Appendix

A Proofs of theoretical results

A.1 Theorem 1

Proof. We can explicitly write down the acceptance probability function as:

$$
\begin{aligned}
a_{\theta,\phi}(\mathbf{z}|\mathbf{x},T) &= e^{-[l_{\theta,\phi}(\mathbf{z}|\mathbf{x},T)]^+} \\
&= \frac{e^T p_\theta(\mathbf{x},\mathbf{z})}{e^T p_\theta(\mathbf{x},\mathbf{z}) + q_\phi(\mathbf{z}|\mathbf{x})}
\end{aligned}
$$

From the above equation, it is easy to see that as $T \to \infty$, we get an acceptance probability close to 1, resulting in an approximate posterior close to the original proposal, $q_\phi(\mathbf{z}|\mathbf{x})$, whereas with $T \to -\infty$, the acceptance probability degenerates to a standard rejection sampler with acceptance probability close to $e^T \frac{p_\theta(\mathbf{x},\mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x})}$, but with potentially untenable efficiency. Intermediate values of T can interpolate between these two extremes.

To prove monotonicity, we first derive the partial derivative of the $\mathtt{KL}$ divergence with respect to T as a covariance of two random variables that are monotone transformations of each other. To get the derivative, we use the fact that the gradient of the $\mathtt{KL}$ divergence is the negative of the ELBO gradient derived in Theorem 1. Recall that the ELBO and the KL divergence add up to a constant independent of T, and that the expressions for the gradients with respect to T and ϕ are functionally the same. We have:

$$
\nabla_T \mathtt{KL}(R_{\theta,\phi}(\mathbf{z}|\mathbf{x},T) \| P_\theta(\mathbf{z}|\mathbf{x})) = -\mathtt{COV}_R \left(A(\mathbf{z}), \nabla_T \log \gamma_r(\mathbf{z}) \right),
$$

where:

$$
\begin{aligned}
A(\mathbf{z}) &= \log p_\theta(\mathbf{x},\mathbf{z}) - \log \gamma_r(\mathbf{z}) \\
&= \log p_\theta(\mathbf{x},\mathbf{z}) - \log q_\phi(\mathbf{z}|\mathbf{x}) \\
&+ [\log q_\phi(\mathbf{z}|\mathbf{x}) - \log p_\theta(\mathbf{x},\mathbf{z}) - T]^+ \\
&= [l_{\theta,\phi}(\mathbf{z}|\mathbf{x},T)]^+ - l_{\theta,\phi}(\mathbf{z}|\mathbf{x},T) - T.
\end{aligned}
$$

For the second term in the covariance, we can use the expressions from Eq. (4) and Eq. (6) to write:

$$
\begin{aligned}
\nabla_T \log \gamma_r(\mathbf{z}) &= -\nabla_T [l_{\theta,\phi}(\mathbf{z}|\mathbf{x},T)]^+ \\
&= -\sigma(l_{\theta,\phi}(\mathbf{z}|\mathbf{x},T)) \nabla_T l_{\theta,\phi}(\mathbf{z}|\mathbf{x},T) \\
&= \sigma(l_{\theta,\phi}(\mathbf{z}|\mathbf{x},T)),
\end{aligned}
$$

where $\sigma(x) \triangleq 1/(1+e^{-x})$ is the sigmoid function. Putting the two terms together, we have:

$$
\begin{aligned}
&\nabla_T \mathtt{KL}(R_{\theta,\phi}(\mathbf{z}|\mathbf{x},T) \| P_\theta(\mathbf{z}|\mathbf{x})) = \\
&- \mathtt{COV}_R([l_{\theta,\phi}(\mathbf{z}|\mathbf{x},T)]^+ - l_{\theta,\phi}(\mathbf{z}|\mathbf{x},T) - T,\ \sigma(l_{\theta,\phi}(\mathbf{z}|\mathbf{x},T))).
\end{aligned}
$$

