Low-precision Sampling for Probabilistic Deep Learning

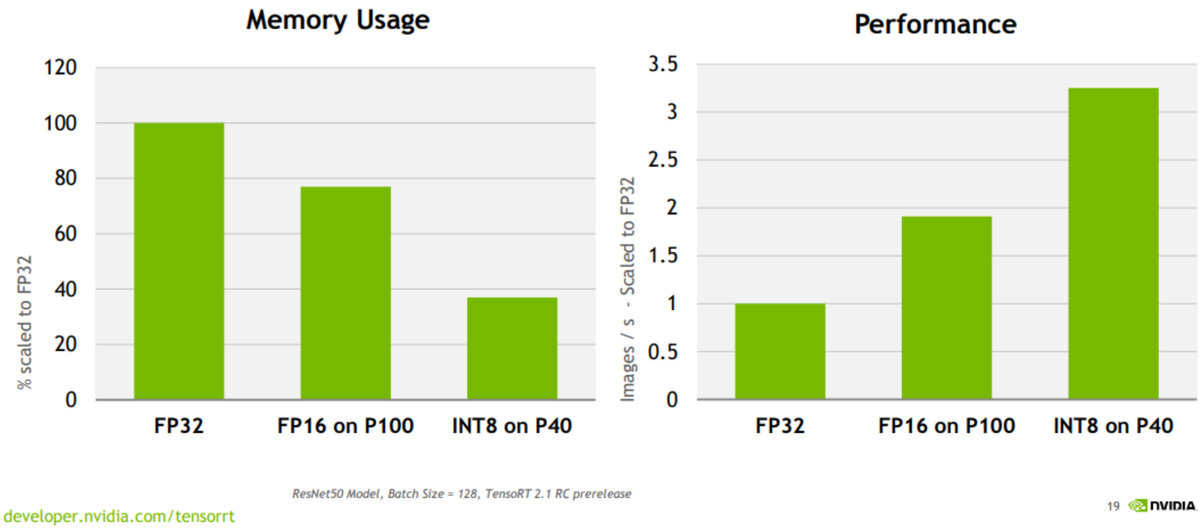
Ruqi Zhang

Department of Computer Science, Purdue University

MLNCP Workshop @ NeurIPS 2023

Low-precision Optimization

- Use fewer bits to represent numbers, e.g. 8 bits
- Significantly reduce memory and latency consumption



- Supported on new emerging chips including CPUs, GPUs, TPUs

Low-precision sampling remains largely unexplored

Sampling: obtain samples from a target distribution $\pi(\theta)$

The task of sampling is ubiquitous in ML

- Probabilistic inference: $\pi(\theta)$ is a parameter distribution (e.g. the posterior of deep neural network weights)
- Generative modeling: $\pi(\theta)$ is a data distribution (e.g. energy-based models, diffusion models)
- Representation learning: $\pi(\theta)$ is a latent variable distribution (e.g. restricted Boltzmann machine)
- ……

Low-precision sampling remains largely unexplored

Stochastic gradient decent (SGD)

$$\theta_{k+1} = \theta_k - \alpha \nabla \tilde{U}(\theta_k)$$

Stochastic gradient Langevin dynamics (SGLD)

$$\theta_{k+1} = \theta_k - \alpha \nabla \tilde{U}(\theta_k) + \sqrt{2\alpha}\xi_{k+1}$$

where $\xi_{k+1} \sim \mathcal{N}(0, I)$

Our work: first comprehensive study for low-precision SGLD

Pros of sampling:

✓ Characterize complex and multi-modal DNN loss landscape

✓ Provide state-of-the-art generalization accuracy and calibration

✓ **Robust to system noise, especially suited for low-precision!**

Bayesian Learning via Stochastic Gradient Langevin Dynamics. M Welling and Y Teh. ICML 2011

Low-precision SGLD

The update of SGLD is

$$\theta_{k+1} = \theta_k - \alpha \nabla \tilde{U}(\theta_k) + \sqrt{2\alpha}\xi_{k+1}, \quad \xi_{k+1} \sim \mathcal{N}(0, I)$$

Apply quantization functions

Key questions:

