations consistently performs best, supporting the joint design.

Selection of local information. The covariance construction in Sec. 3.4 retains dominant directions while discarding eigenvalue magnitudes and sharing a common component trace. We test this choice against variants that additionally transfer normalized local eigenvalue ratios or density-dependent component scales, keeping anchors, mixture weights, coupling, and the overall source second moment fixed. In particular, density adaptation changes component scale rather than regional sampling mass. Table 6 shows that projector-only Local-Subspace achieves lower SWD and MMD than Local-Iso and both richer variants across all tested domains. These results support the constrained transfer of local orientation without additional spectrum or scale adaptation in the evaluated settings.

Table 7: **Anchor resolution.** Local-Subspace (S) achieves lower SWD than Local-Iso (I) at every tested anchor count K. Local-Subspace improves or remains unchanged as K increases. Lower is better.

	ProteinConformers		CLIP (I2T)		HyCoCLIP (I2T)	
K	I	S	I	S	I	S
1	0.3577	0.3250	0.0053	0.0041	0.0043	0.0016
8	0.3180	0.2744	0.0029	0.0028	0.0020	0.0015
16	<u>0.3006</u>	0.2558	0.0028	0.0026	0.0015	<u>0.0013</u>
32	0.3007	<u>0.2374</u>	**0.0026**	<u>0.0024</u>	**0.0014**	<u>0.0013</u>
64	**0.2385**	**0.2051**	**0.0026**	**0.0023**	**0.0014**	**0.0011**

Table 8: **Projector-to-anchor correspondence.** Correct Local-Subspace outperforms shuffled Local-Subspace and Local-Iso on all reported metrics. Shuffling preserves the projector set and permutes its assignment to anchors.

	ProteinConformers		CLIP (I2T)		HyCoCLIP (I2T)	
Variant	SWD ↓	MMD ↓	SWD ↓	MMD ↓	SWD ↓	MMD ↓
Local-Iso	0.3007	0.0034	0.0026	0.0036	0.0015	<u>0.0022</u>
Shuffled Local-Subspace	<u>0.2768</u>	<u>0.0030</u>	<u>0.0024</u>	<u>0.0016</u>	<u>0.0012</u>	0.0025
Correct Local-Subspace	**0.2374**	**0.0021**	**0.0023**	**0.0014**	**0.0010**	**0.0018**

4.4 Analysis

Anchor resolution. Table 7 varies K and shows lower SWD for Local-Subspace than Local-Iso at every tested resolution, with Local-Subspace improving or remaining unchanged as K increases. This trend is consistent with the spatial approximation term in Remark 1, which does not imply monotonic improvement in generation metrics.

Spatial correspondence. To test the region-specific assignment of P_i in Eq. (11), we shuffle projectors across anchors while preserving component centers, mixture weights, the projector set, and the common variance budget. Correct assignments outperform shuffled assignments and Local-Iso on all reported metrics (Table 8). This supports the importance of associating each subspace with its corresponding region, beyond introducing anisotropy alone.

Additional analyses. Appendix Tables 10, 11 show that local PCA is a strong default relative to an RMM-derived estimator (Bamberger et al., 2026a), and LSSA remains compatible with RFM (Chen & Lipman, 2024).

5 Conclusion

Spatially pooled source statistics cannot specify how dominant subspaces vary across target regions. LSSA addresses this gap by assigning region-specific principal subspaces to anchor-based Gaussian components under a fixed second-moment budget. The analysis separates finite localization from projector-estimation error without claiming a generation guarantee. Across ProteinConformers and CLIP/HyCoCLIP embeddings, LSSA achieves the lowest reported SWD and MMD among the compared methods and improves over the Gaussian on MNIST. Ablations show that localization and orientation are complementary and that correct region–subspace correspondence matters. Together, these results support a focused principle: an adaptive source need not reproduce the full target distribution; retaining *where* structure changes and *which directions* dominate can suffice.

