ON THE EFFICIENCY OF A RANDOM SEARCH METHOD

G. Pérez-Lechuga[1], J.C.S. Tuoh-Mora [1], E. Morales-Sánchez [1], M. M. Álvarez Suárez[1],

[1] Centro de Investigación Avanzada en Ingeniería Industrial
Universidad Autónoma del Estado de Hidalgo
ICBI, CIAII, Carretera Pachuca-Tulancingo Km.4.5, Pachuca Hgo, México.

Abstract:- This paper studies the efficiency of the random search reported by Rubinstein [1] and widely studied by Pérez-Lechuga [2]. We proof that the efficiency of the selected random search algorithm is a linear function both of the step size and the direction of the descent movement. We report the theoretical results.

Key Words:- Stochastic optimization, random search, stochastic approximation, gradient approximation, quasigradient method.

1. Introduction

Stochastic approximation algorithms can be used in system optimization problems for which only noisy measurements of the system are available without knowing the gradient of the objective function. This type of problem can be found in adaptative control, neural network training,experimental design, stochastic optimization and many other areas.

The main idea of the stochastic quasigradient methods is to solve a wide class of optimization problems with a complex nature of objective functions and constraints. These methods are stochastic algorithmic procedures for solving general constrained problems with nondifferentiable, nonconvex functions.

For stochastic programming problems, these techniques generalize the well known stochastic approximation method for unconstrained optimization of the expectation of a random function to problems involving general contraints.

Consider the general stochastic programming problem

$$\text{Minimize}\, F_0(x) = \mathbf{E}\,[f_0(x,\omega)], \qquad (1)$$

Subject to $x \in \mathcal{S} \subset \mathbb{R}^n$, where

$$\mathcal{S} = \{x \mid f_i(x,\omega) \leq 0,\ i = 1,\ldots,m\}, \quad (2)$$

$\mathbf{E}$ is the operation of mathematical expectation with respect of some probability –

space $(\Omega, \mathcal{A}, \mathbf{P})$, and $\omega \in \Omega$.
The more trouble on solving the problem (1) to (2) is that, it is only feasible to calculate the exact values of the functions

$$F_i(x) = \mathbf{E}\left[f_i(x,\omega)\right] = \int f_i(x,\omega)\mathbf{P}(d\omega),$$

$$i = 0, \ldots, m$$

in exceptional cases for special types of probability measures $\mathbf{P}(\omega)$.
If the functions $F_i(x)$ have uniformly bounded second derivatives at $x \in \{x_s\}_{s=0}^{\infty}$ then for the random vectors $\xi_i(s)$ defined as (see [3])

$$\sum_{j=1}^{n} \frac{f_i(x_s + \Delta_s e^j, \omega_{sj}) - f_i(x_s, \omega_{s0})}{\Delta_s} e^j, \quad (3)$$

we would have $\mathbf{E}[\xi_i(s) \mid x_s] = F_i(x_s) + b_i(s)$, $\parallel b_i(s) \parallel \leq \Delta_s$. Where e^j is the unit vector on the jth axis and Δ_s is a positive constant.
The random sequence $x_{s+1} = x_s - \rho_s \xi_s$, $s = 0, 1, \ldots$ converges with probability 1 to the solution of (1) if the following conditions are satisfied with probability 1. For the step size: $\rho_s \geq 0$, $\sum_s \rho_s = \infty$, $\sum_s \mathbf{E}\left[\rho_s \parallel \Delta_s \parallel + \rho_s^2\right] < \infty$. For the quasi-gradient ξ_s, $\mathbf{E}\left[\xi_s \mid x^s\right] = \nabla F_0(x^s) + o(\Delta_s)$.