- *How to apply quantization to the update of SGLD?*
- *How will low-precision arithmetic affect the convergence?*
- *How does low-precision SGLD compare with low-precision SGD?*

Low-Precision Stochastic Gradient Langevin Dynamics. R Zhang, A Wilson, C De Sa. ICML 2022

SGLD with full-precision gradient accumulators

SGLDLP-F: a full-precision weight buffer to accumulate gradient updates

$$\theta_{k+1} = \theta_k - \alpha Q_G\left(\nabla\tilde{U}(Q_W(\theta_k))\right) + \sqrt{2\alpha}\xi_{k+1}$$

Theorem (informal): On strongly log-concave distributions, *SGLDLP-F converges under the 2-Wasserstein distance to the target distribution*

Takeaway

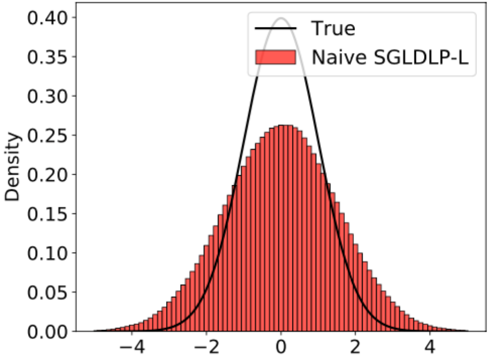
- SGLDLP-F is convergent and can be safely used in practice
- SGLDLP-F is more robust to the quantization error than its SGD counterpart

SGLD with low-precision gradient accumulators

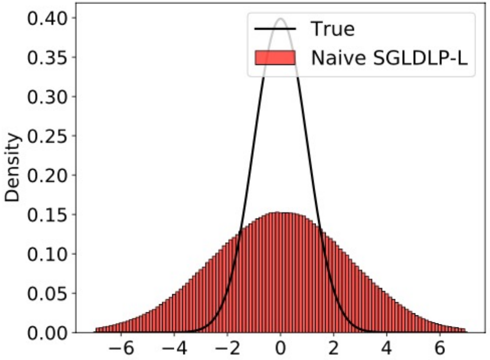
SGLDLP-L: the weight is always represented in low-precision

$$\theta_{k+1} = Q_W\left(\theta_k - \alpha Q_G\left(\nabla \tilde{U}(\theta_k)\right) + \sqrt{2\alpha}\xi_{k+1}\right)$$

☹ **Theorem** (informal): As the stepsize decreases, the W2 *distance between the stationary distribution of SGLDLP-L and the target distribution may increase*



