

# **Applications of Maximum Entropy Principle in Formulating Solution of Constrained Non-Linear Programming Problem**

**Om Parkash<sup>1</sup>, Mukesh<sup>2</sup> and Radhey S. Singh<sup>3</sup>**

## **Abstract**

The maximum entropy principle finds its applications in a variety of disciplines related with Operations Research. The use of the principle can be extended to obtain the solution of a mathematical programming problem. In the present communication, an algorithm has been developed for solving constrained non-linear programming problem through the equivalent surrogate problem. In the solution process of so formed surrogate problem, the surrogate multipliers need to be updated and therefore, maximum entropy formulation has been used for the update of the multipliers. For the implementation and well understanding of the algorithm, an example has been discussed which confirms that entropy based updates for the surrogate multipliers lead to a fast and accurate solution to the problem.

---

<sup>1</sup> Department of Mathematics, Guru Nanak Dev University, Amritsar-143005, India.  
E-mail: omparkash777@yahoo.co.in

<sup>2</sup> Department of Mathematics, Guru Nanak Dev University, Amritsar-143005, India.  
E-mail: sarangal.mukesh@yahoo.co.in

<sup>3</sup> Department of Statistics and Actuarial Science, University of Waterloo, Ontario N2L 3G1, Canada. E-mail: rssingh@uwaterloo.ca

**Mathematics Subject Classification:** 94A15; 94A17

**Keywords:** Entropy; Maximum entropy principle; Non-linear programming; Saddle point; Surrogate problem

## 1 Introduction

Information theory provides a constructive role for setting up probability distributions when partial information is available and leads to a type of statistical inference which is called the maximum-entropy estimate. This estimate is least biased as it is maximally noncommittal with regard to missing information. The principle of maximum entropy introduced by Jaynes [1] provides a unique solution for the posterior probability distribution based on the intuition that the information gain consistent with assumptions and evidence should be minimal.

In real life situations as well as in several disciplinary pursuits, we are very frequently concerned with constrained optimization. For such pursuits, Kapur [2] explained the principle of maximum entropy with the view that at any stage, we may have infinity of probability distributions consistent with the given constraints, such as,

$$\sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i g_{ri} = a_r, r = 1, 2, \dots, m, m+1 < n, p_1 \geq 0, \dots, p_n \geq 0 \quad (1.1)$$

Out of these distributions, one has maximum uncertainty  $H_{\max}$  and the uncertainty  $H$  of any other distribution is less than  $H_{\max}$  and if we desire to reduce uncertainty, then can be done only with the help of some additional information. Thus, the use of any distribution other than the maximum uncertainty distribution implies the use of some information in addition to that given by equation (1.1). Now, according to Jaynes [1] maximum entropy principle which is nothing but the principle of scientific objectivity and it implies that we should use all the information given to us and avoid using any information not given to us.

For the deeper study of maximum entropy principle to a variety of disciplines, Kapur [2] provided the applications of Shannon's entropy [3], given by

$$H(P) = - \sum_{i=1}^n p_i \log p_i \quad (1.2)$$

But there exist a variety of information theoretic entropies-parametric as well non-parametric investigated by various researchers. To cite a few, we have

$${}_{\alpha}H(P) = \frac{1}{1-\alpha} \log \left( \frac{\sum_{i=1}^n p_i^{\alpha}}{\sum_{i=1}^n p_i} \right), \alpha \neq 1, \alpha > 0 \quad (1.3)$$

which is Renyi's entropy [4].

Following this development, Havrada-Charvat [5] introduced a non-additive entropy, given by

$$H^{\alpha}(P) = \frac{\left[ \sum_{i=1}^n p_i^{\alpha} \right] - 1}{2^{1-\alpha} - 1}, \alpha \neq 1, \alpha > 0 \quad (1.4)$$

Parkash and Mukesh [6, 7] investigated and introduced new generalized parametric measures of entropy for the real parameters, given by the following expressions

$$H_{\alpha}(P) = \frac{1}{\alpha} \sum_{i=1}^n (1 - p_i^{\alpha p_i}), \quad \alpha \neq 0, \alpha > 0 \quad (1.5)$$

$${}^{\alpha}H(P) = \frac{1}{2(1-\alpha)} \sum_{i=1}^n [\alpha(p_i^{1-\alpha} - 1) + \log p_i^{1-\alpha}], \quad \alpha \neq 1, \alpha > 0 \quad (1.6)$$

Furthermore, Parkash and Mukesh [8] provided the applications of these generalized measures to the fields of Statistics for the development of interrelationships between information measures and Chi-square variate. Moreover, Parkash and Mukesh [7] investigated the study of maximum entropy principle to the field of queueing theory.

