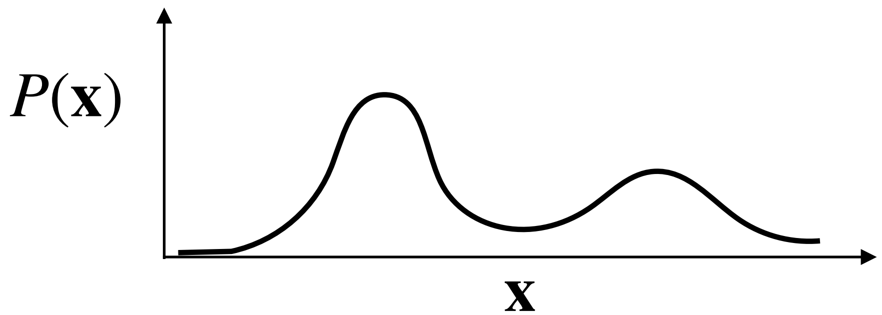
Gaussian-Bernoulli RBMs Without Tears

Renjie Liao*[1], Simon Kornblith[2], Mengye Ren[3], David J. Fleet[2,4,5], Geoffrey Hinton[2,4,5]
University of British Columbia[1], Google Research, Brain Team[2],
New York University[3], University of Toronto[4], Vector Institute[5]
`rjliao@ece.ubc.ca, mengye@cs.nyu.edu`
`{skornblith, davidfleet, geoffhinton}@google.com`

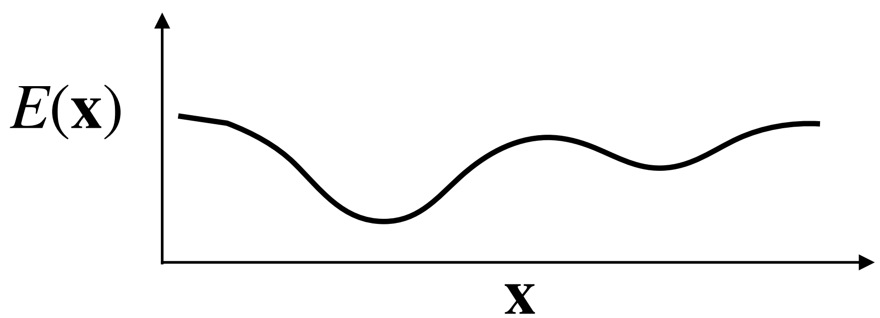
Abstract

We revisit the challenging problem of training Gaussian-Bernoulli restricted Boltzmann machines (GRBMs), introducing two innovations. We propose a novel Gibbs-Langevin sampling algorithm that outperforms existing methods like Gibbs sampling. We propose a modified contrastive divergence (CD) algorithm so that one can generate images with GRBMs starting from noise. This enables direct comparison of GRBMs with deep generative models, improving evaluation protocols in the RBM literature. Moreover, we show that modified CD and gradient clipping are enough to robustly train GRBMs with large learning rates, thus removing the necessity of various tricks in the literature. Experiments on Gaussian Mixtures, MNIST, FashionMNIST, and CelebA show GRBMs can generate good samples, despite their single-hidden-layer architecture. Our code is released at: `https://github.com/lrjconan/GRBM`

The neural sampling hypothesis: *exploit intrinsic variability to sample from probability distributions*.



$$P(\mathbf{x}) = \frac{1}{Z} e^{-E(\mathbf{x})/T}$$



Langevin dynamics:

$$\tau \dot{\mathbf{x}} = -\nabla_{\mathbf{x}} E(\mathbf{x}) + \sqrt{2T\tau}\,\mathbf{n} \quad \Rightarrow \quad \mathbf{x} \sim P(\mathbf{x})$$

- **Build analog electronics to implement** $\tau \dot{\mathbf{x}} = -\nabla_{\mathbf{x}} E(\mathbf{x})$ **.**
- **Add noise** $\mathbf{n} \sim \mathcal{N}(0, \mathbf{I})$**.**

$\Rightarrow$

- Noisy (low power) electronics can be exploited to do probabilistic inference.
- Deterministic, high-precision circuits are wasted on most problems of AI which involve reasoning under uncertainty.

