Selected References on Stochastic Rounding

What is Stochastic Rounding?

> N. Higham, "What Is Stochastic Rounding?," *Nick Higham*, Jul. 07, 2020. https://nhigham.com/2020/07/07/what-is-stochastic-rounding/.

Stochastic rounding: implementation, error analysis and application

Stochastic rounding (SR) randomly maps a real number x to one of the two nearest values in a finite precision number system. The probability of choosing either of these two numbers is 1 minus their relative distance to x. … If used to compute the inner product of two vectors of length n in floating-point arithmetic, it yields an error bound with constant sqrt(n)*u with high probability, where u is the unit round-off. This is not necessarily the case for round to nearest (RN), for which the worst-case error bound has constant nu. A particular attraction of SR is that, unlike RN, it is immune to the phenomenon of stagnation, whereby a sequence of tiny updates to a relatively large quantity is lost. We survey SR .. with a focus on machine learning and the numerical solution of differential equations.

> M. Croci, M. Fasi, N. J. Higham, T. Mary, and M. Mikaitis, "Stochastic rounding: implementation, error analysis and applications," *Royal Society Open Science*, vol. 9, no. 3, p. 211631, Mar. 2022, doi: 10.1098/rsos.211631.

The Positive Effects of Stochastic Rounding in Numerical Algorithms

Recently, stochastic rounding (SR) has been implemented in specialized hardware but most current computing nodes do not yet support this rounding mode. Several works empirically illustrate the benefit of stochastic rounding in various fields such as neural networks .. . For some algorithms, such as summation, inner product or matrix vector multiplication, it has been proved that SR provides probabilistic error bounds better than the traditional deterministic bounds.

> E.-M. El Arar, D. Sohier, P. de Oliveira Castro, and E. Petit, "The Positive Effects of Stochastic Rounding in Numerical Algorithms," in *29th IEEE Symposium on Computer Arithmetic ARITH 2022*, Sep. 2022. https://hal.science/hal-03716058

Deep Learning with Limited Numerical Precision

Training of large-scale deep neural networks is often constrained by the available computational resources. We study the effect of limited precision data representation and computation on neural network training. Within the context of low-precision fixed-point computations, we observe the rounding scheme to play a crucial role in determining the network's behavior during training. Our results show that deep networks can be trained using only 16-bit wide fixed-point number representation when using stochastic rounding, and incur little to no degradation in the classification accuracy. …

> S. Gupta, A. Agrawal, K. Gopalakrishnan, and P. Narayanan, "Deep Learning with Limited Numerical Precision." arXiv, Feb. 09, 2015. doi: 10.48550/arXiv.1502.02551.

Stochastic Rounding: Algorithms and Hardware Accelerator

Algorithms and a hardware accelerator for performing stochastic rounding (SR) are presented. The main goal is to augment the ARM M4F based multi-core processor SpiNNaker2 with a more flexible rounding functionality than is available in the ARM processor itself. The motivation of adding such an accelerator in hardware is based on our previous results showing improvements in numerical accuracy of ODE solvers in fixed point arithmetic with SR, compared to standard round to nearest or bit truncation rounding modes. Furthermore, performing SR purely in software can be expensive, due to requirement of a pseudorandom number generator (PRNG), multiple masking and shifting instructions, and an addition operation. Also, saturation of the rounded values is included, since rounding is usually followed by saturation, which is especially important in fixed point arithmetic due to a narrow dynamic range of representable values. …

> M. Mikaitis, "Stochastic Rounding: Algorithms and Hardware Accelerator." arXiv, Jun. 29, 2020. http://arxiv.org/abs/2001.01501

ESRU: Extremely Low-Bit and Hardware-Efficient Stochastic Rounding

Stochastic rounding is crucial in the low-bit (e.g., 8-bit) training of deep neural networks (DNNs) to achieve high accuracy. One of the drawbacks of prior studies is that they require a large number of high-precision stochastic rounding units (SRUs) to guarantee low-bit DNN accuracy, which involves considerable hardware overhead. In this paper, we use extremely low-bit SRUs (ESRUs) to save a large number of hardware resources during low-bit DNN training.

