Massachusetts Institute of Technology

NAP: Noise-Based Sensitivity Analysis for Programs

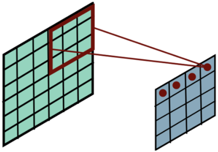
Jesse Michel*, ***Sahil Verma****, *Benjamin Sherman and Michael Carbin*

CSAIL, MIT

June 22, 2019

Approximate computing

- Trade off between speed and accuracy
- Broadly applicable
- NAP gives sensitivity analyses for approximate computing

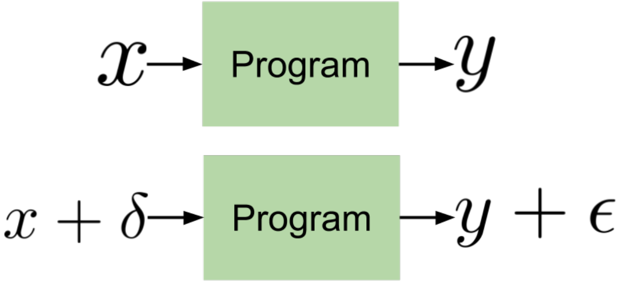


Image/video processing

Energy savings
For data centers

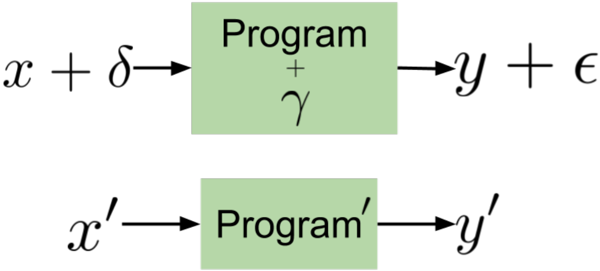
Navigating the space of approximations

- Sensitivity analysis of numerical programs



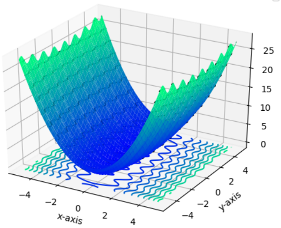
Region-based Sensitivity

- If $\delta, \gamma \sim \mathcal{N}(0, \vec{\sigma}^2)$ then termed noisy programs
- Approximate program from noisy programs



Region-based sensitivities

- Variances of Gaussian distributions
- Compared with ADAPT and Hessian sensitivity [1,3]
- Derivatives may be too local!



The plot of $z = x^2 + \frac{1}{10}y^2 + sin(5y)$

Adding noise

- Quadratic equation

$$f(x) = x^2 + 2x + 1.$$

- Indexing operators gives

$$f(x, \{\times_0, +_1, \times_2, +_3\}) = x \times_0 x +_1 2 \times_2 x +_3 1.$$

- Adding in noise parameters

$$g(x; \vec{\epsilon}) = [(x + \epsilon_x)(x + \epsilon_x) + \epsilon_0] + \epsilon_1 + [2(x + \epsilon_x) + \epsilon_2] + \epsilon_3 + 1.$$

where

$$\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$$

Optimization problem

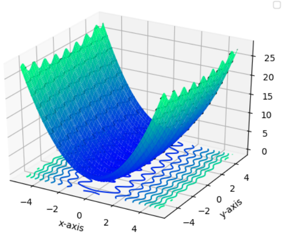
$$\max_{\vec{\sigma}} \Big(\underbrace{\sum_i \log(\sigma_i)}_{\text{noise envelope}} - \lambda \underbrace{\mathop{\mathbb{E}}_{x\sim\mathcal{D},\vec{\epsilon}\sim\mathcal{N}(\vec{0},\vec{\sigma}^2)} \mathcal{L}(x,\vec{\epsilon})}_{\text{minimize error}} \Big)$$

where $x \sim \mathcal{D}$ and the domain-specific loss is

$$\mathcal{L}(x,\vec{\epsilon}) = \Big(g(x;\vec{0}) - g(x;\vec{\epsilon}) \Big)^2$$

Region-based sensitivities

- Variances of Gaussian distributions
- Compared with ADAPT and Hessian sensitivity [1,3]
- Derivatives may be too local!



