Enforcing Sparsity on Latent Space for Robust and Explainable Representations Supplementary File

Hanao Li
Stevens Institute of Technology
Hoboken, NJ
hli136@stevens.edu

Tian Han
Stevens Institute of Technology
Hoboken, NJ
than6@stevens.edu

A. Derivation of Equations

A.1. Log Gradient of Spike and Slab Regularization

The spike and slab distribution can be viewed as:

$$p_{ss}(z) = \alpha_1 N(0, \sigma_1^2) + \alpha_2 N(0, \sigma_2^2) \tag{1}$$

$$= \alpha_1 \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{z^2}{2\sigma_1^2}} + \alpha_2 \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{z^2}{2\sigma_2^2}} \tag{2}$$

Let $\alpha_2 = 1 - \alpha_1$ and take the log of both sides, we can obtain:

$$\mathrm{log} p_{ss}(z) = \log(\alpha_1 \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{z^2}{2\sigma_1^2}} + (1-\alpha_1) \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{z^2}{2\sigma_2^2}}) \tag{3}$$

$$= \log(\frac{\alpha_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{z^2}{2\sigma_1^2}} (1 + \frac{\frac{1-\alpha_1}{\sqrt{2\pi}\sigma_2} e^{-\frac{z^2}{2\sigma_2^2}}}{\frac{\alpha_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{z^2}{2\sigma_1^2}}})) \tag{4}$$

$$= \log(\frac{\alpha_1}{\sqrt{2\pi}\sigma_1}) + \log(e^{-\frac{z^2}{2\sigma_1^2}}) + \log(1 + \frac{\frac{1-\alpha_1}{\sqrt{2\pi}\sigma_2} e^{-\frac{z^2}{2\sigma_2^2}}}{\frac{\alpha_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{z^2}{2\sigma_1^2}}}) \tag{5}$$

$$= \log(\frac{\alpha_1}{\sqrt{2\pi}\sigma_1}) - \frac{z^2}{\sigma_1^2} + \log(1 + \frac{(1-\alpha_1)\sigma_1}{\alpha_1\sigma_2} e^{\frac{(\sigma_2^2-\sigma_1^2)z^2}{2\sigma_1^2\sigma_2^2}}) \tag{6}$$

We then take the partial derivative with respect to z and get:

$$\frac{\partial}{\partial z} \log p_{ss}(z) = 0 - \frac{2z}{2\sigma_1^2} + \frac{\frac{(1-\alpha_1)\sigma_1}{\alpha_1\sigma_2} e^{\frac{(\sigma_2^2-\sigma_1^2)z^2}{2\sigma_1^2\sigma_2^2}} \frac{2(\sigma_2^2-\sigma_1^2)z}{2\sigma_1^2\sigma_2^2}}{1 + \frac{(1-\alpha_1)\sigma_1}{\alpha_1\sigma_2} e^{\frac{(\sigma_2^2-\sigma_1^2)z^2}{2\sigma_1^2\sigma_2^2}}} \tag{7}$$

Multiplying $e^{-\frac{(\sigma_2^2-\sigma_1^2)z^2}{2\sigma_1^2\sigma_2^2}}$ on the numerator and denominator of the third term, we can simplify the equation:

$$\frac{\partial}{\partial z} \log p_{ss}(z) = -\frac{z}{\sigma_1^2} + \frac{\frac{(1-\alpha_1)\sigma_1}{\alpha_1\sigma_2} \frac{(\sigma_2^2-\sigma_1^2)z}{\sigma_1^2\sigma_2^2}}{e^{-\frac{(\sigma_2^2-\sigma_1^2)z^2}{2\sigma_1^2\sigma_2^2}} + \frac{(1-\alpha_1)\sigma_1}{\alpha_1\sigma_2}} \tag{8}$$

$$= -\frac{z}{\sigma_1^2} + \frac{R_1 R_2 z}{e^{-R_2 \frac{z^2}{2}} + R_1} \tag{9}$$

where $R_1 = \frac{(1-\alpha_1)\sigma_1}{\alpha_1\sigma_2}$ and $R_2 = \frac{\sigma_2^2-\sigma_1^2}{\sigma_1^2\sigma_2^2}$. With such derivation, we can prevent exponent overflow when training the model.

A.2. Posterior Component

Given a Gaussian mixture model with two components and prior probability of component $p(C_i) = \alpha_i$, we can represent its posterior below:

$$p(C_i|z) = \frac{p(C_i, z)}{p(z)} \tag{10}$$

$$= \frac{p(C_i)p(z|C_i)}{p_{ss}(z)} \tag{11}$$

$$= \frac{\alpha_i N(0, \sigma_i^2)}{\alpha_i N(0, \sigma_i^2) + \alpha_j N(0, \sigma_j^2)} \tag{12}$$

$$= \frac{\alpha_i \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{z^2}{2\sigma_i^2}}}{\alpha_i \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{z^2}{2\sigma_i^2}} + \alpha_j \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{z^2}{2\sigma_j^2}}} \tag{13}$$

$$= \frac{\frac{\alpha_i}{\sigma_i}}{\frac{\alpha_i}{\sigma_i} + \frac{\alpha_j}{\sigma_j} e^{-\frac{z^2}{2\sigma_j^2} + \frac{z^2}{\sigma_i^2}}} \tag{14}$$

$$= \frac{\frac{\alpha_i}{\sigma_i}}{\frac{\alpha_i}{\sigma_i} + \frac{\alpha_j}{\sigma_j} e^{z^2(\frac{1}{2\sigma_i^2} - \frac{1}{2\sigma_j^2})}} \quad i, j \in \{1, 2\}; i \neq j \tag{15}$$

