Theorem 1 (Monotonicity of LV–EBM ELBO over variational families)**.** Fix a dataset $\{x_i\}_{i=1}^N$ and a latent-variable energy-based model

$$p_\theta(x,z) \;=\; \frac{1}{Z(\theta)} \exp\big(-E_\theta(x,z)\big),$$

with $Z(\theta) < \infty$ and such that, for each i, the conditional density $p_\theta(z \mid x_i)$ exists and is strictly positive on its support.

Let $\mathcal{Q}$ be the set of probability densities $q(z)$ on the latent space with finite differential entropy, and let $\tilde{\mathcal{Q}}$ be the analogous set of joint densities $\tilde{q}(x,z)$ with finite differential entropy. For $q \in \mathcal{Q}$ and $\tilde{q} \in \tilde{\mathcal{Q}}$ define

$$A_i(q) \;:=\; -\mathbb{E}_{q(z)}\big[E_\theta(x_i,z)\big] + H(q), \qquad B(\tilde{q}) \;:=\; -\mathbb{E}_{\tilde{q}(x,z)}\big[E_\theta(x,z)\big] + H(\tilde{q}),$$

and for any collections of admissible families $S_{\text{pos}} = \{S_{\text{pos}}(x_i)\}_{i=1}^N$ with $S_{\text{pos}}(x_i) \subseteq \mathcal{Q}$ and $S_{\text{neg}} \subseteq \tilde{\mathcal{Q}}$ define

$$\text{ELBO}(\theta; S_{\text{pos}}, S_{\text{neg}}) \;:=\; \frac{1}{N}\sum_{i=1}^N \sup_{q\in S_{\text{pos}}(x_i)} A_i(q) \;-\; \sup_{\tilde{q}\in S_{\text{neg}}} B(\tilde{q}).$$

Assume that, for all θ and all admissible choices of $S_{\text{pos}}, S_{\text{neg}}$, the above suprema are finite. Let

$$S^1_{\text{pos}} = \big\{S^1_{\text{pos}}(x_i)\big\}_{i=1}^N, \qquad S^2_{\text{pos}} = \big\{S^2_{\text{pos}}(x_i)\big\}_{i=1}^N$$

be two collections of positive-phase families with $S^1_{\text{pos}}(x_i) \subseteq S^2_{\text{pos}}(x_i) \subseteq \mathcal{Q}$ for all i, and let $S^1_{\text{neg}} \subseteq S^2_{\text{neg}} \subseteq \tilde{\mathcal{Q}}$ be two admissible negative-phase families. Then, for every θ,

$$\text{ELBO}\big(\theta; S^1_{\text{pos}}, S^1_{\text{neg}}\big) \;\le\; \text{ELBO}\big(\theta; S^2_{\text{pos}}, S^2_{\text{neg}}\big) \;\le\; \frac{1}{N}\sum_{i=1}^N \log p_\theta(x_i). \tag{1}$$

Moreover, the first inequality in (1) is strict for a given θ whenever at least one of the following holds:

(i) For some i, every maximizer of $\sup_{q\in S^2_{\text{pos}}(x_i)} A_i(q)$ lies outside $S^1_{\text{pos}}(x_i)$.

(ii) Every maximizer of $\sup_{\tilde{q}\in S^2_{\text{neg}}} B(\tilde{q})$ lies outside S^1_{neg}.

Proof. By the Donsker–Varadhan variational formula applied to $f_x(z) := -E_\theta(x,z)$ and $f(x,z) := -E_\theta(x,z)$, we have

$$\log p_\theta(x_i) = \sup_{q\in\mathcal{Q}} A_i(q),$$
$$\log Z(\theta) = \sup_{\tilde{q}\in\tilde{\mathcal{Q}}} B(\tilde{q}),$$

with maximizers $q^\star(\cdot) = p_\theta(\cdot \mid x_i)$ and $\tilde{q}^\star = p_\theta(x,z)$, respectively. Equivalently, using the identities

$$\mathbb{E}_{\tilde{q}}[E_\theta] - H(\tilde{q}) = \text{KL}(\tilde{q}\|p_\theta) - \log Z(\theta), \qquad \mathbb{E}_q[E_\theta(x_i,\cdot)] - H(q) = \text{KL}(q\|p_\theta(\cdot \mid x_i)) - \log p_\theta(x_i),$$

we obtain

$$A_i(q) = \log p_\theta(x_i) - \text{KL}\big(q \,\|\, p_\theta(\cdot \mid x_i)\big),$$
$$B(\tilde{q}) = \log Z(\theta) - \text{KL}\big(\tilde{q} \,\|\, p_\theta\big).$$