Fang, Rajasekera and Tsao [9] remarked that entropy optimization is a useful combination of classical entropy theory with mathematical optimization. The resulting entropy optimization models have proved their usefulness with

successful applications in a variety of mathematical areas. Greenberg and Pierskalla [10] presented an approach for solving arbitrary mathematical programs. The authors observed that the surrogate mathematical program is a lesser constrained problem that, in some cases, may be solved with dynamic programming. A similar study of the surrogate mathematical program has been made by Glover [11].

Erlander [12] provided a treatment of entropy constrained linear programs from modelling as well as computational aspects. Some other contributors, due to Templeman and Xingsi [13], Das, Mazumder and Kajal [14] and Yin-Xing [15], concerned the development of many information-theoretic algorithms for solving constrained linear and non-linear programming problems based upon the principle of minimum cross-entropy and surrogate mathematical programming.

In the present communication, a solution of general constrained non-linear programming problem has been provided by formulating it into an equivalent surrogate problem which reached the solution point iteratively by using the maximum entropy formulation. Further a numerical example is given for solving a simple convex non-linear programming problem which ensures the advantage of using maximum entropy technique for solving the problem.

## 2 General Non-linear Programming Problem

Many problems in engineering sciences involve the optimization of non-linear functions  $f$  and/or  $g$  and almost invariably active constraints. Consider such a problem of a generalized constrained non-linear programming problem given by

### **Problem-I:**

$$\text{Minimize } f(x) \tag{2.1}$$

subject to the constraints

$$g_j(x) \leq 0, \quad j = 1, 2, \dots, m \quad (2.2)$$

where  $x$  denotes the vector of real valued variable  $x_i (i = 1, 2, \dots, n)$  and the constraint functions  $g_j (j = 1, 2, \dots, m)$  represent the constraint vector  $g$ .

There are widespread applications of the Problem-I and plenty of techniques for the solution of this problem. The numerical solutions of this type of problems are of great importance for engineering design, synthesis and analysis. Thus, Problem-I has an equivalent surrogate form expressed in the following.

**Problem-II:**

Minimize  $f(x)$

subject to a single constraint

$$\sum_{j=1}^m \mu_j g_j(x) = 0 \quad (2.3)$$

where  $\mu_j (j = 1, 2, \dots, m)$  are non-negative surrogate multipliers termed as surrogate multipliers. Without any loss of generality, the surrogate multipliers  $\mu$  may be assumed to be normalized to unity, that is,

$$\sum_{j=1}^m \mu_j = 1 \quad (2.4)$$

The solution process is set in a probabilistic context by considering the Problem-I with estimating the probability of assigning the event that each constraint is active at the optimum. Denoting these probabilities by  $\mu_j (j = 1, 2, \dots, m)$ , it is obvious that at least one of the constraints must be active so that the condition (2.4) must hold. We reach the solution  $x^*$  of the Problem-I indirectly through a sequence of solutions of the Problem-II. Thus, the problems I and II are equivalent at the solution point.

The Lagrange's function of the Problem-II is given by

$$\Omega = f(x) + \lambda \sum_{j=1}^m \mu_j g_j(x) \quad (2.5)$$

where  $\lambda$  is the Lagrange multiplier associated with the constraint (2.3) in the Problem-II. An essential condition to be satisfied is that  $(x^*, \mu^*, \lambda^*)$  should be a saddle point of the Lagrange's function  $\Omega$  of the Problem-II, that is,

$$\Omega(x, \mu^*, \lambda^*) \geq \Omega(x^*, \mu^*, \lambda^*) \geq \Omega(x^*, \mu, \lambda) \quad (2.6)$$

The left hand inequality of (2.6) gives the minimization over variables  $x$  for specified  $\mu$  and  $\lambda$  whereas the right hand inequality implies the maximization over  $\mu$  and  $\lambda$  for specified  $x$ . In fact  $x$ ,  $\mu$  and  $\lambda$  are related through the stationarity of the Lagrange's function  $\Omega$  with respect to  $x$  as shown in the following expression

$$\frac{\partial \Omega}{\partial x_i} = \frac{\partial}{\partial x_i} f(x) + \lambda \sum_{j=1}^m \mu_j \frac{\partial}{\partial x_i} g_j(x), \quad i = 1, 2, \dots, n \quad (2.7)$$

