

Paper of the Month, October 2021: Bottou (1998). Stochastic approximations as a mathematical discipline started in 1950's with its origin in computer science and engineering, partly due to the need to overcome the shortage of computing power and data storage. In the age of big data, stochastic approximations have become a dominant approach to parameter optimization for large-scale learning systems. The paper by Bottou (1998) helped popularize the approach in the modern machine learning community. However, its fundamental idea of stochastic gradient descent dates back to Robbins and Monro (1951), and its use of martingale convergence dates back to Gladyshev (1965).

Basic ideas of stochastic approximations in the context of large-scale learning will be discussed. Some attempt will be made to explain why martingale convergence is such a useful tool.

References

- Silvère Bonnabel. Stochastic gradient descent on Riemannian manifolds. *IEEE Trans. Automat. Control*, 58(9):2217–2229, 2013.
- Léon Bottou. Online learning and stochastic approximations. In D Saad, editor, *Online Algorithms and Stochastic Approximations*. Cambridge University Press, Cambridge, U.K., 1998.
- E. G. Gladyshev. On stochastic approximation. *Theory Probab. & Appl.*, 10:297–300, 1965.
- Herbert Robbins and Sutton Monro. A stochastic approximation method. *Ann. Math. Statistics*, 22:400–407, 1951.