2. The random search algorithms

In this section we introduce the random search algorithms for optimization problems, in which the computing cost of the random search at a point increases as the point tends to satisfy appropriate optimality conditions. The algorithms progress faster beginning from initial conditions far away from an optimal or suboptimal point, and they gain precision with expense of efficiency as such a point is approached. The algorithms start at some $x_0 \in \mathcal{S}$, and they generate a sequence $x_0, x_1, \ldots, \in \mathcal{S}$. These are descent algorithms, in the sense that the sequences $F(x_0), F(x_1) \ldots,$ are monotone decreasing. They generate x_{i+1} from x_i by random search techniques, using the sufficient descent principle. This principle enables that the algorithms do not adopt the first point $y \in \mathbb{R}^n$ found by random search satisfying $F(y) < F(x_i)$ as x_i, but they rather wait to find a point y for which $F(x_i) - F(y)$ is large by some criteria [4]. The amount of descent $F(x_{i+1}) - F(x_i)$, like the amount of time the algorithm spent for the random search at x_i, depends on the desired extent for x_i; the less desirable x_i, the larger the descent will tend to be.
Random search techniques have been an object of research for quite some time. The concept has been initially introduced by Anderson [5] and then developed by Rastrigin [6]. The idea is to determine a descent direction at random, by using a distribution on the unit sphere around x_i, and then, to find a suitable step size. The step size is used by minimizing F along the descent direction. This is determined adaptively, based on the ratio of successful to failed attempts (by random searches) to reduce F. The basic property of this algorithms is that x_i reaches a solution of (1) with probability 1 using a prescribed tolerance as $i \to \infty$. Thus in descent algorithms, one random search is conducted at x_i to generate x_{i+1}. For the convergence states, the probability of x_i satisfying $f(x_i) \leq \inf\{f(x) \mid x \in \mathcal{S}\} + \epsilon$ (for a given $\epsilon > 0$) approaches 1 as $i \to \infty$.

2.1. Stochastic Quasigradient Methods

Stochastic quasigradient methods are a set of techniques useful to solve optimization problems with objective functions and constrains of such a complex nature which make impossible to calculate the precise values of these functions (let alone of their derivatives). The basic idea is to use statistical estimates for the values of the functions rather than precise values. For stochastic programming problems these methods generalize the well-known stochastic approximation method for unconstrained optimization of the expectation of a random function. This stochastic problem can be defined as follows.

$$\min\{\mathbf{E}_\omega f(x,\omega) : x \in \mathcal{S}\}, \tag{4}$$

where x represents the variable to be chosen optimally, $\mathcal{S}$ is a set of constraints, and ω is a random variable belonging to some probabilistic space $(\Omega, \mathcal{B}, \mathbf{P})$. Here $\mathcal{B}$ is a Borel field and $\mathbf{P}$ is a probabilistic measure. In this problem we assume that $\mathcal{S}$ is defined in such a way that the projection operation $x \rightarrow \Pi_{\mathcal{S}}(x)$ is comparatively inexpensive from a computational point of view, where

$$\Pi_{\mathcal{S}}(x) = \mathrm{argmin}_{Z\in\mathcal{S}} \parallel x - Z \parallel$$

In this case it is possible to implement a stochastic quasigradient algorithm of the following type

$$x_{i+1} = \Pi_{\mathcal{S}}(x_i - \rho_i \varphi_i), \tag{5}$$

Here x_i is the current approximation of the optimal solution, ρ_i is the steep size, and φ_i is a random step direction. This step direction may, for instance, be a statistical estimate of the gradient (or subgradient in the nondifferentiable case) of $f(x)$, then $\varphi_i \equiv \xi_i$, such that

$$\mathbf{E}(\xi_i \mid x_1, \ldots x_i) = \nabla F_i(x_i) + a_i, \tag{6}$$

$$i = 0, \ldots, m$$

where a_i decreases for an increasing number of iterations, the vector ξ_i is called a *stochastic quasigradient* of functions $F_i(x)$, and $\nabla F(x)$ is the subgradient of $F(x)$ in each point x_i. Usually $\rho_i \rightarrow 0$ as $i \rightarrow \infty$ and therefore $\parallel x_{i+1} - x_i \parallel \rightarrow 0$.

Algorithm (5) can also resolve with problems with more general constraints formulated in terms of mathematical expectations $\mathbf{E}_\omega[f_i(x,\omega) \leq 0]$, $i = 1, \ldots, m$, by making use of penalty functions or Lagrangians.

3. Problem definition

The idea of efficiency was introduced by Rubinstein et al. [1] as follows. Let x_{i+1} be the point reached after one single iteration, and $\Delta F_i = F_{i+1} - F_i$ the increment of the value of F. The efficiency of the random search algorithms is defined as

$$C = -\frac{\mathbf{E}(\Delta F_i)}{\mathbf{E}(N_i)}, \tag{7}$$

$$D = C\,[\mathbf{Var}\,\Delta f_i]^{-1/2}, \tag{8}$$

where N_i is the number of observations (measurements) of the convex function $F(x)$ required for the algorithm at the ith step.