The saddle point condition (2.6), which is written in terms of Lagrange's function  $\Omega$ , may however be satisfied by iterative means using Problem-II itself, with alternative iterations in  $x$ -space and  $\mu$ -space. The algorithm is as follows:

Choose the initial set of the surrogate multipliers  $\mu^0$  and solve the Problem-II to obtain the values of  $x^0$  (by minimization, corresponding to the left hand inequality in (2.6)). The surrogate multipliers are then updated to  $\mu^1$  (by maximization for fixed  $x^0$ , corresponding to the left hand inequality in (2.6)) and the Problem-II is solved again to give  $x^1$ . The process is repeated until the sequence  $(\mu^0, x^0), (\mu^1, x^1), \dots, (\mu^k, x^k)$ , converges to the solution of the Problem-II and hence of the Problem-I, at  $(\mu^*, x^*)$ . It is evident that the updated surrogate multipliers  $\mu^{k+1}$  must satisfy the normality condition and the surrogate constraint of the Problem-II. Thus

$$\sum_{j=1}^m \mu_j^{k+1} = 1 \quad (2.8)$$

and

$$\sum_{j=1}^m \mu_j^{k+1} g_j(x^{k+1}) = 0 \quad (2.9)$$

For the condition (2.9) to be satisfied we require the values for  $g_j(x^{k+1})$ ,  $j = 1, 2, \dots, m$  which are not known yet. Therefore, we assume that the best estimates for  $g_j(x^{k+1})$  are the values  $g_j(x^k)$  to be used in the equation (2.9), which is modified to the following equation

$$\sum_{j=1}^m \mu_j^{k+1} g_j(x^k) = \varepsilon \quad (2.10)$$

where  $\varepsilon$  is the unknown error introduced due to the replacement of the values  $g_j(x^{k+1})$  by the values  $g_j(x^k)$ ,  $j = 1, 2, \dots, m$ . Furthermore, we would expect  $\varepsilon$  to approach zero as the sequence of the iterations  $x^k, x^{k+1}, \dots$  approaches to  $x^*$ . Therefore, we reached the following assignment problem

Assign  $\mu_j^{k+1}$ ,  $j = 1, 2, \dots, m$  subject to the conditions

$$\sum_{j=1}^m \mu_j^{k+1} = 1 \text{ and } \sum_{j=1}^m \mu_j^{k+1} g_j(x^k) = \varepsilon \text{ with } \varepsilon \rightarrow 0^+$$

Since the surrogating has no meaning for a single constraint, that is,  $m = 1$ , therefore the cases for  $m \geq 2$  are of interest. For the special case  $m = 2$ , the equations (2.8) and (2.10) give a unique assignment

$$\mu_1^{k+1} = \frac{[\varepsilon - g_2(x^k)]}{[g_1(x^k) - g_2(x^k)]} \quad (2.11)$$

and

$$\mu_2^{k+1} = \frac{[g_1(x^k) - \varepsilon]}{[g_1(x^k) - g_2(x^k)]} \quad (2.12)$$

which depends upon the unknown error term  $\varepsilon$ . For the general case  $m > 2$ , the equations (2.8) and (2.10) are insufficient to yield a unique assignment  $\mu^{k+1}$ . In the absence of any explicit criterion of  $\mu$ -maximization (necessary to satisfy the

saddle point condition (2.6)) which would lead to a unique assignment, an artificial criterion must be imposed. However, on the basis of the information available, there is no logical justification for a criterion which favours one specific assignment rather than another. The artificial criterion should correspond to a maximization process in  $\mu$ -space with the conditions (2.8) and (2.10) satisfied alongside, otherwise the criterion should introduce minimize bias into the assignment of  $\mu^{k+1}$ . Hence, it is logical to maximize the entropy of the multipliers  $\mu^{k+1}$  by using the maximum entropy formulation. Thus, the unsolvable assignment problem is replaced by the following solvable problem in the context of Havrada-Charvat's entropy [5].

