

# A Stochastic Augmented Lagrangian Equality Constrained-Based Algorithm for Global Optimization

Ana Maria A. C. Rocha and Edite M. G. P. Fernandes

*Department of Production and Systems, University of Minho, 4710-057 Braga, Portugal*

**Abstract.** This paper presents a numerical study of a stochastic augmented Lagrangian algorithm to solve continuous constrained global optimization problems. The algorithm approximately solves a sequence of bound constrained subproblems whose objective function penalizes equality and inequality constraints violation and depends on the Lagrange multiplier vectors and a penalty parameter. Each subproblem is solved by a population-based method that uses an electromagnetism-like mechanism to move points towards optimality. A comparison with another stochastic technique is also reported.

**Keywords:** Global optimization, augmented Lagrangian, electromagnetism-like mechanism

**PACS:** 02.60.Pn

## INTRODUCTION

This paper presents a numerical study of a stochastic augmented Lagrangian methodology, where the subproblems are solved by a stochastic population-based algorithm, for continuous constrained global optimization. We aim to address the problem in the form:

$$\min f(x) \quad \text{subject to} \quad g(x) \leq 0, \quad h(x) = 0, \quad x \in \Omega, \quad (1)$$

where  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g: \mathbb{R}^n \rightarrow \mathbb{R}^p$  and  $h: \mathbb{R}^n \rightarrow \mathbb{R}^m$  are nonlinear continuous functions and  $\Omega = \{x \in \mathbb{R}^n : lb \leq x \leq ub\}$ . We do not assume that the objective function  $f$  is convex. There may be many local minima in the feasible region. This class of global optimization problems arises frequently in engineering applications. Specially for large scale problems, derivative-free and stochastic methods are the most well-known and used methods.

Methods based on penalty functions are the most used in literature to handle constraints [1, 4, 6, 7]. The constraints violation is combined with the objective function to define a penalty function. This function aims to penalize infeasible solutions by increasing their fitness values proportionally to their level of constraints violation. Static, dynamic, annealing and adaptive penalties are the most popular. Methods based on augmented Lagrangians are common in deterministic type methods for local optimization, but rare when combined with heuristics that rely on a population of points to converge to a global solution.

Here, we aim to show the functionality of an augmented Lagrangian methodology to handle the equality and inequality constraints of the problem (1), where the subproblems are approximately solved by a stochastic global population-based algorithm. Due to its simplicity, the electromagnetism-like (EM) algorithm proposed in [3] is used to obtain the solution of each subproblem. Since the EM algorithm has been designed to find a minimizer which satisfies  $x \in \Omega$ , our subproblem has bound constraints. We implemented the MaxQ method to convert the inequality constraints of the problem (1) into equality constraints [7]. The usual  $l_2$  augmented Lagrangian function is then applied separately to the original equalities and converted equalities. The bound constrained subproblems are approximately solved by the EM algorithm.

## STOCHASTIC AUGMENTED LAGRANGIAN METHOD

Most stochastic methods for global optimization are developed primarily for unconstrained or simple bound constrained problems. Then they are extended to constrained optimization problems using, for example, a penalty technique. This type of technique transforms the constrained problem into a sequence of unconstrained subproblems by penalizing the objective function when constraints are violated. The objective penalty function, in the unconstrained subproblem, depends on a positive penalty parameter that must be updated throughout the iterative process. With most

penalty functions, the solution of the constrained problem is reached for an infinite value of the penalty parameter. An augmented Lagrangian is a more sophisticated penalty function for which a finite penalty parameter value is sufficient to yield convergence to the solution of the constrained problem [2]. We now show the functionality of an augmented Lagrangian function when solving constrained global optimization problems. Practical and theoretical issues from the augmented Lagrangian methodology are used with a stochastic population-based algorithm, the EM algorithm as proposed in [3], to compute approximate solutions of the sequence of bound constrained subproblems.

### Formulation based on equality constraints

Our proposal takes the general problem (1) and uses a method, proposed in [7], therein denoted by MaxQ, to convert inequality constraints into equality constraints as follows:

$$g_j(x) \leq 0 \Leftrightarrow Q_j(x) \equiv [\max(0, g_j(x))]^{q_j} = 0 \quad (2)$$

where  $q_j \geq 1$ ,  $j = 1, \dots, p$  are parameters that depend on the level of constraint violation, since they are crucial to the rate of convergence and solution accuracy. Note that when the constraint is satisfied,  $Q_j(x)$  is zero. The corresponding augmented Lagrangian function of the new reformulation is