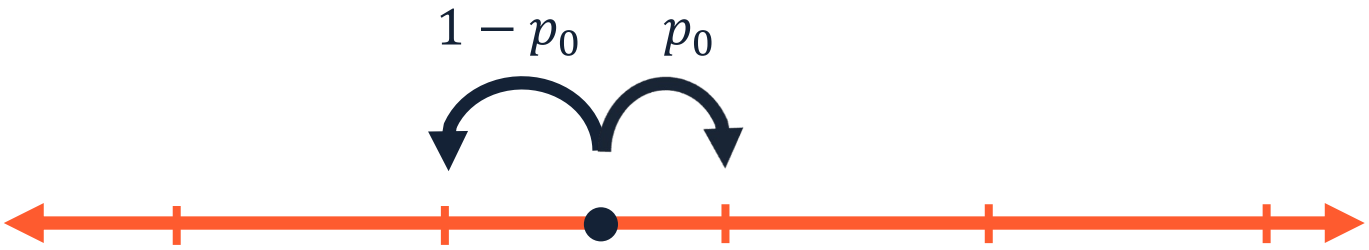
Stepsize = 0.001



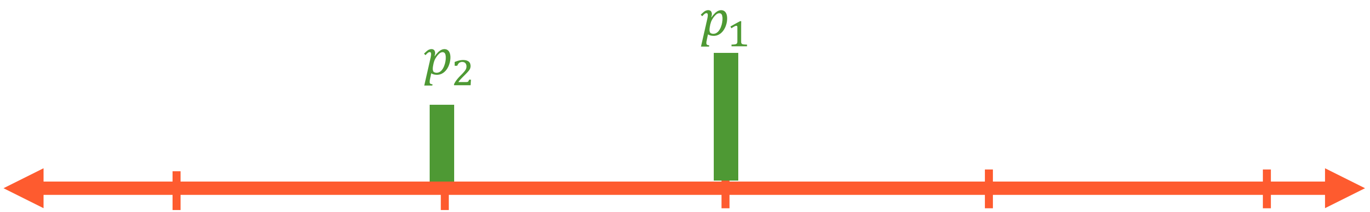
Stepsize = 0.0001

Variance-Corrected Quantization Function

- Reason: wrong variance of $\theta_{k+1} = Q_W\left(\theta_k - \alpha Q_G\left(\nabla \tilde{U}(\theta_k)\right) + \sqrt{2\alpha}\xi_{k+1}\right)$

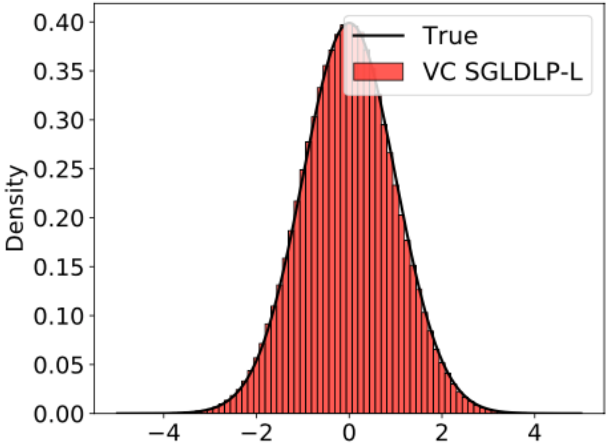


- Our solution: new quantization function to keep correct mean and variance

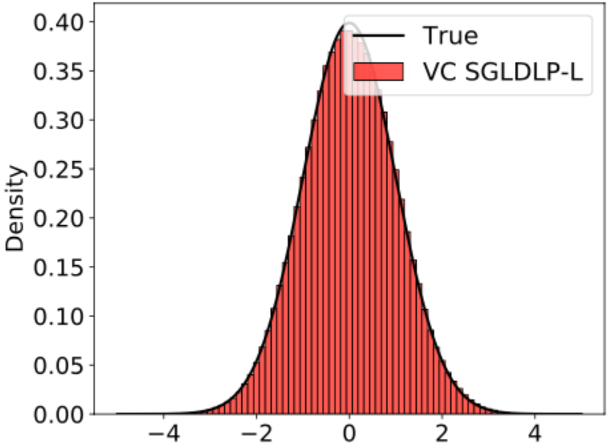


Variance-Corrected SGLD (VC SGLDLP-L)

☺ **Theorem** (informal): On strongly log-concave distributions, VC *SGLDLP-L converges under the 2-Wasserstein distance to the target distribution*



Stepsize = 0.001



Stepsize = 0.0001

Experimental Results: Prediction and Calibration

Table 1. Test errors (%).

	CIFAR-10	CIFAR-100	IMDB
32-BIT FLOATING POINT			
SGLDFP	4.65 ±0.06	22.58 ±0.18	13.43 ±0.21
SGDFP	4.71 ±0.02	22.64 ±0.13	13.88 ±0.29
cSGLDFP	4.54 ±0.05	21.63 ±0.04	13.25 ±0.18
8-BIT FIXED POINT			
NAÏVE SGLDLP-L	7.82 ±0.13	27.25 ±0.13	16.63 ±0.28
VC SGLDLP-L	7.13 ±0.01	26.62 ±0.16	15.38 ±0.27
SGDLP-L	8.53 ±0.08	28.86 ± 0.10	19.28 ±0.63
SGLDLP-F	5.12 ±0.06	23.30 ±0.09	15.40 ±0.36
SGDLP-F	5.20 ±0.14	23.84 ±0.12	15.74 ±0.79
8-BIT BLOCK FLOATING POINT			
NAÏVE SGLDLP-L	5.85 ±0.04	26.38 ±0.13	14.64 ±0.08
VC SGLDLP-L	5.51 ±0.01	25.22 ±0.18	13.99 ±0.24
SGDLP-L	5.86 ±0.18	26.19 ±0.11	16.06 ±1.81
SGLDLP-F	4.58 ±0.07	22.59 ±0.18	14.05 ±0.33
SGDLP-F	4.75 ±0.05	22.9 ±0.13	14.28 ±0.17
VC cSGLDLP-L	4.97 ±0.10	22.61 ±0.12	13.09 ±0.27
cSGLD-F	4.32 ±0.07	21.50 ±0.14	13.13 ±0.37