**Problem-III:**

$$\text{Maximize } H^\alpha(\mu) = \frac{1}{2^{1-\alpha} - 1} \left[ \sum_{j=1}^m (\mu_j^{k+1})^\alpha - 1 \right] \quad (2.13)$$

subject to the constraints

$$\sum_{j=1}^m \mu_j^{k+1} = 1 \quad \text{and} \quad \sum_{j=1}^m \mu_j^{k+1} g_j(x^k) = \varepsilon$$

where  $\mu_j^{k+1} \geq 0$ ,  $j = 1, 2, \dots, m$

The Lagrange's function of the Problem-III is given by

$$\Delta = \frac{1}{2^{1-\alpha} - 1} \left[ \sum_{j=1}^m (\mu_j^{k+1})^\alpha - 1 \right] + \nu \left[ \sum_{j=1}^m \mu_j^{k+1} - 1 \right] + \eta \left[ \sum_{j=1}^m \mu_j^{k+1} g_j(x^k) - \varepsilon \right] \quad (2.14)$$

where  $\nu$  and  $\eta$  are the Lagrange's multipliers associated with the constraints.

Differentiating equation (2.14) with respect to  $\mu_j^{k+1}$  and setting equal to zero, we get

$$(\mu_j^{k+1})^{\alpha-1} = \frac{1 - 2^{1-\alpha}}{\alpha} [\nu + \eta g_j(x^k)] \quad (2.15)$$

Now, taking in particular  $\alpha = 2$ , the equation (2.15) becomes



$$\mu_j^{k+1} = \frac{1}{4} [\nu + \eta g_j(x^k)] \quad (2.16)$$

In view of the equations (2.8), (2.10) and (2.16), we eliminate the Lagrange's multipliers  $\nu$  and  $\eta$  to get the updated surrogate multipliers (See Appendix A), given by

$$\mu_j^{k+1} = \frac{1}{m} \left\{ 1 - \frac{\left[ \sum_{j=1}^m g_j(x^k) - m g_j(x^k) \right] \left[ \sum_{j=1}^m g_j(x^k) - m \varepsilon \right]}{\left[ \sum_{j=1}^m g_j(x^k) \right]^2 - m \sum_{j=1}^m g_j^2(x^k)} \right\}, \quad j = 1, 2, \dots, m \quad (2.17)$$

The equation (2.17) is proposed to be a relationship for updating the surrogate multipliers in the iterative algorithm described above. The error term  $\varepsilon$  is unknown but should approach zero from positive side as iterations approach the optimum. Thus  $\varepsilon$  may be considered as a control parameter whose value should be chosen at each iteration with the desired behaviour.

In the special case of Problem-I with two constraints ( $m = 2$ ), equations (2.11) and (2.12) represent a unique updating formula for  $\mu_1$  and  $\mu_2$  quite independently of the maximum entropy formulation. It may be noted that for  $m = 2$ , the entropy based update formula (2.17) yields the same results as the update in the equations (2.11) and (2.12). This is not surprising, but it does permit the  $\mu$  update formula (2.17) to be viewed as a least biased extension of the unique result represented by the set of the equations (2.11) and (2.12) for  $m = 2$  in the more general realm of  $m > 2$ . This concludes the derivation of an entropy based update formula for the surrogate multipliers.

### 3 Numerical Example

We consider the following example in which an analytical solution for the  $x$ -phase minimization is available for any set of values of  $\mu$ . Consequently, the

example focuses attention on the  $\mu$ -phase iterations which are the centre of the present method. The example is as follows:

$$\text{Minimize } f(x) = \frac{1}{x_1 x_2 x_3} \quad (3.1)$$

subject to the constraints

$$g_1(x) = 2x_1 + x_2 + 3x_3 - 1 \leq 0 \quad (3.2)$$

$$g_2(x) = x_1 + x_2 + x_3 - 1 \leq 0 \quad (3.3)$$

$$g_3(x) = x_1 + 3x_2 + 2x_3 - 1 \leq 0 \quad (3.4)$$

$$x_1, x_2, x_3 > 0$$

The surrogate form of the given problem is

$$\text{Minimize } f(x) = \frac{1}{x_1 x_2 x_3} \text{ subject to}$$

$$\mu_1(2x_1 + x_2 + 3x_3 - 1) + \mu_2(x_1 + x_2 + x_3 - 1) + \mu_3(x_1 + 3x_2 + 2x_3 - 1) = 0$$

which rearranges to

$$\mu_1(2x_1 + x_2 + 3x_3) + \mu_2(x_1 + x_2 + x_3) + \mu_3(x_1 + 3x_2 + 2x_3) - 1 = 0 \quad (3.5)$$

The corresponding Lagrange's function is given by

$$\Omega = \frac{1}{x_1 x_2 x_3} + \lambda [\mu_1(2x_1 + x_2 + 3x_3) + \mu_2(x_1 + x_2 + x_3) + \mu_3(x_1 + 3x_2 + 2x_3) - 1] \quad (3.6)$$