A.3. Log Gradient of Spike and Slab Prior from Posterior Perspective

We start to take the derivative with respect to z without any reparametrization from Equation 3:

$$\frac{\partial}{\partial z} \log p_{ss}(z) \tag{16}$$

$$= \frac{\partial}{\partial z}(\log(\alpha_1 \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{z^2}{2\sigma_1^2}} + (1-\alpha_1)\frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{z^2}{2\sigma_2^2}})) \tag{17}$$

$$= \frac{-\alpha_1 \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{z^2}{2\sigma_1^2}\frac{2z}{2\sigma_1^2}} - (1-\alpha_1)\frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{z^2}{2\sigma_2^2}} \frac{2z}{2\sigma_2^2}}{\alpha_1 \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{z^2}{2\sigma_1^2}} + (1-\alpha_1)\frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{z^2}{2\sigma_2^2}}} \tag{18}$$

$$= -\frac{\alpha_1 \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{z^2}{2\sigma_1^2}\frac{z}{\sigma_1^2}}}{\alpha_1 \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{z^2}{2\sigma_1^2}} + (1-\alpha_1)\frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{z^2}{2\sigma_2^2}}} - \frac{(1-\alpha_1)\frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{z^2}{2\sigma_2^2}} \frac{z}{\sigma_2^2}}{\alpha_1 \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{z^2}{2\sigma_1^2}} + (1-\alpha_1)\frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{z^2}{2\sigma_2^2}}} \tag{19}$$

Substituting with Equation 14, we can have:

$$\frac{\partial}{\partial z} \log p_{ss}(z) = -p(C_1|z)\frac{z}{\sigma_1^2} - p(C_2|z)\frac{z}{\sigma_2^2} \tag{20}$$

B. Additional Experiments

B.1. Analysis of Sparsity Level

We plotted the model's PSNR performance with various sparsity levels on CelebA dataset [2]. As we can see from Figure 1, the PSNR does not drop significantly even with a very sparse latent representation. With the use of maximum likelihood sampling, we are able to achieve better results than the dense short-run model [3].

B.2. Ablation on Maximum Sampling

We also tested the performance using multiple MCMC chains and a single chain with skip steps on MNIST dataset [1]. We tested the PSNR and the cost of time for training each epoch with a batch size equal to 100. From the table, we can see even though running multiple independent MCMC chains can lead to slightly better PSNR, the cost of time is much larger. We showed that running a single chain with K skip steps can be efficient and lead to comparative performance.

B.3. Generation Results

To maintain fair comparison, we adopted the structure from [5] and we have shown that our model can achieve better performance compared to the existing sparse latent variable models. Even though our aim is to learn explainable and robust sparse latent representations rather than focusing on the image generation, we can still incorporate the diffusion model and follow the procedure in [4] to produce sharp generation results to demonstrate our model's generalization ability on larger models. From Figure 2, we can see our model can generate sharp images and can benefit from modern architecture.

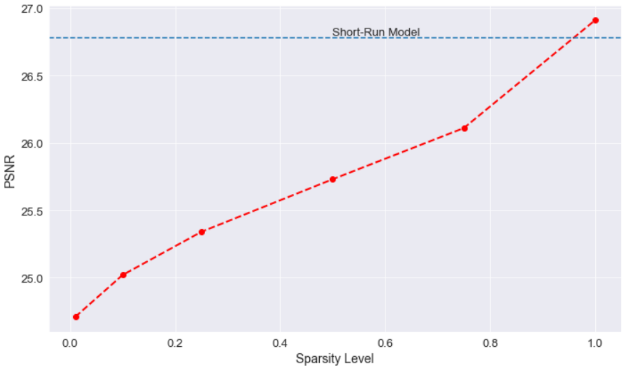

Figure 1. Different sparsity level vs PSNR.

Model	K = 3		K = 5		K = 7	
	PSNR	Time (s)	PSNR	Time (s)	PSNR	Time (s)
Multiple Chains	19.44	50.58	19.82	72.16	19.90	104.91
Single Chain	18.56	22.31	19.77	30.44	19.84	36.12

Table 1. PSNR and cost of time on ablation models.

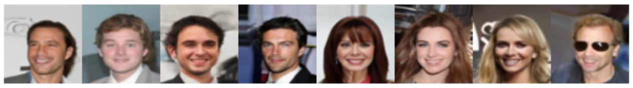

Figure 2. Generated Images from CelebA.

References

[1] Yann LeCun, Corinna Cortes, and CJ Burges. Mnist handwritten digit database, 2010. 2

[2] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015. 2

[3] Erik Nijkamp, Bo Pang, Tian Han, Linqi Zhou, Song-Chun Zhu, and Ying Nian Wu. Learning multi-layer latent variable model via variational optimization of short run MCMC for approximate inference. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK,*

August 23-28, 2020, Proceedings, Part VI, volume 12351 of *Lecture Notes in Computer Science*, pages 361–378. Springer, 2020. 2

[4] Kushagra Pandey, Avideep Mukherjee, Piyush Rai, and Abhishek Kumar. Diffusevae: Efficient, controllable and high-fidelity generation from low-dimensional latents. *arXiv preprint:2201.00308*, 2022. 2

[5] Francesco Tonolini, Bjørn Sand Jensen, and Roderick Murray-Smith. Variational sparse coding. In Amir Globerson and Ricardo Silva, editors, *Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence, UAI 2019, Tel Aviv, Israel, July 22-25, 2019*, volume 115 of *Proceedings of Machine Learning Research*, pages 690–700. AUAI Press, 2019. 2