Therefore, for any $S_{\text{pos}}, S_{\text{neg}}$,

$$\sup_{q\in S_{\text{pos}}(x_i)} A_i(q) = \log p_\theta(x_i) - \inf_{q\in S_{\text{pos}}(x_i)} \text{KL}\big(q \,\|\, p_\theta(\cdot \mid x_i)\big),$$
$$\sup_{\tilde{q}\in S_{\text{neg}}} B(\tilde{q}) = \log Z(\theta) - \inf_{\tilde{q}\in S_{\text{neg}}} \text{KL}\big(\tilde{q} \,\|\, p_\theta\big),$$

and hence

$$\text{ELBO}(\theta; S_{\text{pos}}, S_{\text{neg}}) = \frac{1}{N}\sum_{i=1}^N \log p_\theta(x_i) - \log Z(\theta) - \frac{1}{N}\sum_{i=1}^N \inf_{q\in S_{\text{pos}}(x_i)} \text{KL}\big(q \,\|\, p_\theta(\cdot \mid x_i)\big) + \inf_{\tilde{q}\in S_{\text{neg}}} \text{KL}\big(\tilde{q} \,\|\, p_\theta\big).$$

Now take two nested pairs $(S^1_{\mathrm{pos}}, S^1_{\mathrm{neg}})$ and $(S^2_{\mathrm{pos}}, S^2_{\mathrm{neg}})$ as in the statement. For each i,

$$\inf_{q\in S^2_{\mathrm{pos}}(x_i)} \mathrm{KL}\big(q \,\|\, p_\theta(\cdot \mid x_i)\big) \;\le\; \inf_{q\in S^1_{\mathrm{pos}}(x_i)} \mathrm{KL}\big(q \,\|\, p_\theta(\cdot \mid x_i)\big),$$

and likewise

$$\inf_{\tilde q\in S^2_{\mathrm{neg}}} \mathrm{KL}\big(\tilde q \,\|\, p_\theta\big) \;\le\; \inf_{\tilde q\in S^1_{\mathrm{neg}}} \mathrm{KL}\big(\tilde q \,\|\, p_\theta\big),$$

since infima over supersets cannot be larger. Substituting these inequalities into the expression for $\mathrm{ELBO}(\theta;\cdot,\cdot)$ yields

$$\mathrm{ELBO}\big(\theta; S^1_{\mathrm{pos}}, S^1_{\mathrm{neg}}\big) \;\le\; \mathrm{ELBO}\big(\theta; S^2_{\mathrm{pos}}, S^2_{\mathrm{neg}}\big)$$

for all θ, with strict inequality whenever at least one of the two infima is strictly smaller over $S^2_{\mathrm{pos}}, S^2_{\mathrm{neg}}$ than over $S^1_{\mathrm{pos}}, S^1_{\mathrm{neg}}$; this is exactly conditions (i)–(ii).

Finally, taking $S^2_{\mathrm{pos}}(x_i) = \mathcal{Q}$ and $S^2_{\mathrm{neg}} = \tilde{\mathcal{Q}}$, we recover $\sup_{q\in\mathcal{Q}} A_i(q) = \log p_\theta(x_i)$ and $\sup_{\tilde q\in\tilde{\mathcal{Q}}} B(\tilde q) = \log Z(\theta)$, so

$$\mathrm{ELBO}(\theta; \mathcal{Q}, \tilde{\mathcal{Q}}) = \frac{1}{N}\sum_{i=1}^{N} \log p_\theta(x_i),$$

which gives the upper bound in (1). □

Corollary 1 (LV–EBM gradient-flow families dominate parametric VI)**.** In the setting of Theorem 1, fix θ and let $S^{\mathrm{VI}}_{\mathrm{pos}}(x_i) \subseteq \mathcal{Q}$ and $S^{\mathrm{VI}}_{\mathrm{neg}} \subseteq \tilde{\mathcal{Q}}$ be any finite-dimensional "variational inference" families (e.g. diagonal Gaussian conditionals and a parametric joint). Define the corresponding VI objective by

$$\mathrm{ELBO}_{\mathrm{VI}}(\theta) \;:=\; \mathrm{ELBO}\big(\theta; S^{\mathrm{VI}}_{\mathrm{pos}}, S^{\mathrm{VI}}_{\mathrm{neg}}\big).$$