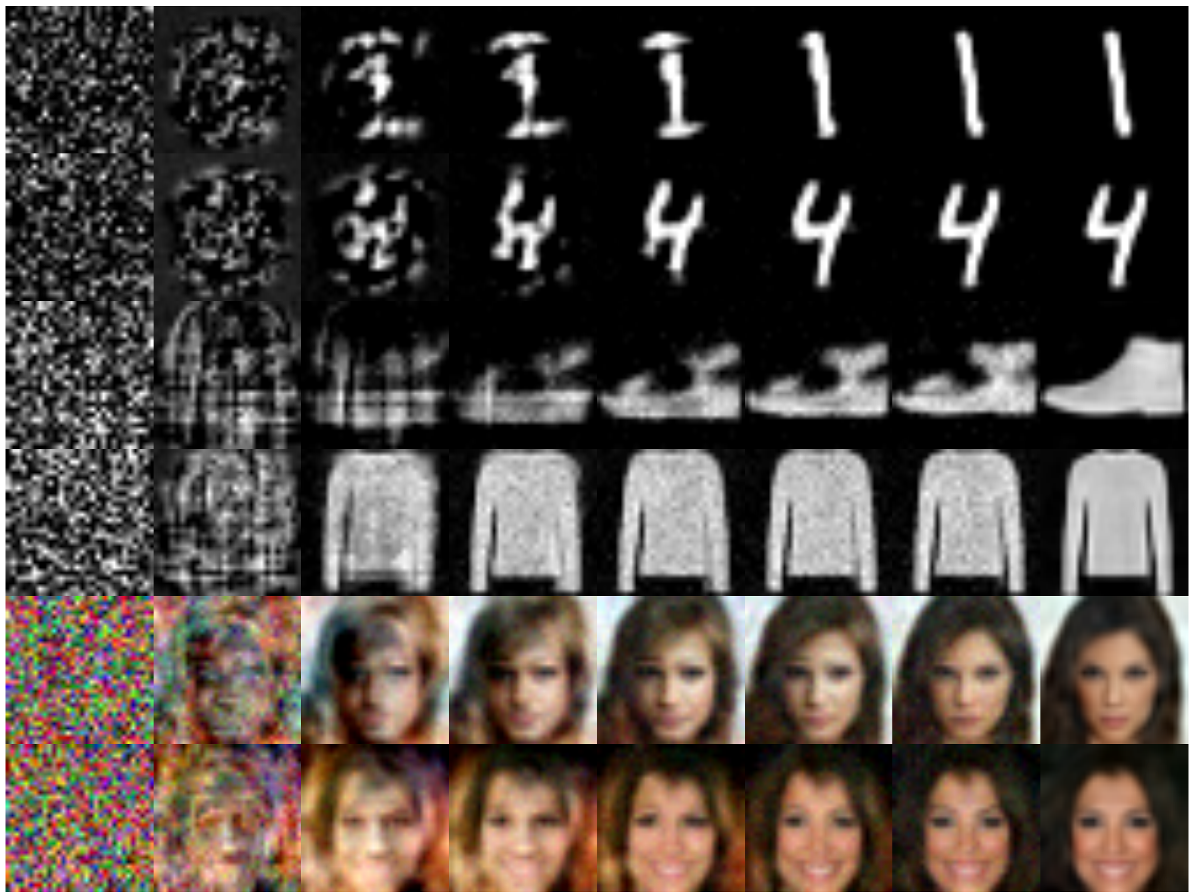
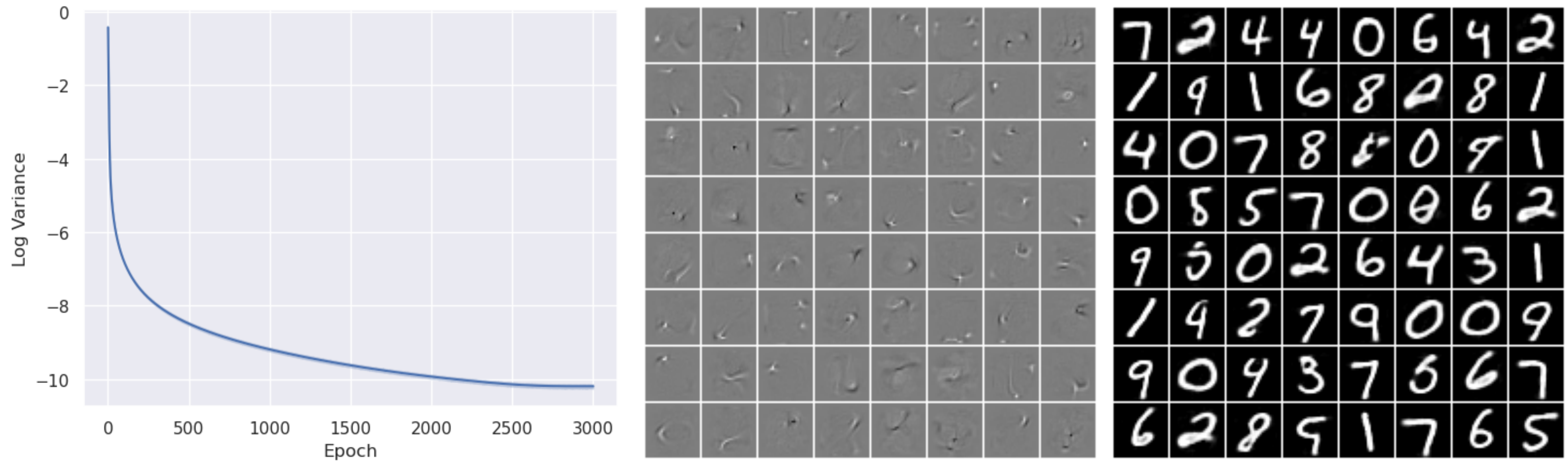


Figure 2: Intermediate samples from Gibbs-Langevin sampling.