In this paper we are interested in evaluate the efficiency of the following algorithm [2]:

$$x_{i+1} = x_i - \alpha_i \gamma_i \xi_i, \quad (9)$$

where α_i is the step size, γ_i is a normalization factor proposed by [1], $\xi_i = \Upsilon^0_{im} B^0_{il}$, and $\Upsilon^0_{il} = \text{mín}\{\Upsilon_{i1}, \ldots, \Upsilon_{il}\}$ denotes the difference

$$\Upsilon_{il} = f(x_i + B_{il}, W_{il}) - f(x_i, W_{i0}), \quad (10)$$

$$l = 1, \ldots, \mathcal{H}$$

here, W_{il} and W_{i0} are the realizations observed from the noise at points x_i and $x_i + B_{il}$ respectively. B^0_{il} denotes the vector in which the minimum increment is produced, and $\mathcal{H}$ is the number of points generated on the surface of the n-dimensional unit hipersphere.

Note that B_{il} are independent and uniformly distributed vectors on the surface of such sphere. We assume that $f(x, W) = f(x) + W$, where $W \sim N(0, \sigma^2)$. From the convexity of f we have

$$f(x_i + \Delta x_i) - f(x_i) \geq \langle \Delta x_i, \nabla f(x_i) \rangle$$

or equivalently

$$f(x_{i+1}) = f(x_i + \Delta x_i) =$$

$$f(x_i) + \langle \Delta x_i, \nabla f(x_i) \rangle + \delta(\Delta x_i) = f(x_i)$$

$$+ \parallel \Delta x_i \parallel \parallel \nabla f(x_i) \parallel \cos\theta + \delta(\Delta x_i),$$

where $\cos\theta$ is the angle between the unit vectors Δx_i and $\nabla f(x_i)$, and $\delta(\Delta x_i) \to 0$ as $\parallel \Delta x_i \parallel \to 0$.

We analyze two cases. In the first, we consider the noise $W = 0$, and in the second, $W \sim N(0, \sigma^2)$.

First case: $W = 0$. From (9), note that

$$\Delta x_i = x_{i+1} - x_i = -\alpha \gamma_i \Upsilon^0_{il} B^0_{il}, \quad (11)$$

Substituting (10) in (9) we obtain

$$\begin{aligned} f(x_{i+1}) &= f(x_i) + \alpha_i \gamma_i \Upsilon^0_{il} \parallel B^0_{il} \parallel \\ &\quad \parallel \nabla f(x_i) \parallel \cos\theta + \delta(\Delta x_i) \end{aligned}$$

Taking into account that $\parallel B^0_{ig} \parallel = 1$, and for Δx_i sufficiently small, then

$$f(x_{i+1}) = f(x_i) + \alpha_i \gamma_i \Upsilon^0_{il} \cos\theta,$$

therefore

$$\Delta f_i = \alpha_i \gamma_i \Upsilon^0_{il} \cos\theta.$$

Thus, by (7)

$$\begin{aligned} C_i &= \frac{\mathbf{E}\left[f(x_i) - f(x_{i+1})\right]}{\mathcal{H}} = \frac{\mathbf{E}\left[\Delta f_i\right]}{\mathcal{H}} \\ &= \frac{1}{\mathcal{H}} \alpha_i \gamma_i \Upsilon^0_{il} \, \mathbf{E}\left[\cos\theta\right] \quad (12) \end{aligned}$$

where the probabilty density function of the random angle is given by (see [7])

$$\zeta_n(\theta) = \frac{\sin^{n-2}(\theta)}{\int_0^\pi \sin^{n-2}(\theta) d\theta}$$

$$= \varrho_n \sin^{n-2}(\theta), \; -\pi/2 \leq \theta \leq \pi/2, \quad (13)$$

where

$$\varrho = \frac{\Gamma(n/2)}{\sqrt{\pi}\,\Gamma[(n-1)/2]},$$

and Γ denotes the gamma function.