Now  $\frac{\partial \Omega}{\partial x_i} = 0$ ,  $i = 1, 2, 3$  gives the following set of equations

$$\frac{1}{x_1^2 x_2 x_3} = \lambda(2\mu_1 + \mu_2 + \mu_3) \quad (3.7)$$

$$\frac{1}{x_1 x_2^2 x_3} = \lambda(\mu_1 + \mu_2 + 3\mu_3) \quad (3.8)$$

$$\frac{1}{x_1 x_2 x_3^2} = \lambda(3\mu_1 + \mu_2 + 2\mu_3) \quad (3.9)$$

In view of the equations (3.5), (3.7), (3.8) and (3.9), we get the following values of

$x_1$ ,  $x_2$  and  $x_3$  in terms of  $\mu_1$ ,  $\mu_2$  and  $\mu_3$ .

$$x_1 = \frac{1}{3(2\mu_1 + \mu_2 + \mu_3)} \quad (3.10)$$

$$x_2 = \frac{1}{3(\mu_1 + \mu_2 + 3\mu_3)} \quad (3.11)$$

$$x_3 = \frac{1}{3(3\mu_1 + \mu_2 + 2\mu_3)} \quad (3.12)$$

Table 1: Iterations with the updated  $\mu_i$ .

$k$	$\varepsilon$	$\mu_1$	$\mu_2$	$\mu_3$	$x_1$	$x_2$	$x_3$	$g_1$	$g_2$	$g_3$
0	--	1/3	1/3	1/3	0.25	0.20	0.16667	0.20	-0.38333	0.18333
1	0.09	0.41494	0.17691	0.40814	0.23558	0.18352	0.14894	0.10151	-0.43196	0.08403
2	0.0008	0.41624	0.17540	0.40835	0.23536	0.18348	0.14875	0.10047	-0.43240	0.08332
3	0.00004	0.41621	0.17533	0.40846	0.23537	0.18346	0.14875	0.10045	-0.43242	0.08325

Now, we choose the initial set of the surrogate multipliers  $\mu_1$ ,  $\mu_2$  and  $\mu_3$  which would be used in the equations (3.10), (3.11) and (3.12) to obtain the values of  $x_1$ ,  $x_2$  and  $x_3$ . Jaynes maximum entropy formulation shows that in the absence of any extra information about the problem, the least-biased assumption one can choose is that all the constraints should be equally weighted. Thus we have taken  $\mu_1 = \mu_2 = \mu_3 = \frac{1}{3}$  for the initial iteration.

The surrogate multipliers are then updated by using the relationship (2.17) for the repetition of the process till we reach the desired solution of the problem as shown in the Table 1.

Table 1 gives the iterative result and is as exact as the four decimal places accuracy and therefore the minimum value of  $f(x)$  is found to be 155.6858.

## 4 Conclusion

The present work explores the use of information theory in the mathematical programming problem. An algorithm has been developed for solving constrained non-linear programming problem where the original problem is formulated into equivalent surrogate problem and surrogate multipliers are updated iteratively with the use of maximum entropy formulation. The illustrated example confirms that entropy based updates for the surrogate multipliers lead to a fast and accurate solution of the Problem-I. The example which have discussed above is relatively simple and convex problem in which the convergence rate is dependent upon the sequence of the decreasing values specified for  $\varepsilon$ . Other sequences than that used above have been tested and confirmed for convergence. Thus, the entropy maximization appears to be quantitatively sound. With similar arguments, one can provide the applications of entropy measures either by using the standard measures of entropy or by developing new measures of entropy for the solutions of mathematical programming problems. Moreover, in the literature there exist a variety of methods for obtaining the solutions of linear and non-linear programming problems; we have provided the solution of non-linear programming problem only whereas the solutions for the linear programming problems can similarly be provided.

**ACKNOWLEDGEMENTS.** The authors are thankful to the University Grants Commission, New Delhi, for providing financial assistance for the preparation of the manuscript.

