

Stationary Density of Stochastic Search Processes

Arturo Berrones

Universidad Autónoma de Nuevo León, Mexico

Dexmont Peña

Universidad Autónoma de Nuevo León, Mexico

Ricardo Sánchez

Universidad Autónoma de Nuevo León, Mexico

INTRODUCTION

The optimization of a cost function which has a number of local minima is a relevant subject in many important fields. For instance, the determination of the weights of learning machines depends in general on the solution of global optimization tasks (Haykin, 1999). A feature shared by almost all of the most common deterministic and stochastic algorithms for continuous non – linear optimization is that their performance is strongly affected by their starting conditions. Depending on the algorithm, the correct selection of an initial point or set of points have direct consequences on the efficiency, or even on the possibility to find the global minima. Of course, adequate selection of seeds implies prior knowledge on the structure of the optimization task. In the absence of prior information, a natural choice is to draw seeds from a uniform density defined over the search space. Knowledge on the problem can be gained through the exploration of this space.

In this contribution is presented a method to estimate probability densities that describe the asymptotic behavior of general stochastic search processes over continuously differentiable cost functions. The relevance of such densities is that they give a description of the residence times over the different regions of the search space, after an infinitely long exploration. The preferred regions are those which minimize the cost globally, which is reflected in the asymptotic densities. In first instance, the resulting densities can be used to draw populations of points that are consistent with the global properties of the associated optimization tasks.

BACKGROUND

Stochastic strategies for optimization are essential to most of the heuristic techniques used to deal with complex, unstructured global optimization problems (Pardalos, 2004). The roots of such methods can be traced back to the Metropolis algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller & Teller, 1953), introduced in the early days of scientific computing to simulate the evolution of a physical system to thermal equilibrium. This process is the base of the simulated annealing technique (Kirkpatrick, Gellat & Vecchi, 1983), which makes use of the convergence to a global minimum in configurational energy observed in physical systems at thermal equilibrium as the temperature goes to zero.

The method presented in this contribution is rooted in similar physical principles as those on which simulated annealing type algorithms are based. However, in contrast with other approaches (Suykens, Verrelst & Vandewalle, 1998) (Gidas, 1995) (Parpas, Rustem & Pistikopoulos, 2006), the proposed method considers a density of points instead of Markov transitions of individual points. The technique is based in the interplay between Langevin and Fokker – Planck frameworks for stochastic processes, which is well known in the study of out of equilibrium physical systems (Risken, 1984) (Van Kampen, 1992). Fokker - Planck equation has been already proposed for its application in search algorithms, in several contexts. For instance, it has been used to directly study the convergence of populations of points to global minima (Suykens, Verrelst & Vandewalle, 1998), as a tool to demonstrate the convergence of simulated annealing type algorithms (Parpas, Rustem & Pistikopoulos, 2006) (Geman & Hwang, 1986), or

as a theoretical framework for Boltzmann type learning machines (Movellan & McClelland, 1993) (Kosmatopoulos & Christodoulou, 1994). In the context of global optimization by populations of points, it has been proposed that the populations evolve under the time – dependent version of the Fokker – Planck equation, following a schedule for the reduction of the diffusion constant D (Suykens, Verrelst & Vandewalle, 1998).

In our approach, the stationary version of the Fokker – Planck equation is used to learn the long – term probability density of a general stochastic search process. This is achieved using linear operations and a relatively small number of evaluations of the given cost function.

STATIONARY DENSITY ESTIMATION ALGORITHM

Consider the minimization of a cost function of the form $V(x_1, x_2, \dots, x_n, \dots, x_N)$ with a search space defined over $L_{1,n} \leq x_n \leq L_{2,n}$. A stochastic search process for this problem is modeled by

$$\frac{dx_n}{dt} = -\frac{\partial V}{\partial x_n} + e(t) \quad (1)$$

where $e(t)$ is an additive noise with zero mean. Equation (1), known as Langevin equation in the Statistical Physics literature (Risken, 1984) (Van Kampen, 1992), captures the basic properties of a general stochastic search strategy. Under an uncorrelated Gaussian noise with constant strength, Eq. (1) represents a search by diffusion, while a noise strength that is slowly varying in time gives a simulated annealing process. Notice that choosing an external noise of infinite amplitude, the dynamical influence of the cost function over the exploration process is lost, leading to a blind search. The model given by Eq. (1) can be interpreted as a nonlinear dynamical system composed by N interacting particles. The temporal evolution of the probability density of such a system is described by a linear differential equation, the Fokker – Planck equation (Risken, 1984) (Van Kampen, 1992),

$$\frac{dp}{dt} = \frac{\partial}{\partial x} \left[\frac{\partial V}{\partial x} \right] + D \frac{\partial^2 p}{\partial x^2} \quad (2)$$