If $h_i = \alpha_i \gamma_i \Upsilon^0_{il}$, then substituting (13) in (12) we have that

$$\mathbf{E}\left[\Delta f_i\right] = h_i \int_{-\pi/2}^{\pi/2} \cos(\theta) \; \zeta_i(\theta) d\theta =$$

$$2\varrho h_i \int_0^{\pi/2} \cos\theta \sin^{n-2}(\theta) d\theta = \frac{2\varrho h_i}{n-1},$$

therefore (7) takes the form

$$-\frac{\mathbf{E}(\Delta f_i)}{\mathbf{E}(N_i)} = \frac{2\varrho h_i}{\mathcal{H}(n-1)}, \quad (14)$$

Second case: $W \sim N(0,\sigma^2)$. Consider the event

$$x_{i+1} = \begin{cases} x_i - h_i B^0_{il}, & \text{if } \Upsilon = \min\{\Upsilon_{il}\} \\ x_i, & \text{in other case} \end{cases} \tag{15}$$

note that for any i iteration, Υ_{il} is such that

$$\begin{aligned} \Upsilon_{il} &= \psi(x_i + B_{il}, W_{il}) - \psi(x_i, W_{i0}) \\ &= f(x_i + B_{il}) + W_{il} - f(x_i) - W_{i0} \\ &= f(x_i + B_{il}) - f(x_i) \end{aligned} \tag{16}$$

where W_{il} and W_{i0} are the realizations of the noise observed at points $x_i + B_{il}$ and x_i. The probability of this event is (see [2])

$$\mathbf{P} = \frac{1}{2}\left(1 + \phi\left(\frac{|\Delta f|}{2\sigma}\right)\right).$$

where $\phi(y) = \int_0^y 2\pi^{-1/2} e^{-t^2} dt$. Thus

$$\mathbf{P} = \frac{1}{2}\left(1 + \phi\left(\frac{|h_i \cos\theta|}{2\sigma}\right)\right).$$

As in [1], let $\mathcal{Q}$ be the random variable defined by

$$\mathcal{Q} = \begin{cases} \cos\theta, & \text{with probability } (\mathbf{P}) \\ -\cos\theta & \text{with probability } (1-\mathbf{P}) \end{cases}$$

for $-\pi/2 \le \theta \le \pi/2$. Then, taking the mathematical expectation in $\mathcal{Q}$ we obtain

$$\begin{aligned} &\mathbf{E}[\mathcal{Q}] = \int_{-\pi/2}^{\pi/2} \varrho_n \mathbf{P}[\cos\theta \sin^{n-2}(\theta)]d\theta - \\ &\int_{-\pi/2}^{\pi/2} \varrho_n (1-\mathbf{P})[\cos\theta \sin^{n-2}(\theta)]d\theta = \\ &2\varrho \int_0^{\pi/2} \mathbf{P}[\cos\theta \sin^{n-2}(\theta)]d\theta - \\ &2\varrho \int_0^{\pi/2} (1-\mathbf{P})[\cos\theta \sin^{n-2}(\theta)]d\theta. \end{aligned}$$

After some algebraic manipulations, wehave

$$\frac{\mathbf{E}[\mathcal{Q}]}{2\varrho_n} = \int_0^{\pi/2} \cos\theta \sin^{n-2}(\theta)\,\phi\left(\frac{|\Delta f_i|}{2\sigma}\right) d\theta = \int_0^{\pi/2} \cos\theta \sin^{n-2}(\theta)\,\phi\left(\frac{|h_i\cos\theta|}{2\sigma}\right) d\theta. \tag{17}$$

Then, C_i takes the form

$$\frac{2\varrho}{\mathcal{H}} \int_0^{\pi/2} \cos\theta \sin^{n-2}(\theta)\phi\left(\frac{|h_i\cos\theta|}{2\sigma}\right) d\theta. \tag{18}$$

In the final analysis we let us estimate the variance of Δf_i from $\mathbf{E}[\mathcal{Q}^2] - [\mathbf{E}[\mathcal{Q}]]^2$. Note that

$$\begin{aligned} &\mathbf{E}[\mathcal{Q}^2] = 2\varrho \int_0^{\pi/2} \mathbf{P}\cos^2\theta \,\mathrm{sen}^{n-2}(\theta)d\theta - \\ &2\varrho \int_0^{\pi/2} (1-\mathbf{P})\cos^2\theta \,\mathrm{sen}^{n-2}(\theta)d\theta = \\ &2\varrho \int_0^{\pi/2} \phi\left(\frac{|h_i\cos\theta|}{2\sigma}\right)\cos^2\theta \,\mathrm{sen}^{n-2}(\theta)d\theta \end{aligned}$$