$$\mathcal{L}_\rho^E(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \left[ \lambda_i h_i(x) + \frac{\rho}{2} h_i(x)^2 \right] + \sum_{j=1}^p \left[ \mu_j Q_j(x) + \frac{\rho}{2} Q_j(x)^2 \right], \quad (3)$$

where  $\lambda \in \mathbb{R}^m$ ,  $\mu \in \mathbb{R}^p$  are the vectors of Lagrange multipliers associated with  $h(x) = 0$  and  $g(x) \leq 0$  respectively, and  $\rho$  is a positive penalty parameter. Hence, the subproblems, at each iteration  $k$ ,

$$\min_x \mathcal{L}_{\rho^k}^E(x, \lambda^k, \mu^k) \text{ subject to } x \in \Omega \quad (4)$$

are approximately solved using a stochastic global optimization method. The penalty parameter is increased whenever the infeasibility is not reduced; otherwise it is not changed (lines 9-10 in Algorithm 1). Using the ideas proposed in [4], the initial value for  $\rho$  is

$$\rho^1 = \max \left\{ 10^{-6}, \min \left\{ 10, 2|f(x^0)| / (\|\max(0, g(x^0))\|^2 + \|h(x^0)\|^2) \right\} \right\} \quad (5)$$

where  $x^0$  is an arbitrary initial approximation. It has been shown that convergence could be accelerated if  $q_j$ ,  $j = 1, \dots, p$  are updated through the iterative process according to the corresponding constraint violation. The original idea is to define a large  $q_j$  if  $g_j(x) \gg 1$  and reduce until  $q_j \approx 1$  when  $g_j(x) = 0$ . The suggestion is  $q_j = s_0 / (1 + \exp(-s_1 g_j(x)))$  where  $s_0$  is a parameter crucial for the convergence precision. The other parameter  $s_1$  is chosen to define  $q_j = 2$  when  $g_j(x) = 1$ , see [7]. The use of the formula (3) requires an updating scheme for the Lagrange multipliers  $\mu_j$  associated with the constraints  $g_j(x) \leq 0$ ,  $j = 1, \dots, p$ :

$$\mu_j^{k+1} = \min \left\{ \max \left\{ 0, \mu_j^k + \rho^k U_j \right\}, \mu^+ \right\}, j = 1, \dots, p \quad (6)$$

$$\text{where } U_j(x^k) = \begin{cases} Q_j(x^k), & \text{if } g_j(x^k) > 0 \\ Q_j(x^k) - \kappa \mu_j^k, & \text{otherwise} \end{cases}.$$

The decay term  $\kappa \mu_j^k$ ,  $\kappa \geq 1$ , only affects the update when the corresponding constraint is satisfied, decreasing the multiplier. However, it does not affect the updating when the constraint is violated, thus increasing the multiplier. This places more weight in the constraint, forcing the point into the feasible region. Furthermore, the Lagrange multipliers  $\lambda$  associated with equality constraints  $h(x) = 0$  are also updated using first order estimates

$$\lambda_i^{k+1} = \max \left\{ \lambda^-, \min \left\{ \lambda_i^k + \rho^k h_i(x^k), \lambda^+ \right\} \right\}, i = 1, \dots, m. \quad (7)$$

Note that the safeguarded schemes (listed in (6) and (7)) where  $\lambda^- \ll -1$ , and  $\mu^+, \lambda^+ \gg 1$ , are crucial issues to maintain the sequences  $\{\lambda^k\}$  and  $\{\mu^k\}$  bounded.

## Augmented Lagrangian algorithm

Our proposed stochastic augmented Lagrangian algorithm is presented in Algorithm 1. Lines 5-8 of the algorithm show details of the inner iterative process to compute an approximation to the solution of subproblem (4). Since the EM algorithm is based on a population of points, the point which yields the least objective function value, denoted by the best point of the population,  $x(\text{best})$ , after stopping, is taken as the next approximation to the problem (1). The inner iteration counter is represented by  $l$ . This process terminates when the difference between the function value at the best point,  $\mathcal{L}_{\rho^k}^E(x(\text{best}), \lambda^k, \mu^k)$ , and the average of the function values of the population,  $\mathcal{L}_{\text{avg}}^E$ , is under a specified tolerance  $\varepsilon^k$ . This tolerance decreases as outer iterations proceed (see line 4 of the algorithm).