The approach proposed in this article is based on the notion of an infinitely long exploration of the search space. In the present model setup for the search, the process converges to a state described by the stationary solution of Eq. (2) (Berrones, 2007). The form of this solution is of the well known Boltzmann type (Risken, 1984) (Van Kampen, 1992). For optimization or deviate generation purposes, its direct use would imply a high computational cost. Instead, a form of Gibbs sampling is proposed in order to estimate the marginal probability density $p(x_n)$ (the details of the following discussion can be consulted in (Berrones, 2007)). The one dimensional projection of Eq. (2) at $t \rightarrow \infty$ leads to the following equation for the conditional cumulative distribution, $y(x_n | \{x_j \neq x_n\})$

$$\frac{d^2 y}{dx_n^2} + \frac{1}{D} \frac{\partial V}{\partial x_n} \frac{dy}{dx_n} = 0$$

$$y(L_{1,n}) = 0, \quad y(L_{2,n}) = 1 \quad (3)$$

Therefore, the estimation of the analytical form of $y(x_n | \{x_j \neq x_n\})$ can be achieved by the substitution of the expansion

$$y = \sum_{l=1}^L a_l \phi_l(x_n) \quad (4)$$

into Eq. (3). The distribution obtained in this way can be used to draw points from the conditional density $p(x_n | \{x_j \neq x_n\})$. According to the principles of Gibbs sampling (Geman & Geman, 1984), the iteration of the previous steps over the N variables will produce a population sampled from the corresponding marginal densities $p(x_n)$. However, in our setup all the information needed to characterize the densities is contained in the coefficients of the expansion (4). In this way, the stationary marginal densities associated to the N variables of the optimization problem, are learned through the averages of the coefficients over the iteration of the random deviate generation process. We call this basic procedure a Stationary Density Estimation Algorithm (SDEA). We have also named the method Stationary Fokker – Planck Machine (SFPM) in (Berrones, 2007), in order to indicate its relation with other methods (Suykens, Verrelst & Vandewalle, 1998) that make use of the Fokker – Planck equation to learn statistical features of stochastic search processes. However, in

3 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/stationary-density-stochastic-search-processes/10431

Related Content

Evaluation of Data Imbalance Algorithms on the Prediction of Credit Card Fraud

Godlove Otoo, Justice Kwame Appati, Winfred Yaokumah, Michael Agbo Tettey Soli, Stephane Jnr Nwolley and Julius Yaw Ludu (2021). *International Journal of Intelligent Information Technologies* (pp. 1-26). www.irma-international.org/article/evaluation-of-data-imbalance-algorithms-on-the-prediction-of-credit-card-fraud/289967

Chaotic Neural Networks

Emilio Del-Moral-Hernandez (2009). *Encyclopedia of Artificial Intelligence* (pp. 275-281). www.irma-international.org/chapter/chaotic-neural-networks/10260

From the Real Ant to the Artificial Ant: Applications in Combinatorial Optimization, Data Clustering, Collective Robotics and Image Processing

Moussa Diaf, Kamal Hammouch and Patrick Siarry (2012). *International Journal of Signs and Semiotic Systems* (pp. 45-68). www.irma-international.org/article/from-the-real-ant-to-the-artificial-ant/101251

Object-Assisted Question Featurization and Multi-CNN Image Feature Fusion for Visual Question Answering

Sruthy Manmadhan and Binsu C. Koor (2023). *International Journal of Intelligent Information Technologies* (pp. 1-19). www.irma-international.org/article/object-assisted-question-featurization-and-multi-cnn-image-feature-fusion-for-visual-question-answering/318671

Using Fuzzy Goal Programming Technique to Solve Multiobjective Chance Constrained Programming Problems in a Fuzzy Environment

Animesh Biswas and Nilkanta Modak (2012). *International Journal of Fuzzy System Applications* (pp. 71-80). www.irma-international.org/article/using-fuzzy-goal-programming-technique/63356