To prove that the two random variables, $[l_{\theta,\phi}(\mathbf{z}|\mathbf{x},T)]^+ - l_{\theta,\phi}(\mathbf{z}|\mathbf{x},T) - T$ and $\sigma(l_{\theta,\phi}(\mathbf{z}|\mathbf{x},T))$ are a monotone transformation of each other, we can use the identity $[x]^+ - x = \log(1+e^x) - x = -\log \sigma(x)$ to rewrite the final expression for the gradient of the $\mathtt{KL}$ divergence as:

$$
\nabla_T \mathtt{KL}(R_{\theta,\phi}(\mathbf{z}|\mathbf{x},T) \| P_\theta(\mathbf{z}|\mathbf{x})) = \mathtt{COV}_R \left(\log \sigma(l_{\theta,\phi}(\mathbf{z}|\mathbf{x},T)) + T,\ \sigma(l_{\theta,\phi}(\mathbf{z}|\mathbf{x},T)) \right)
$$

The inequality follows from the fact that the covariance of a random variable and a monotone transformation (the logarithm in this case) is non-negative. □

A.2 Theorem 2

Before proving Theorem 2, we first state and prove an important lemma[3].

[3] We assume a discrete and finite state space in all proofs below for simplicity/clarity, but when combined with the necessary technical conditions required for the existence of the corresponding integrals, they admit a straightforward replacement of sums with integrals.

Lemma 1. *Suppose $p(x) = \gamma_p(x)/Z_p$ and $r(x) = \gamma_r(x)/Z_R$ are two unnormalized densities, where only R depends on ϕ (the recognition network parameters), but both P and R can depend on θ.[4] Let $\mathcal{A}(x) \triangleq \log \gamma_p(x) - \log \gamma_r(x)$. Then the variational lower bound objective (on $\log Z_P$) and its gradients with respect to the parameters θ, ϕ are given by:*

$$
\begin{aligned}
ELBO(\theta,\phi) &\triangleq \mathbb{E}_R\left[\mathcal{A}(x)\right] + \log Z_R \\
\nabla_\phi ELBO(\theta,\phi) &= \mathtt{COV}_R\left(\mathcal{A}(x), \nabla_\phi \log \gamma_r(x)\right) \\
\nabla_\theta ELBO(\theta,\phi) &= \mathbb{E}_R\left[\nabla_\theta \log \gamma_p(x)\right] \\
&\quad + \mathtt{COV}_R\left(\mathcal{A}(x), \nabla_\theta \log \gamma_r(x)\right).
\end{aligned}
$$

Note that the covariance is the expectation of the product of (at least one) mean-subtracted version of the two random variables. Further, we can also write: $\mathtt{KL}(R\|P) = \log\left(\mathbb{E}_R\left[e^{-\bar{\mathcal{A}}(x)}\right]\right)$, where $\bar{\mathcal{A}}(x) \triangleq \mathcal{A}(x) - \mathbb{E}_R\left[\mathcal{A}(x)\right]$ is the mean subtracted version of the learning signal, $\mathcal{A}(x)$.

Proof. The equation for the ELBO follows from the definition. For the gradients, we can write: $\nabla_\phi ELBO(\theta,\phi) = D_2 - D_1 + D_3$, where:

$$
\begin{aligned}
D_1 &= \nabla_\phi \mathbb{E}_R\left[\log \gamma_r(x)\right] \\
D_2 &= \nabla_\phi \mathbb{E}_R\left[\log \gamma_p(x)\right] \\
D_3 &= \nabla_\phi \log Z_R
\end{aligned}
$$