### Algorithm 1 Stochastic augmented Lagrangian algorithm

- 1: **Given:**  $\mu^+ > 0, \lambda^- < \lambda^+, \rho^+ > 0, \varepsilon^* > 0, 0 < \alpha < 1, \gamma > 1, k_{\max}, l_{\max}, \mu^1 \in [0, \mu^+], \lambda^1 \in [\lambda^-, \lambda^+]$
- 2: choose arbitrary  $x^0$  in  $\Omega$ ; compute  $\rho^1$  using (5); set  $k = 1$
- 3: **while**  $\max\{\|Q(x^{k-1})\|, \|h(x^{k-1})\|\} > \varepsilon^*$  and  $k \leq k_{\max}$  **do**
- 4:      $\varepsilon^k = \max\{\varepsilon^*, 10^{-k}\}$ ; set  $l = 1$
- 5:     **while**  $(\mathcal{L}_{\text{avg}}^E - \mathcal{L}_{\rho^k}^E(x(\text{best}), \lambda^k, \mu^k)) > \varepsilon^k$  and  $l \leq l_{\max}$  **do**
- 6:         compute  $x^k$ , an approximation to the global solution of  $\min_x \mathcal{L}_{\rho^k}^E(x, \lambda^k, \mu^k)$  subject to  $x \in \Omega$ , by EM algorithm
- 7:          $l = l + 1$
- 8:     **end while**
- 9:     **if**  $k = 1$  or  $\max\{\|Q(x^k)\|, \|h(x^k)\|\} \leq \alpha \max\{\|Q(x^{k-1})\|, \|h(x^{k-1})\|\}$  **then**  $\rho^{k+1} = \rho^k$
- 10:     **else**  $\rho^{k+1} = \min\{\rho^+, \gamma\rho^k\}$  **end if**
- 11:     update  $\mu_j^{k+1}$  using (6) and  $\lambda_i^{k+1}$  using (7)
- 12:      $k = k + 1$
- 13: **end while**

## The electromagnetism-like mechanism

In this section, we briefly present the ideas of the EM mechanism, proposed in [3], for solving the subproblems. Here, the objective is to compute an approximate minimizer of  $\mathcal{L}_{\rho^k}^E(x, \lambda^k, \mu^k)$ , for fixed values of the parameters  $\rho^k, \lambda^k$  and  $\mu^k$ . For simplicity we use the notation  $\mathcal{L}^k(x) = \mathcal{L}_{\rho^k}^E(x, \lambda^k, \mu^k)$ . Because EM is a population-based algorithm, the inner iterative process begins with a population of  $p_{\text{size}}$  solutions. We also note that our stochastic EM algorithm uses the approximation  $x^{k-1}$  as one of the points of the population to initialize the EM algorithm. The remaining  $p_{\text{size}} - 1$  points are randomly generated. The best found solution, denoted by  $x(\text{best})$ , and the average of function values, are computed by

$$x(\text{best}) = \arg \min \left\{ \mathcal{L}^k(x(s)) : s = 1, \dots, p_{\text{size}} \right\} \text{ and } \mathcal{L}_{\text{avg}}^k = \sum_{s=1}^{p_{\text{size}}} \mathcal{L}^k(x(s)) / p_{\text{size}}, \quad (8)$$

respectively, where  $x(s), s = 1, \dots, p_{\text{size}}$  represent the points of the population. The EM algorithm simulates the electromagnetism theory of physics by considering each point in the population as an electrical charge. The method uses an attraction-repulsion mechanism to move a population of points towards optimality. The main steps of the EM mechanism are shown in Algorithm 2. Details of each step can be found in [3].

### Algorithm 2 EM algorithm

- 1: **Given:**  $x(s), s = 1, \dots, p_{\text{size}}$
- 2: evaluate the population and select  $x(\text{best})$
- 3: compute the charges of the points
- 4: compute the forces between pairs of points
- 5: compute the total force exerted on each point by the others in the population
- 6: move the points except  $x(\text{best})$
- 7: evaluate the new population and select  $x(\text{best})$

**TABLE 1.** Comparison of our results with the best of 5 variants in [1].

Prob.	$f^*$	our study		[1]	
		$f_{\text{best}}$	$f_{\text{avg}}$	$f_{\text{best}}$	$f_{\text{avg}}$
g01	-15.0000	<b>-14.9999</b>	-14.7574	-14.9998	<b>-14.9989</b>
g02	-0.80362	-0.47048	-0.27538	<b>-0.79252</b>	<b>-0.72555</b>
g03	-1.00050	<b>-0.99986</b>	<b>-0.99511</b>	-0.99725	-0.77797
g04	-30665.54	<b>-30665.53</b>	<b>-30665.42</b>	-30665.32	-30578.55
g05	5126.497	5127.503	<b>5159.328</b>	<b>5126.779</b>	5323.866
g06	-6961.814	-6942.999	-6426.668	<b>-6961.448</b>	<b>-6805.229</b>
g07	24.3062	<b>24.3156</b>	<b>24.3328</b>	24.5450	27.8486
g08	-0.09583	<b>-0.09583</b>	<b>-0.09583</b>	<b>-0.09583</b>	-0.08769
g09	680.630	<b>680.633</b>	<b>680.673</b>	680.681	681.470
g10	7049.25	7092.93	<b>7434.13</b>	<b>7070.56</b>	8063.29
g11	0.74990	<b>0.75000</b>	<b>0.77000</b>	0.75217	0.88793

In [1]:  $p_{\text{size}} = 100$ , runs = 25, maximum number of generations = 1000.