## References

- [1] E.T. Jaynes, Information theory and statistical mechanics, *The Physical Review*, **106**, (1957), 620-630.
- [2] J.N. Kapur, *Maximum Entropy Models in Science and Engineering*, Wiley Eastern Limited, New Delhi, 1989.
- [3] C.E. Shannon, A mathematical theory of communication, *Bell System Technical Journal*, **27**, (1948), 379-423, 623-659.
- [4] A. Renyi, On measures of entropy and information, *Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability*, **1**, (1961), 547-561.
- [5] J.H. Havrada and F. Charvat, Quantification methods of classification process: Concept of structural  $\alpha$ -entropy, *Kybernetika*, **3**, (1967), 30-35.
- [6] O. Parkash and Mukesh, New generalized parametric measure of entropy and cross-entropy, *American Journal of Mathematics and Sciences*, **1**, (2012), 91-96.
- [7] O. Parkash and Mukesh, Contribution of maximum entropy principle in the field of queueing theory, *Communications in Statistics--Theory and Methods*, (2014). (Accepted)
- [8] O. Parkash and Mukesh, Relation between information measures and chi-square distribution, *International Journal of Pure and Applied Mathematics*, **84**, (2013), 517-524.
- [9] S-C. Fang, J.R. Rajasekera and H-S.J. Tsao, *Entropy Optimization and Mathematical Programming*, Kluwer Academic Publishers, USA, 1997.
- [10] H.J. Greenberg and W.P. Pierskalla, Surrogate mathematical programming, *Operations Research*, **18**, (1970), 924-939.
- [11] F. Glover, Surrogate constraints, *Operations Research*, **16**, (1986), 741-749.
- [12] S. Erlander, Entropy in linear programs, *Mathematical Programming*, **21**, (1979), 137-150.

- [13] A.B. Templeman and L.A. Xingsi, Maximum-entropy approach to constrained non-linear programming, *Engineering Optimization*, **12**(3), (1987), 191-205.
- [14] N.C. Das, S.K. Mazumder and D.E. Kajal, Constrained non-linear programming: A minimum cross-entropy algorithm, *Engineering Optimization*, **31**(4), (1999), 479-487.
- [15] L. Yin-Xing, Maximum entropy method for linear programming with bounded variables, *Journal of Northwest University*, **35**(5), (2005), 507-510.

## Appendix A.

Calculation for  $\mu_j^{k+1}$ .

Using equation (2.16) in equations (2.8) and (2.10), we get

$$m\nu + \eta \sum_{j=1}^m g_j(x^k) = 4 \quad (\text{A1})$$

$$\nu \sum_{j=1}^m g_j(x^k) + \eta \sum_{j=1}^m g_j^2(x^k) = 4\varepsilon \quad (\text{A2})$$

Solving equations (A1) and (A2), we get

$$\eta = \frac{4 \left[ \sum_{j=1}^m g_j(x^k) - m\varepsilon \right]}{\left[ \sum_{j=1}^m g_j(x^k) \right]^2 - m \sum_{j=1}^m g_j^2(x^k)} \quad (\text{A3})$$

and

$$\nu = \frac{4}{m} \left\{ 1 - \frac{\left[ \sum_{j=1}^m g_j(x^k) \right]^2 - m\varepsilon \sum_{j=1}^m g_j(x^k)}{\left[ \sum_{j=1}^m g_j(x^k) \right]^2 - m \sum_{j=1}^m g_j^2(x^k)} \right\} \quad (\text{A4})$$

Substituting the vales of  $\nu$  and  $\eta$  in equation (2.16), we get

$$\begin{aligned} \mu_j^{k+1} &= \frac{1}{m} \left\{ 1 - \frac{\left[ \sum_{j=1}^m g_j(x^k) \right]^2 - m\varepsilon \sum_{j=1}^m g_j(x^k)}{\left[ \sum_{j=1}^m g_j(x^k) \right]^2 - m \sum_{j=1}^m g_j^2(x^k)} \right\} + \frac{\left[ \sum_{j=1}^m g_j(x^k) - m\varepsilon \right] g_j(x^k)}{\left[ \sum_{j=1}^m g_j(x^k) \right]^2 - m \sum_{j=1}^m g_j^2(x^k)} \\ &= \frac{1}{m} + \frac{\left[ \sum_{j=1}^m g_j(x^k) - m\varepsilon \right] \left[ g_j(x^k) - \frac{\sum_{j=1}^m g_j(x^k)}{m} \right]}{\left[ \sum_{j=1}^m g_j(x^k) \right]^2 - m \sum_{j=1}^m g_j^2(x^k)} \end{aligned}$$

Therefore, we get

$$\mu_j^{k+1} = \frac{1}{m} \left\{ 1 - \frac{\left[ \sum_{j=1}^m g_j(x^k) - m g_j(x^k) \right] \left[ \sum_{j=1}^m g_j(x^k) - m \varepsilon \right]}{\left[ \sum_{j=1}^m g_j(x^k) \right]^2 - m \sum_{j=1}^m g_j^2(x^k)} \right\}$$