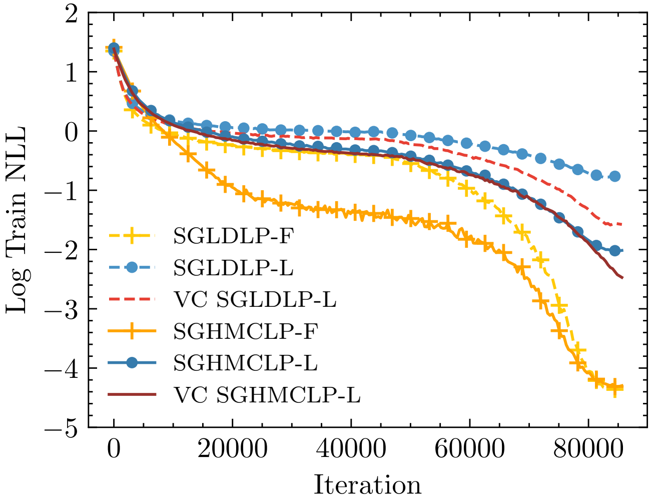
Table 2. ECE ↓ (%).

	CIFAR-10	CIFAR-100
32-BIT FLOATING POINT		
SGLD	1.11	3.92
SGD	2.53	4.97
cSGLDFP	0.66	1.38
8-BIT FIXED POINT		
VC SGLDLP-L	0.6	3.19
SGDLP-L	3.4	10.38
SGLDLP-F	1.12	4.42
SGDLP-F	3.05	6.80
8-BIT BLOCK FLOATING POINT		
VC SGLDLP-L	0.6	5.82
SGDLP-L	4.23	12.97
SGLDLP-F	1.19	3.78
SGDLP-F	2.76	5.2
VC cSGLDLP-L	0.51	1.39
cSGLD-F	0.56	1.33

- Low-precision SGLD matches full-precision SGLD with only 8 bits
- Significantly outperforms low-precision SGD

Low-precision Hamiltonian Monte Carlo

	Gradient Complexity	Achieved 2-Wasserstein
Low-precision SGLD	$\tilde{\mathcal{O}}\left(\epsilon^{-4}\lambda^{*-1}\log^{5}\left(\epsilon^{-1}\right)\right)$	$\tilde{\mathcal{O}}\left(\epsilon+\log\left(\epsilon^{-1}\right)\sqrt{\Delta}\right)$
Low-precision SGHMC	$\tilde{\mathcal{O}}\left(\epsilon^{-2}\mu^{*-2}\log^{2}\left(\epsilon^{-1}\right)\right)$	$\tilde{\mathcal{O}}\left(\epsilon+\sqrt{\log\left(\epsilon^{-1}\right)\Delta}\right)$



Low-precision SGHMC converges faster than low-precision SGLD

Enhancing Low-Precision Sampling via Stochastic Gradient Hamiltonian Monte Carlo.
Z Wang, Y Chen, Q Song, R Zhang. Arxiv 2023

Takeaways

- Sampling is especially compatible with low-precision arithmetic due to its inherent randomness
- Low-precision sampling is convergent and is more robust to quantization noise than SGD counterpart
- We develop low-precision Langevin dynamics, low-precision Hamiltonian dynamics, and a new quantization function

Other new compute paradigms in probabilistic DL

- Binary neural networks

  A Langevin-like Sampler for Discrete Distributions.

  R Zhang, X Liu, Q Liu. ICML 2022

- Calibrated sparse neural networks

  Calibrating the Rigged Lottery: Making All Tickets Reliable.

  B Lei, R Zhang, D Xu, B Mallick. ICLR 2023

- Sparse Bayesian neural networks

  Training Bayesian Neural Networks with Sparse Subspace Variational Inference.

  J Li, Z Miao, Q Qiu, R Zhang. NeurIPS Workshop 2023

Thank you!