Available online at www.sciencedirect.com

SCIENCE DIRECT®

NEUROCOMPUTING

Neurocomputing 68 (2005) 130–160

www.elsevier.com/locate/neucom

Unsupervised learning with stochastic gradient

Harold Szu, Ivica Kopriva*

Department of Electrical and Computer Engineering, George Washington University, 801, 22nd St. NW, Washington, DC 20052, USA

Received 24 February 2004; received in revised form 19 October 2004; accepted 23 November 2004
Available online 25 February 2005
Communicated by T. Heskes

Abstract

A stochastic gradient is formulated based on deterministic gradient augmented with Cauchy simulated annealing capable to reach a global minimum with a convergence speed significantly faster when simulated annealing is used alone. In order to solve space-time variant inverse problems known as blind source separation, a novel Helmholtz free energy contrast function, $H = E - T_0S$, with imposed thermodynamics constraint at a constant temperature T_0 was introduced generalizing the Shannon maximum entropy S of the closed systems to the open systems having non-zero input–output energy exchange E. Here, only the input data vector was known while source vector and mixing matrix were unknown. A stochastic gradient was successfully applied to solve inverse space-variant imaging problems on a concurrent pixel-by-pixel basis with the unknown mixing matrix (imaging point spread function) varying from pixel to pixel.
Published by Elsevier B.V.

Keywords: Stochastic optimization; Cauchy annealing; Blind source separation; Helmholtz free energy

1. Introduction

Gradient optimization is generally incapable of reaching global minimum of the functional with multiple minimums [32]. A stochastic optimization known as simulated annealing [1,39,40], is guaranteed to reach global minimum but with the

*Corresponding author. Tel.: +202 994 5508; fax: +202 994 0227.
E-mail address: ikopriva@gwu.edu (I. Kopriva).

0925-2312/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.neucom.2004.11.010

very low speed of convergence. Geman and Geman proved in 1984 the convergence guaranteed to find the global minimum by means of classical Gaussian annealing to be an exceedingly slow admissible cooling schedule, $T_a(t) = T_0/\log t$. Their proof used the Metropolis annealing algorithm to generate random walks based on Gaussian distribution [17]. Szu in 1986 [39,40] had extended Geman–Geman convergence proof for the case of the Cauchy noise with unbounded variance combining naturally both Gaussian random walks with Levi random flights achieving the admissible cooling schedule $T_c(t) = T_0/t$. In this paper, we have augmented the classical gradient optimization with the fast fluctuating term using the rapid Cauchy annealing cooling schedule. We coined this approach the stochastic gradient optimization. One important application of the derived stochastic gradient optimization aimed to be unsupervised learning applied to the solution of the highly non-stationary linear inverse problems known as blind source separation (BSS) [11,6,4,3,10,9,12,13,23,24,26,33,46]. In this regard we have introduced a novel Helmholtz free energy contrast function, $H = E - T_0 S$, with the imposed thermodynamics constraint at a constant temperature T_0 generalizing the Shannon maximum entropy S of the closed systems to the open systems having non-zero input–output energy exchange E. Following BSS terminology for linear data models, only the input data vector was known while the source vector and mixing matrix were unknown. In comparison with a number of the cost functions for BSS already proposed we have demonstrated a feature of the Helmholtz free energy cost function to have global minimum at the solution of the linear inverse problem. That enabled the applicability of the proposed cost function to solve the BSS problems when the unknown mixing matrix varied from measurement to measurement. In this paper, we have successfully applied a stochastic gradient optimization to solve inverse space-variant imaging problems on a concurrent pixel-by-pixel basis with the unknown mixing matrix (imaging point spread function) varying from pixel to pixel.

The organization of the paper is as follows. In Section 2, we have introduced the BSS problem as well as the Helmholtz free energy cost function with the classical gradient solution for the BSS problem. Section 3 gives the convergence proofs for both Cauchy and Gaussian annealing, along with their differences in free space and in gradient potential wells. The stochastic gradient is also introduced in Section 3. Performances of the 2-dimensional Cauchy and Gaussian annealing search algorithms as well as performances of stochastic gradient algorithm with Cauchy and Gaussian cooling schedule were compared with multiple minimums on the objective function. Section 4 gives more detailed description and illustration of the Helmholtz free energy $H = E - T_0 S$ applied on the solution of both linear and nonlinear BSS problems. Comparison has been carried out with the adaptive independent component analysis (ICA) algorithms for linear [3,6,10,33] and post-nonlinear [46] mixtures. The conclusion is given in Section 5. For readers' convenience, Appendix A provides a derivation of the higher-dimensional Cauchy annealing algorithm based on the transformation of the higher-dimensional Cauchy pdf from Cartesian to hyper-spherical coordinates. Biological conjecture of the unsupervised learning based on the minimum of the Helmholtz free energy is given in Appendix B.