**A short study on minima distribution**

Loc Nguyen

Loc Nguyen’s Academic Network, Vietnam

Email: ng\_phloc@yahoo.com

Homepage: www.locnguyen.net

**Abstract**

Global optimization is the imperative development of local optimization because there are many problems in artificial intelligence and machine learning requires highly acute solutions over entire domain. There are many methods to resolve the global optimization, which can be classified into three groups such as analytic methods (purely mathematical methods), probabilistic methods, and heuristic methods. Especially, heuristic methods like particle swarm optimization and ant bee colony attract researchers because their effective and practical techniques which are easy to be implemented by computer programming languages. However, these heuristic methods are lacking in theoretical mathematical fundamental. Fortunately, minima distribution establishes a strict mathematical relationship between optimized target function and its global minima. In this research, I try to study minima distribution and apply it into explaining convergence and convergence speed of optimization algorithms. Especially, weak conditions of convergence and monotonicity within minima distribution are drawn so as to be appropriate to practical optimization methods.

|  |  |
| --- | --- |
|  |  |

**1. Introduction**

Section 2 draws weak conditions of convergence and monotonicity whereas section 3 continues to mention convergence speed. Section 4 describe an experiment on convergence speed associated with particle swarm optimization (PSO) algorithm. Section 5 is the conclusions.

**2. Weak conditions of convergence and monotonicity**

Two most important properties of minima distribution are convergence and monotonicity. Stability and shrinkage are derived from the two properties. As a summary, Luo (Luo, 2019, p. 5) proposed following conditions for satisfying convergence and monotonicity of minima distribution:

* Function *τ*(*x*) is defined based on function *ρ* such that *τ*(*x*) = *ρ*(*f*(*x*)) where *ρ* is monotonically decreasing and *ρ* is positive, for instance, *ρ*(*y*) > 0 for all *x* belonging the domain *X* such that *y* = *f*(*x*).
* Function *τ*(*x*) is defined as *τ*(*x*) = exp(–*f*(*x*)) so as to easily obtain the monotonicity.

However, it is not necessary to strictly define *τ*(*x*) based on *ρ* and *f*(*x*) for convergence. Following the proofs of theorem 1 in (Luo, 2019, p. 7), a *weak convergence condition* is drawn that:

* Function *τ*(*x*)is positive for all *x* belonging to domain *X*.
* Given any *x* which is not a maximizer of *f*(*x*), there always exists an open set having nonzero Lebesgue measure such that *τ*(*t*) > *τ*(*x*) for all *t* belonging to this open set.

This condition is easily proved by proofs of theorem 1 (vi, vii) and theorem 2 (Luo, 2019, pp. 7-8). Of course, if *τ*(*x*) is defined as *τ*(*x*) = *ρ*(*f*(*x*)) where *ρ* is monotonically decreasing and *ρ* is positive, the weak convergence condition is satisfied. This condition is useful in practical cases that *τ*(*x*) is defined as a nonconstant, positive, and differential function which has no actual minimizer with note that some infima may not be actual minima, for example, *τ*(*x*) = exp(–*x*), *τ*(*x*) = exp(–*x*2), etc.

Within the weak convergence condition, theorem 3 for nonnegativity (Luo, 2019, p. 8) is obtained because *τ*(*x*)is positive.

Monotonicity can be achieved if the first order derivative of with regard to *k* is greater than or equal to 0. Suppose *m*(*k*)(*x*) is differentiable, we have:

Suppose positive function *τk*(*x*) is differentiable with regard to *k* and its derivatives with regard to *k* is formulated as follows:

In other words, *g*(*x*) is specified as follows:

|  |  |
| --- | --- |
|  | (1.1) |

Following the proof of theorem 1 (v) (Luo, 2019, p. 7), we have:

|  |  |
| --- | --- |
|  | (1.2) |

Where:

Therefore, we obtain (Luo, 2019, p. 9):

|  |  |
| --- | --- |
|  | (1.3) |

Where:

According to Gurland’s inequality (Luo, 2019, p. 6), if *f*(*x*) and *g*(*x*) are inversely proportional, for instance, *f*(*x*) is nonincreasing and *g*(*x*) is nondecreasing in every open set, and vice versa then, *E*(*k*)(*yg*) ≤ (*E*(*k*)(*y*))(*E*(*k*)(*g*)). As a result, a *weak monotonicity condition* is that positive function *τ*(*x*) is defined such that *f*(*x*) and *g*(*x*) are inversely proportional where *g*(*x*) is specified by equation 1.1. Note that the weak monotonicity condition follows the weak convergence condition.