Let $(\mathcal{T}^x_t)_{t\ge 0}$ and $(\tilde{\mathcal{T}}_t)_{t\ge 0}$ denote the Markov semigroups associated with the conditional and joint Langevin dynamics used in LV–EBMs, assumed to be well-defined and Feller under the regularity assumptions on E_θ in the main text. For any fixed time horizon $t \ge 0$, define the (infinite-dimensional) gradient-flow families

$$S^{\mathrm{flow}}_{\mathrm{pos}}(x_i) \;:=\; \big\{\mathcal{T}^{x_i}_t(q_0) : q_0 \in \mathcal{Q}\big\} \subseteq \mathcal{Q}, \qquad S^{\mathrm{flow}}_{\mathrm{neg}} \;:=\; \big\{\tilde{\mathcal{T}}_t(\tilde q_0) : \tilde q_0 \in \tilde{\mathcal{Q}}\big\} \subseteq \tilde{\mathcal{Q}}.$$

Assume that every VI density can be realized as an admissible initial distribution for the semigroups, and that we include $t = 0$ in the definition of the flow families (so that $\mathcal{T}^x_0$ and $\tilde{\mathcal{T}}_0$ are the identity maps on $\mathcal{Q}$ and $\tilde{\mathcal{Q}}$). Then

$$S^{\mathrm{VI}}_{\mathrm{pos}}(x_i) \;\subseteq\; S^{\mathrm{flow}}_{\mathrm{pos}}(x_i), \qquad S^{\mathrm{VI}}_{\mathrm{neg}} \;\subseteq\; S^{\mathrm{flow}}_{\mathrm{neg}},$$

and, for all θ,

$$\mathrm{ELBO}_{\mathrm{VI}}(\theta) \;\le\; \mathrm{ELBO}\big(\theta; S^{\mathrm{flow}}_{\mathrm{pos}}, S^{\mathrm{flow}}_{\mathrm{neg}}\big) \;\le\; \frac{1}{N}\sum_{i=1}^{N} \log p_\theta(x_i). \tag{2}$$

Moreover, the first inequality in (2) is strict whenever, for the given θ, at least one of the VI families $S^{\mathrm{VI}}_{\mathrm{pos}}(x_i)$ or $S^{\mathrm{VI}}_{\mathrm{neg}}$ is misspecified in the sense that its best-approximation KL divergence to the true posterior $p_\theta(z \mid x_i)$ or joint $p_\theta(x, z)$ is strictly positive.

Proof. By definition of the semigroups, $\mathcal{T}^x_0$ and $\tilde{\mathcal{T}}_0$ act as the identity on $\mathcal{Q}$ and $\tilde{\mathcal{Q}}$, so any VI density $q_\phi(z \mid x_i)$ or $q_\psi(x, z)$ can be written as $\mathcal{T}^{x_i}_0(q_0)$ or $\tilde{\mathcal{T}}_0(\tilde q_0)$ with $q_0 = q_\phi$ and $\tilde q_0 = q_\psi$. Hence $S^{\mathrm{VI}}_{\mathrm{pos}}(x_i) \subseteq S^{\mathrm{flow}}_{\mathrm{pos}}(x_i)$ and $S^{\mathrm{VI}}_{\mathrm{neg}} \subseteq S^{\mathrm{flow}}_{\mathrm{neg}}$. Applying Theorem 1 with $(S^1_{\mathrm{pos}}, S^1_{\mathrm{neg}}) = (S^{\mathrm{VI}}_{\mathrm{pos}}, S^{\mathrm{VI}}_{\mathrm{neg}})$ and $(S^2_{\mathrm{pos}}, S^2_{\mathrm{neg}}) = (S^{\mathrm{flow}}_{\mathrm{pos}}, S^{\mathrm{flow}}_{\mathrm{neg}})$ gives the inequalities in (2). Strictness follows from the strict part of Theorem 1 whenever at least one of the VI families cannot realize the exact posterior or joint. □