AI USE STATEMENT

In this work, we used generative AI tools to aid or polish writing . We have not used generative AI tools for other tasks with required disclosure, and the rest of the required disclosure tasks are not applicable to this work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Jacob Bamberger, Adam Gosztolai, Pierre Vandergheynst, Michael Bronstein, and Iolo Jones. Riemannian metric matching for scalable geometric modeling of distributions. In *International Conference on Machine Learning*, 2026a. Oral.

Jacob Bamberger, Iolo Jones, Dennis Duncan, Michael M. Bronstein, Pierre Vandergheynst, and Adam Gosztolai. Carré du champ flow matching: better quality-generalisation tradeoff in generative models. In *International Conference on Learning Representations*, 2026b.

Guillermo Canas, Tomaso Poggio, and Lorenzo Rosasco. Learning manifolds with K-means and K-flats. In *Advances in Neural Information Processing Systems*, volume 25, 2012. URL https://proceedings.neurips.cc/paper_files/paper/2012/hash/b20bb95ab626d93fd976af958fbc61ba-Abstract.html.

Jannis Chemseddine, Gregor Kornhardt, Richard Duong, and Gabriele Steidl. Adapting noise to data: Generative flows from 1D processes. In *International Conference on Machine Learning*, 2026.

Chen Chen, Pengsheng Guo, Liangchen Song, Jiasen Lu, Rui Qian, Tsu-Jui Fu, Xinze Wang, Wei Liu, Yinfei Yang, and Alex Schwing. CAR-flow: Condition-aware reparameterization aligns source and target for better flow matching. 2025.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations*, 2024.

Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, 5th edition, 2019. doi: 10.1017/9781108591034.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 2012.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, 2017.

Noam Issachar, Mohammad Salama, Raanan Fattal, and Sagie Benaim. Designing a conditional prior distribution for flow-based generative models. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856.

Junwan Kim, Jiho Park, Seonghu Jeon, and Seungryong Kim. Better source, better flow: Learning condition-dependent source distribution for flow matching. *arXiv preprint arXiv:2602.05951*, 2026.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998.

Junho Lee, Kwanseok Kim, and Joonseok Lee. Is there a better source distribution than gaussian? exploring source distributions for image flow matching. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In *European Conference on Computer Vision*, 2014.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *International Conference on Learning Representations*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *International Conference on Learning Representations*, 2023.

Stuart P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28 (2):129–137, 1982. doi: 10.1109/TIT.1982.1056489.

Fouad Oubari and Mathilde Mougeot. Correcting source mismatch in flow matching with radial-angular transport. *arXiv preprint arXiv:2604.04291*, 2026.

Avik Pal, Max van Spengler, Guido Maria D'Amely di Melendugno, Alessandro Flaborea, Fabio Galasso, and Pascal Mettes. Compositional entailment learning for hyperbolic vision-language models. 2025.

Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *Scale Space and Variational Methods in Computer Vision*, volume 6667 of *Lecture Notes in Computer Science*, pp. 435–446. Springer, 2012. doi: 10.1007/978-3-642-24785-9_37.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. *International Conference on Machine Learning*, 2021.

Michael E. Tipping and Christopher M. Bishop. Mixtures of probabilistic principal component analysers. *Neural Computation*, 11(2):443–482, 1999.

Alexander Tong, Kilian FATRAS, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856.

Yi Yu, Tengyao Wang, and Richard J. Samworth. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015. doi: 10.1093/biomet/asv008.

Zhenyue Zhang and Hongyuan Zha. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM Journal on Scientific Computing*, 26(1):313–338, 2004. doi: 10.1137/S1064827502419154.

Yihang Zhou, Chen Wei, Minghao Sun, Jin Song, Yang Li, Lin Wang, and Yang Zhang. Proteinconformers: Benchmark dataset for simulating protein conformational landscape diversity and plausibility. In *Advances in Neural Information Processing Systems*, 2025.