Because the target function f(x) is too complicated to determine whether *f*(*x*) and *g*(*x*) are inversely proportional, we concern the case that *g*(*x*) can be specified as *g*(*x*) = *h*(*y*) where *y* = *f*(*x*). In other words, if *g* is function of *f*(*x*), we have:

|  |  |
| --- | --- |
|  | (1.4) |

Where:

Given *y* and *h*(*y*) where *y* is increasing function, the weak monotonicity condition is simplified that positive function *τ*(*x*) is defined such that *h*(*y*) is nonincreasing function with regard to *y*.

As usual, *τk*(*x*) is the *kth*-power function of *τ*(*x*) as *τk*(*x*) = (*τ*(*x*))*k*. Here *τk*(*x*) is generalized as *τk*(*x*) = *wk*(*τ*(*x*)) where *wk*(*τ*) is called power function for *τ* which has two properties as follows:

Note, the power function *wk*(*τ*) is function of *k*, which is determined based on the value *τ*(*x*). In trivial cases, we have *wk*(*τ*) = (*τ*(*x*))*k*, *wk*(*τ*) = , etc. With the two properties of *wk*(*τ*), it is easy to assert the weak convergence condition by the proofs of theorem 1 (vi, vii) (Luo, 2019, p. 7). Based on two weak conditions of convergence and monotonicity, convergence speed is proposed in the next section.

**3. Convergence speed**

For all real *k* and real Δ*k* > 0, two successive integrals within the weak convergence condition are determined as follows:

Convergence speed is defined as the absolute differential of two successive integrals as follows:

It is easy to recognize that the convergence speed is the absolute value of the first order derivative of specified by equation 1.3.

|  |  |
| --- | --- |
|  | (2.1) |

Note, *g*(*x*) is defined by equation 1.1. The magnitude of *cτ* is proportional to the difference between *f*(*x*) and *g*(*x*) which is determined by the following derivative:

Here we can ignore the target function *f*(*x*) because it is not the parameters for all optimization methods. Therefore, let *Qτ* be the metric that measures the convergence speed, which is defined as follows:

|  |  |
| --- | --- |
|  | (2.2) |

Actually, *Qτ* is absolutely slope of function *g*. The steeper *Qτ* is, the larger *cτ* is, which in turn, the faster convergence speed. For example, if the definition of function *τ* is associated with normal distribution given mean 0 and variance 1, it becomes:

Therefore,

If the definition of function *τ* is associated with Gumbel distribution for extreme value given location parameter 0 and scale parameter 1, it becomes:

Therefore,

Obviously, convergence speed of *τG*(*x*) is faster than *τN*(*x*) because *gG*(*y*) is steeper than *gN*(*y*). Given normal distribution with mean *μ* and variance *σ*2 then, *τN*(*x*) is redefined as follows:

Variance *σ*2 reflects the steeper of *τN*(*x*). Therefore, the smaller variance *σ*2 is, the steeper function *τN*(*y*) is. The next section describes an experiment with varying variance *σ*2.

**4. Experiment with PSO algorithm**

In this research, convergence speed of minima distribution is calculated, which in turn derives the conclusion that function *τ*(*x*) = *ρ*(*f*(*x*)) should be more steeper regarding *f*(*x*) in order to improve the convergence speed. In particle swarm optimization (PSO) algorithm, movement of particles obeys normal distribution. Therefore, this section describes an experiment by varying variance of suchnormal distribution with note that *τ*(*x*) here is associated with the PSO normal distribution.

**4. Conclusions**

Some optimization algorithms like PSO take advantages of distribution of *x* instead of taking advantages of *f*(*x*). In other words, they define implicitly *τ*(*x*) as function of *x* instead of function of *f*(*x*) like *τ*(*x*) = *ρ*(*f*(*x*)). Therefore, it is reasonable to assert their convergence by the weak convergence condition. The convergence speed also depends on the acuteness of the power function *wk*(*τ*), besides the slope *Qτ* of *τ*(*x*). It is inferred that *wk*(*τ*) implies the knowledge amount of given optimization algorithm after each iteration. For PSO, heuristic movement of particles after each iteration reflects the power function *wk*(*τ*). In the future trend, I will research minima distribution with Bayesian optimization because Bayesian optimization takes full advantages of prior information which is the knowledge amount associated with the power function.

**References**

Luo, X. (2019, May 24). Minima distribution for global optimization. *arXiv preprint*. doi:10.48550/arXiv.1812.03457