

NITISH SHIRISH KESKAR

<http://users.iems.northwestern.edu/~nitish>

2145 Sheridan Road ◊ Room C210 ◊ Evanston, IL – 60208

(+1) 805 312 8841 ◊ keskar.nitish@u.northwestern.edu

SUMMARY

Ph.D. candidate with research experience in deep learning, machine learning, statistics and mathematical optimization; actively seeking full-time opportunities starting June, 2017.

EDUCATION

Northwestern University 07/2012 – Present

Ph.D. Candidate, Department of Industrial Engineering and Management Sciences

Advisors: Jorge Nocedal and Andreas Wächter

Northwestern University 07/2012 – 12/2013

M.S., Department of Industrial Engineering and Management Sciences

GPA: 3.86/4

Veermata Jijabai Technological Institute, Mumbai, India 07/2008 – 06/2012

B.Tech., Department of Mechanical Engineering

GPA: 9.3/10

Ranked 1st in graduating class

TECHNICAL SKILLS

Computer Languages Python, C, C++, MATLAB, R, Haskell, Lua, SQL, Fortran, Gurobi.

Software & Tools NumPy, SciPy, Theano, Tensorflow, Torch, scikit-learn, Caffe, Spark, Hadoop.

WORK EXPERIENCE

Research Assistant, Northwestern University, IL 04/2013 – Present

- Currently working on the development of higher-order methods for nonsmooth optimization and deep learning.
- Designed a stochastic quasi-Newton algorithm for training recurrent neural networks.
- Designed and implemented a second-order method for ℓ_1 -regularized convex optimization with applications in machine learning.

Data Center Group Intern, Intel Corporation, CA 06/2016 – 09/2016

- Worked on the design and implementation of highly-scalable training algorithms for deep learning.
- Investigated the causes for the generalization failure of deep learning models trained using large-batch methods.

Summer Intern, IBM T.J. Watson Research Center, NY 06/2014 – 09/2014

- Worked on improving convergence rate of Stochastic Gradient Descent (SGD) for training deep neural networks for automatic speech recognition.
- Proposed and implemented an adaptive non-monotone learning rate strategy to improve convergence and impart robustness to SGD training.

Summer Intern, IISc, Bangalore, India 05/2011 – 07/2011

- Worked on the implementation of Krylov subspace methods for solving dense linear systems on high performance computing environments.
- Conducted extensive numerical experiments to study the convergence behavior for various test cases.

Project Assistant, IIT, Mumbai, India 11/2009 – 01/2010

- Designed and implemented algorithms for the guidance, navigation and control of launch vehicles during the ascent phase.
- Optimized launch vehicle trajectories to achieve significant savings in fuel and time required for the space mission.

PUBLICATIONS

Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M. & Tang, P.T.P. (2016). *On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima*. arXiv preprint arXiv:1609.04836.

Keskar, N. S., & Berahas, A. S. (2016). *adaQN: An Adaptive Quasi-Newton Algorithm for Training RNNs*. European Conference Machine Learning and Knowledge Discovery in Databases, (ECML PKDD 2016), Part I, Vol 9851, 1-16.

Keskar, N. S., Nocedal, J., Oztoprak, F. & Wächter, A. (2015). *A Second-Order Method for Convex ℓ_1 -Regularized Optimization with Active Set Prediction*. Optimization Methods and Software 31.3 (2016): 605-621.

Keskar, N. S., & Saon, G. (2015). *A Nonmonotone Learning Rate Strategy for SGD training of Deep Neural Networks*. In Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on (pp. 4974-4978). IEEE.

CONFERENCES AND INVITED TALKS

“Deep Learning Scalability Wall: On the Failure of Large-Batch Methods for Deep Learning”: Intel Corporation (Santa Clara, USA) 2016.

“adaQN: An Adaptive Quasi-Newton Algorithm for Training RNNs”: Amazon Fall 2015 Graduate Research Symposium (Seattle, USA).

“Quasi-Newton Methods for Nonsmooth Optimization”: INFORMS 2015 (Philadelphia, USA).

“A Second-Order Method for Convex ℓ_1 -Regularized Optimization with Active Set Prediction”: INFORMS 2014 (San Francisco, USA), ISMP 2015 (Pittsburgh, USA), ACNW 2015 (Chicago, USA).

“A Nonmonotone Learning Rate Strategy for SGD training of Deep Neural Networks”: IBM Corporation (Yorktown Heights, USA) 2014, ICASSP 2015 (Brisbane, Australia).

AWARDS AND FELLOWSHIPS

Recipient of the Northwestern University Fellowship for the 2012-2013 academic year to pursue Ph.D.

Recipient of Sir Ratan Tata Merit Scholarship, 2009.

Recipient of the Late C.K. Radhakrishnan Scholarship for securing 1st rank in graduating class, 2010.

TEACHING EXPERIENCE

Industrial Engineering Design (IEMS 393-1): Winter 2015, Winter 2016, Fall 2016, *teaching assistant*.

Robust Optimization (IEMS 490): Spring 2016, *grader*.

Mathematical Programming II (IEMS 450-2): Winter 2014, *grader*.

RELEVANT COURSEWORK

Machine Learning, Statistical Pattern Recognition, Intermediate Statistics, Applied Dynamic Programming, Applied Econometrics, Linear and Nonlinear Optimization, Convex Optimization, Robust Optimization, Integer Programming, Stochastic Simulation.

SERVICE

President, Northwestern University INFORMS Chapter, 2014 – 2015.

Secretary, Northwestern University INFORMS Chapter, 2013 – 2014.