The plot of $z = x^2 + \frac{1}{10}y^2 + sin(5y)$

Breaking from tradition

Expected **better than** maximum error:

- Improved common case
- Less sensitive to extremes
- Specialize to a data distribution

Expected **worse than** maximum error:

- Not for mission-critical
- More difficult to reason about

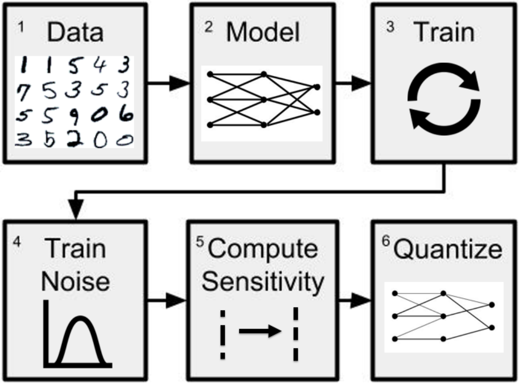
$$\max_{\vec{\sigma}} \Bigg(\underbrace{\sum_i \log(\sigma_i)}_{\text{noise envelope}} - \lambda \underbrace{\mathop{\mathbb{E}}_{x\sim\mathcal{D},\vec{\epsilon}\sim\mathcal{N}(\vec{0},\vec{\sigma}^2)} \mathcal{L}(x, \vec{\epsilon})}_{\text{minimize error}} \Bigg)$$

Comparison with FPTuner/FPTaylor

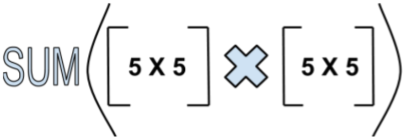
Generated mixed-precision approximate version of numerical programs from FPBench benchmarks.

Benchmarks	FPTuner	RMSE	Mean Bits
verlhulst	3.79e-16	3.72e-16	50
sineOrder3	1.17e-15	7.90e-16	50
predPrey	1.99e-16	1.73e-16	50
sine	8.73e-16	8.34e-17	51
doppler1	1.82e-13	7.75e-14	51
doppler2	3.20e-13	1.07e-13	51
doppler3	1.02e-13	5.10e-14	51
rigidbody1	3.86e-13	1.37e-13	51
sqroot	7.45e-16	4.00e-16	50
rigidbody2	5.23e-11	6.08e-12	51
turbine2	4.13e-14	2.35e-14	50
carbon gas	1.51e-08	3.01e-09	49
turbine1	3.16e-14	1.13e-14	51
turbine3	1.73e-14	1.40e-14	50
jet	2.68e-11	1.07e-11	50

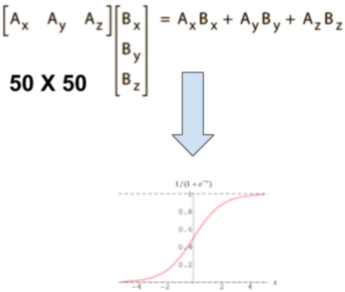
Neural Network Quantization



FPTuner's scalibility



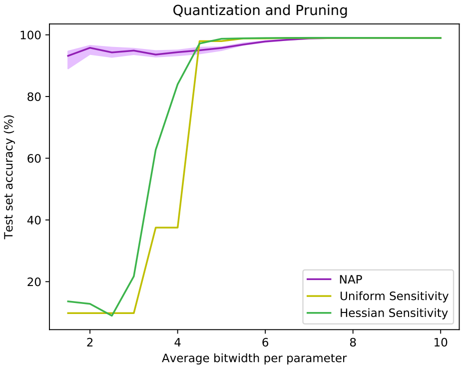
(A)



(B)

FPTuner timed-out after 15 hours in both programs

Space-accuracy tradeoff



$f(\sigma) = \lfloor c - \log(\sigma) \rfloor$ for a constant c

Conclusions

- Broadly applicable sensitivity analysis
- Provides expected error over a data distribution
- Scales well to large computations
- Future work to incorporate generalization bounds.

Harshitha Menon, Michael O. Lam, Daniel Osei-Kuffuor, Markus Schordan, Scott Lloyd, Kathryn Mohror, and Jeffrey Hittinger. Adapt: Algorithmic differentiation applied to floating-point precision tuning. *SC18: International Conference for High Performance Computing, Networking, Storage and Analysis*, 2018. (1)

Wei-Fan Chiang, Mark Baranowski, Ian Briggs, Alexey Solovyev, Ganesh Gopalakrishnan, and Zvonimir Rakamarić. Rigorous floating-point mixed-precision tuning. *SIGPLAN Not.*, 52(1):300–315, January 2017. (2)

Yann Le Cun, John S. Denker, and Sara A. Solla. Optimal brain damage. In *Advances in Neural Information Processing Systems*, pages 598–605. Morgan Kaufmann, 1990. (3)