## NUMERICAL EXPERIMENTS AND CONCLUSIONS

In this section, we report the results of our numerical study, after running a set of 11 benchmark constrained global problems, described in full detail in [5].

Since the algorithm relies on some random parameters and variables, we solve each problem 30 times and take average of the obtained solutions, herein denoted by  $f_{\text{avg}}$ . The best of the solutions found after all runs is denoted by  $f_{\text{best}}$ . The size of the population depends on  $n$  and we choose  $p_{\text{size}} = \min\{200, 10n\}$ . The fixed parameters are set in this study as follows:  $\lambda^+ = \mu^+ = \rho^+ = 10^{12}$ ,  $\varepsilon^* = 10^{-6}$ ,  $\alpha = 0.5$ ,  $\gamma = 2$ ,  $\lambda^- = -10^{12}$ . We define  $k_{\text{max}} = 50$  and  $l_{\text{max}} = 30$  so that a maximum of 1500 iterations are allowed. We remark that the other conditions in the stopping criteria of the Algorithm 1 (in the outer and inner iterative processes) may cause the termination of the algorithm before reaching the 1500 iterations. The initial multiplier vectors are set to the null vectors. In the augmented Lagrangian of the MaxQ method, we set  $\kappa = 1$ ,  $s_0 = 3$  and  $s_1 = -\ln(s_0/2 - 1)$  as used in [7].

To compare the performance of our stochastic augmented Lagrangian algorithm with another stochastic penalty technique in the literature [1], we report in Table 1 our results and those of the cited paper. In [1], a genetic algorithm combined with an adaptive penalty function is implemented. Table 1 reports  $f_{\text{best}}$  and  $f_{\text{avg}}$  obtained by our study and those of [1] for the eleven problems therein registered (g01-g11). We have better performance (both in  $f_{\text{best}}$  and  $f_{\text{avg}}$ ) than the adaptive penalty algorithm of [1] in six problems (in boldface).

From our preliminary numerical tests, we may conclude that the proposed stochastic augmented Lagrangian algorithm is able to effectively solve constrained problems. The MaxQ method that converts inequality to equality constraints produces solutions with good accuracy and seems to be competitive with a penalty based algorithm. Practical engineering problems, for example, those reported in [1], will be solved in the near future. We also aim to test our algorithms with a point-to-point search yet stochastic method, when solving the bound constrained subproblems.

## REFERENCES

1. H.J.C. Barbosa and A.C.C. Lemonge, *An adaptive penalty method for genetic algorithms in constrained optimization problems*, in *Frontiers in Evolutionary Robotics*, H. Iba (ed.) 34 pages. 2008 (ISBN: 978-3-902613-19-6) I-Tech Education Publ., Austria.
2. D.P. Bertsekas, *Nonlinear Programming*, 2nd edn. Athena Scientific, Belmont, 1999.
3. S.I. Birbil and S.-C. Fang, *An electromagnetism-like mechanism for global optimization*, *Journal of Global Optimization*, 25 (2003), pp. 263–282.
4. E.G. Birgin, R. Castillo, and J.M. Martinez, *Numerical comparison of Augmented Lagrangian algorithms for nonconvex problems*. *Computational Optimization and Applications*, 31 (2004), pp. 31–56.
5. J.J. Liang, T.P. Runarsson, E. Mezura-Montes, M. Clerc, P.N. Suganthan, C.A.C. Coello, and K. Deb, *Problem definitions and evaluation criteria for the CEC2006 special session on constrained real-parameter optimization 2006*.
6. A.M.A.C. Rocha and E.M.G.P. Fernandes, *Self-adaptive penalties in the electromagnetism-like algorithm for constrained global optimization problems*, in *Proceedings of the 8th World Congress on Structural and Multidisciplinary Optimization*, CD 10 pages, Lisbon 2009.
7. T. Wang and B.W. Wah, *Handling inequality constraints in continuous nonlinear global optimization*, in *Integrated Design and Process Science* (1996), pp. 267–274.