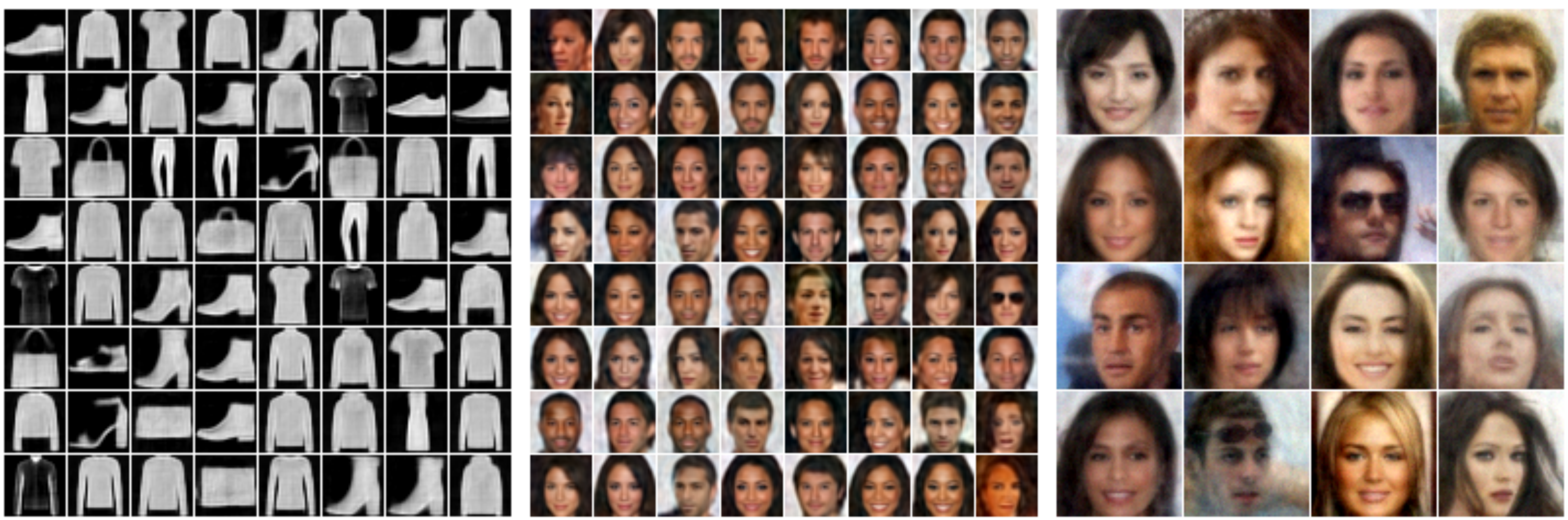
Methods	FID
VAE	16.13
2sVAE (Dai & Wipf, 2019)	12.60
PixelCNN++ (Salimans et al.)	11.38
WGAN (Arjovsky et al., 2017)	10.28
NVAE (Vahdat & Kautz, 2020)	7.93
GRBMs	
Gibbs	47.53
Langevin wo. Adjust	43.80
Langevin w. Adjust	41.24
Gibbs-Langevin wo. Adjust	17.49
Gibbs-Langevin w. Adjust	19.27

Table 1: Results on MNIST dataset.



(a) (b) (c)

Figure 3: (a) Learning curve of (natural) log variances, (b) learned filters, and (c) samples on MNIST.



(a) (b) (c)

Figure 4: Samples from GRBMs on (a) FashionMNIST, (b) CelebA-32, and (c) CelebA-2K-64.

Learning and Inference in Sparse Coding Models With Langevin Dynamics

Michael Y.-S. Fang
michaelfang@berkeley.edu
Department of Physics, University of California, Berkeley, Berkeley, CA 94720, U.S.A., and Redwood Center for Theoretical Neuroscience, University of California, Berkeley, Berkeley, CA 94720, U.S.A.

Mayur Mudigonda
mudigonda@berkeley.edu
Redwood Center for Theoretical Neuroscience, University of California, Berkeley, Berkeley, CA 94720, U.S.A.

Ryan Zarcone
ryanzarcone@gmail.com
Redwood Center for Theoretical Neuroscience, University of California, Berkeley, Berkeley, CA 94720, U.S.A., and Biophysics Graduate Group, University of California, Berkeley, Berkeley, CA 94720, U.S.A.

Amir Khosrowshahi
khosra@gmail.com
Redwood Center for Theoretical Neuroscience, University of California, Berkeley, Berkeley, CA 94720, U.S.A., and Intel Corporation, Santa Clara, CA 95054, U.S.A.

Bruno A. Olshausen
baolshausen@berkeley.edu
Redwood Center for Theoretical Neuroscience, University of California, Berkeley, Berkeley, CA, U.S.A., and Helen Wills Neuroscience Institute and School of Optometry, University of California, Berkeley, Berkeley, CA 94720, U.S.A.