> S.-E. Chang *et al.*, "ESRU: Extremely Low-Bit and Hardware-Efficient Stochastic Rounding Unit Design for Low-Bit DNN Training," in *2023 Design, Automation & Test in Europe (DATE)*, Belgium: IEEE, pp. 1–6. doi: 10.23919/DATE56975.2023.10137222.

Stochastic Rounding Variance and Probabilistic Bounds: A New Approach

Stochastic rounding (SR) is an idea proposed in the 1950s by von Neumann and Goldstine [28]. First, it can be used to estimate empirically the numerical error of computer programs; SR introduces a random noise in each floating-point operation and then a statistical analysis of the set of sampled outputs can be applied to estimate the effect of rounding errors. To make this simulation available, various tools such as Verificarlo [13], Verrou [15] and Cadna [22] have been developed. Second, SR can be used as a replacement for the default deterministic rounding mode in numerical computations. It has been demonstrated that in multiple domains such as neural networks, ODEs, PDEs, and Quantum mechanics [9], SR provides better results compared to the IEEE-754 default rounding mode [3]. Connolly et al. [8] show that SR successfully prevents the phenomenon of stagnation that takes place in various applications such as neural networks, ODEs and PDEs. In particular, Gupta et al show in [16] that deep neural networks are prone to stagnation during the training phase. … .

> E.-M. E. Arar, D. Sohier, P. de O. Castro, and E. Petit, "Stochastic rounding variance and probabilistic bounds" arXiv, Jun. 05, 2023. doi: 10.48550/arXiv.2207.10321.

Additional references

How the way we're taught to round numbers in school falls short

R. Crowell, "How the way we're taught to round numbers in school falls short," *Science News*, Mar. 22, 2022. https://www.sciencenews.org/article/rounding-numbers-stochastic-machine-learning-quantum-computing

Algorithms for Stochastically Rounded Elementary Arithmetic Operations

We present algorithms for performing [these] arithmetic operations (+, −, ×, ÷, and √) in floating point arithmetic with stochastic rounding, and demonstrate the value of these algorithms by discussing various applications where stochastic rounding is beneficial …

> M. Fasi and M. Mikaitis, "Algorithms for Stochastically Rounded Elementary Arithmetic Operations in IEEE 754 Floating-Point Arithmetic," *IEEE Trans. Emerg. Topics Comput.*, vol. 9, no. 3, pp. 1451–1466, Jul. 2021, doi: 10.1109/TETC.2021.3069165.

On the Convergence of the Gradient Descent Method with Stochastic Fixed-point Rounding

When training neural networks with low-precision computation, rounding errors often cause stagnation or are detrimental to the convergence of the optimizers; in this paper we study the infuence of rounding errors on the convergence of the gradient descent method for problems satisfying the Polyak Lojasiewicz inequality. Within this context, we show that .. biased stochastic rounding errors may be beneficial since choosing a proper rounding strategy eliminates the vanishing gradient problem and forces the rounding bias in a descent direction.

> L. Xia, M. E. Hochstenbach, and S. Massei, "On the Convergence of the Gradient Descent Method with Stochastic Fixed-point Rounding Errors under the Polyak-Lojasiewicz Inequality." arXiv, Jan. 23, 2023. doi: 10.48550/arXiv.2301.09511.

Effects of round-to-nearest and stochastic rounding .. in low precision

We have investigated the behavior of local and global rounding errors arising from the solution of the heat equation in low precision using both round-to-nearest and stochastic rounding. … we explained why round-to-nearest computations are always affected by stagnation as delta t decreases. … we showed that stochastic rounding successfully prevents stagnation and always approximates the exact solution. Furthermore, the dominant local error terms arising from stochastic rounding have favorable statistical properties …

> M. Croci and M. B. Giles, "Effects of round-to-nearest and stochastic rounding in the numerical solution of the heat equation in low precision," *IMA Journal of Numerical Analysis*, vol. 43, no. 3, pp. 1358–1390, Jun. 2023, doi: 10.1093/imanum/drac012

Faithfully Rounded Floating-Point Arithmetic

M. Lange and S. M. Rump, "Faithfully Rounded Floating-point Computations," *ACM Trans. Math. Softw.*, vol. 46, no. 3, pp. 1–20, Sep. 2020, doi: 10.1145/3290955