Simplifying D_1, D_2, and D_3, we get:

$$
\begin{aligned}
D_1 &= \nabla_\phi \mathbb{E}_R\left[\log \gamma_r(x)\right] \\
&= \sum_x \nabla_\phi\left[r(x) \log \gamma_r(x)\right] \\
&= \sum_x \left(\frac{r(x)}{\gamma_r(x)} \nabla_\phi \gamma_r(x) + \log \gamma_r(x)\ \nabla_\phi r(x)\right) \\
&= \frac{1}{Z_R} \nabla_\phi Z_R + \sum_x r(x) \log \gamma_r(x) \nabla_\phi \log r(x) \\
&= D_3 + \mathbb{E}_R\left[\log \gamma_r(x) \nabla_\phi \log r(x)\right] \\
D_2 &= \nabla_\phi \mathbb{E}_R\left[\log \gamma_p(x)\right] \\
&= \nabla_\phi \sum_x r(x) \log \gamma_p(x) \\
&= \sum_x \log \gamma_p(x) \nabla_\phi r(x) \\
&= \sum_x \log \gamma_p(x) r(x) \nabla_\phi \log r(x) \\
&= \mathbb{E}_R\left[\log \gamma_p(x) \nabla_\phi \log r(x)\right]
\end{aligned}
$$

which implies:

$$
\begin{aligned}
\nabla_\phi ELBO(\theta,\phi) &= D_2 - (D_1 - D_3) \\
&= \mathbb{E}_R\left[\left(\log \gamma_p(x) - \log \gamma_r(x)\right) \nabla_\phi \log r(x)\right].
\end{aligned}
$$

Next, observe that $\mathbb{E}_R\left[\nabla_\phi \log r(x)\right] = 0$, Therefore, using the fact that the expectation of the product of two random variables is the same as their covariance when at least one of the two random variables has a zero mean, we

[4] The dependence for R on θ can happen via some resampling mechanism that is allowed to, for example, evaluate γ_p on the sample proposals before making its accept/reject decisions, as in our case.

get $\nabla_\phi ELBO(\theta, \phi) = \texttt{COV}_R(\mathcal{A}(x), \nabla_\phi \log r(x))$. Next note that we can add an arbitrary constant to either random variable without changing the covariance, therefore this is equal to $\texttt{COV}_R(\mathcal{A}(x), \nabla_\phi \log r(x) - \nabla_\phi \log Z_R) = \texttt{COV}_R(\mathcal{A}(x), \nabla_\phi \log \gamma_r(x))$.

The derivation for the gradient with respect to θ is analogous, except for D_2, which has an additional term $\mathbb{E}_R[\nabla_\theta \log \gamma_p(x)]$ which did not appear in the gradient with respect to ϕ because of our assumption on the lack of dependence of $\log \gamma_p(x)$ on the recognition parameters ϕ. For the identity on the KL divergence, we have:

$$\begin{aligned}
\texttt{KL}(R \| P) &= \log Z_P - \log Z_R + \mathbb{E}_R[\log \gamma_r(x) - \log \gamma_p(x)] \\
&= \log\left(\sum_x \frac{\gamma_p(x)}{Z_R}\right) + \mathbb{E}_R[\log \gamma_r(x) - \log \gamma_p(x)] \\
&= \log\left(\mathbb{E}_R\left[\frac{\gamma_p(x)}{\gamma_r(x)}\right]\right) + \mathbb{E}_R[\log \gamma_r(x) - \log \gamma_p(x)] \\
&= \log\left(\mathbb{E}_R\left[e^{-\mathcal{A}(x)}\right]\right) + \mathbb{E}_R[\mathcal{A}(x)] \\
&= \log\left(\mathbb{E}_R\left[e^{-\bar{\mathcal{A}}(x)}\right]\right).
\end{aligned}$$

□

Using the above lemma, we provide a proof for Theorem 2 below.

Proof. We apply the result of Theorem 1, which computes the ELBO corresponding to the two unnormalized distributions on the latent variable space $\mathbf{z}$ (for fixed $\mathbf{x}, T$), with $\log \gamma_p(.) \triangleq \log p_\theta(\mathbf{z}, \mathbf{x})$ and $\log \gamma_r(.) \triangleq \log q_\phi(\mathbf{z}|\mathbf{x}) - [l_{\theta,\phi}(\mathbf{z}|\mathbf{x}, T)]^+$. This gives: $\nabla_\phi\text{R-ELBO}(\theta, \phi) = \texttt{COV}_R(A_{\theta,\phi}(\mathbf{z}|\mathbf{x}, T), \nabla_\phi \log \gamma_r(\mathbf{z}))$. We can then evaluate $\nabla_\phi \log \gamma_r(\mathbf{z}) = (1 - \sigma(l_{\theta,\phi}(\mathbf{z}|\mathbf{x}, T)))\nabla_\phi \log q_\phi(\mathbf{z}|\mathbf{x})$, where $\sigma()$ is the sigmoid function. Note that this is a consequence of the fact that the derivative of the softplus, $\log(1 + e^x)$, is the sigmoid function, $1/(1 + e^{-x})$. Similarly for the θ gradient, we get:

$$\begin{aligned}
\nabla_\theta\text{R-ELBO}(\theta, \phi) = \mathbb{E}_Q[\nabla_\theta \log p_\theta(\mathbf{x}, \mathbf{z})] \\
+ \texttt{COV}_R(A_{\theta,\phi}(\mathbf{z}|\mathbf{x}, T), \nabla_\theta \log \gamma_r(\mathbf{z}))
\end{aligned}$$

where:

$$\begin{aligned}
\nabla_\theta \log \gamma_r(\mathbf{z}) &= \nabla_\theta [l_{\theta,\phi}(\mathbf{z}|\mathbf{x}, T)]^+ \\
&= \sigma(l_{\theta,\phi}(\mathbf{z}|\mathbf{x}, T))\nabla_\theta l_{\theta,\phi}(\mathbf{z}|\mathbf{x}, T) \\
&= -\sigma(l_{\theta,\phi}(\mathbf{z}|\mathbf{x}, T))\nabla_\theta \log p_\theta(\mathbf{x}, \mathbf{z}).
\end{aligned}$$

□

B Experimental details

B.1 Synthetic

To construct the target distribution, we transform a Poisson distribution of rate $\lambda^* > 0$, denoted $\texttt{Poi}(\lambda^*)$ by removing probability mass near 0. More precisely, this transformation forces a negligible uniform mass, $\epsilon \approx 0$, on $0 \leq z < c$. This leaves the distribution unnormalized, although this fact is not particularly relevant for subsequent discussion. The approximate proposal is parameterized as $Q_\phi \triangleq \texttt{Poi}(e^\phi)$, where ϕ is an unconstrained scalar, and denotes a (unmodified) Poisson distribution with the (non-negative) rate parameter, e^ϕ. Note that for $\texttt{Poi}(e^\phi)$ to explicitly represent a small mass on $z < c$ would require $\phi \to \infty$, but this would be a bad fit for points just above c. As a result, $\{Q_\phi\}$ does not contain candidates close to the target distribution in the sense of $\texttt{KL}$ divergence, even while it may be possible to approximate well with a simple resampling modification that transforms the raw proposal Q_ϕ into a better candidate, R.

The target distribution was set with an optimal parameter $\phi^* = \log(10.0)$ (*i.e.*, the rate parameter is 10.0), and $c = 5$. The optimizer used was SGD with momentum using a mass of 0.5. We observed that the gradients were consistently positive while initialized at a parameter setting less than the true value (as shown in the plots) and similarly consistently negative when initialized to a parameter more than the true value (which we did not present in the paper) due to the consistency of the correlation between the two terms in the covariance specific to this toy example. For resampling, plots show results with learning rate set to 0.01 and $T = 50$. For VIMCO, plots show results with learning rate set to 0.005 and $k = 100$.

B.2 MNIST

We consider the standard $50,000/10,000/10,000$ train/validation/test split of the binarized MNIST dataset. For a direct comparison with prior work, both the generative and recognition networks have the same architecture of stochastic layers. No additional deterministic layers were used for SBNs trained using VRS. The batch size was 50, the optimizer used is Adam with a learning rate of $3e-4$. We ran the algorithm for $5,000,000$ steps updating the resampling thresholds after every $100,000$ iterations based on the threshold selection heuristic corresponding to the top γ quantile. We set $S = 5$ for the unbiased covariance estimates for gradients. The lower bounds on the test set are calculated based on importance sampling with 25 samples for IS or resampling with 25 accepted samples for RS.