Since, for x small,

$$\phi(x) \approx \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\left(x - \frac{x^3}{2\cdot 3} + \ldots\right), \tag{19}$$

then $\mathbf{E}[\mathcal{Q}]$ can be written as

$$\varrho\left[\frac{1}{n-1} + \frac{h_i}{\sigma\sqrt{2\pi}} \int_0^{\pi/2} \cos^2\theta \,\mathrm{sen}^{n-2}(\theta)d\theta\right]$$

and $\mathbf{E}[\mathcal{Q}^2]$ is defined by

$$\varrho\left[\int_0^{\pi/2} \frac{\mathrm{sen}^{n+1}(\theta)d\theta}{n} + \frac{h_i}{\sigma\sqrt{2\pi}} \int_0^{\pi/2} \cos^3\theta \,\mathrm{sen}^{n-1}(\theta)d\theta\right],$$

thus (7) takes the form

$$C_n = \frac{\varrho_n}{\mathcal{H}(n-1)} +$$

$$\frac{h_i \varrho_n}{\mathcal{H}\sigma\sqrt{2\pi}} \int_0^{\pi/2} \cos^2\theta \operatorname{sen}^{n-2}(\theta) d\theta. \quad (20)$$

Finally, and using (19), (7) can be defined as

$$D_n = \frac{\varrho}{\mathcal{H}(n-1)\mathbf{Var}[\mathcal{Q}]} +$$

$$\frac{h_i \int_0^{\pi/2} \cos^2\theta \operatorname{sen}^{n-2}(\theta) d\theta}{\mathcal{H}(n-1)\mathbf{Var}[\mathcal{Q}]}. \quad (21)$$

Where (20) and (21) can be written in the linear form

$$C_n = a_n + b_n h_i, \quad D_n = a'_n + b'_n h_i, \quad (22)$$

where $a_n = \frac{\varrho}{\mathcal{H}(n-1)}$, and $a'_n = \frac{\varrho}{\mathcal{H}(n-1)\mathbf{Var}[\mathcal{Q}]}$,

$b_n = \frac{h_i \varrho_n}{\mathcal{H}\sigma\sqrt{2\pi}} \int_0^{\pi/2} \cos^2\theta \operatorname{sen}^{n-2}(\theta) d\theta$, and

$b'_n = \frac{\int_0^{\pi/2} \cos^2\theta \operatorname{sen}^{n-2}(\theta) d\theta}{\mathcal{H}(n-1)\mathbf{Var}[\mathcal{Q}]}$.

Table 1 shows some values of C_n. Here, $\vartheta_n = \int_0^{\pi/2} \cos^2\theta \, \sin^{n-2}(\theta) d\theta$.

n	ϱ_n	ϑ_n	C_n
2	1.2837	0.7853	0.1283 + 0.0402 h_i
3	0.5641	0.3331	0.0282 + 0.0749 h_i
4	0.5641	0.1963	0.0188 + 0.0441 h_i
5	0.1880	0.1333	4.7 E-3 + (1E-3) h_i
6	0.0940	0.0981	1.8 E-2 + (3.6E-4)h_i

Table 1: Efficiency of C_n

4. Conclusions

For the algorithm presented, the efficiency of the search can be viewed as a linear function of the h_i parameter, and of the evaluated points on the surface of the unit sphere. Where $h_i = \alpha_i \gamma_i \Upsilon^0_{im}$.

5. References

1. Rubinstein, Y. and Samorodnitsky, G., "Efficieny of the random search method", *Mathematics and computers in simulation*, XXIV, pp. 257-268, 1982.

2. Pérez, L. G., *An algorithm for the stochastic optimization of some dynamic models*, Engineering Doctor Dissertation, DEPFI, UNAM, 1993.

3. Yu. M. Ermoliev., "Stochastic quasigradient methods and their application to system optimization", *Stochastics*, Vol. **9**, pp. 1-36, 1983.

4. Wardi, Y., "Random search algorithms with sufficient descent for minimization of functions", *Mathematics of Operations Research*, Vol 14, No. **2**, pp. 343-354, 1989.

5. Anderson, R. L. "Recent advances in finding best operating conditions", *J. Amer. Statist, Assoc.*, **48**, pp. 789-798, 1953.

6. Rastrigin, L. A. "Extremal control by methods of random scanning". *Automat. Remote Control*, **21**, pp. 891-896, 1960.

7. Reuben, Y.R., "Generating random vectors uniformly distributed inside and on the surface of different regions", *Euro., J.O.R.*, Vol. **10**, pp.205-209, 1982.