

A Modified Augmented Lagrangian Method for a Class of Nonlinear Ill-Posed Problems

Mhbm Shariff

Department of Applied Mathematics and Sciences, Khalifa University, Sharjah, UAE
 Email: shariff@kustar.ac.ae

Received 2013

ABSTRACT

A class of nonlinear problems with real parameters is defined. Generally, in this class of problems, when the parametric values are very large, the problems become ill-posed and numerical difficulties are encountered when trying to solve these problems. In this paper, the nonlinear problems are reformulated to overcome the numerical difficulties associated with large parametric values. A novel iterative algorithm, which is suitable for large scale problems and can be easily parallelized, is proposed to solve the reformulated problems. Numerical tests indicate that the proposed algorithm gives stable solutions. Convergence properties of the proposed algorithm are investigated. In the limiting case, when the corresponding constraint is exactly satisfied, the proposed method is equivalent to the standard augmented Lagrangian method.

Keywords: Iterative Method; Nonlinear; Ill-conditioned; Large Parameter Values; Large-scale

1. Introduction

There are a number of physical problems, where penalty terms (large parameter) occur naturally in the problems. For example, in nonlinear isotropic elasticity, the bulk modulus [1] can be considered as a penalty term for the incompressible constraint. When the penalty term is very large the corresponding constraint is nearly satisfied. Generally, numerical difficulties are encountered, see, e.g., [2], when trying to solve problems with very large penalty values.

In this paper we consider a class of parametric problems that can be reformulated in such a way that the numerical difficulties mentioned above can be overcome. The reformulated problem generally yields indefinite system of equations. When such a system is solved using an existing iterative method, its convergence properties are generally not as good as an iterative method design for a symmetric positive definite system. Here, we proposed a novel iterative algorithm to solve the reformulated problem. The proposed algorithm solves a symmetric positive definite system in each iteration. The convergence properties of the proposed algorithm are investigated and numerical tests are implemented. The algorithm is an extension of the algorithm developed by Shariff [3] for quadratic problems with (near) linear constraints. Our proposed method is suitable for large scale problems in the sense that it only uses a handful of vectors in the algorithm. The proposed algorithm is easily

parallelized and is especially suitable for sparse system of equations.

2. A Class of Nonlinear Constrained Problems

Consider a class of nonlinear problems which is of the form

Problem (I): Minimize

$$f(x) + \varepsilon \sum_{i=1}^m \phi(\chi h_i(x)),$$

where χ is a real constant (can be considered as a penalty term), $\varepsilon = \frac{1}{\chi}$ and $\phi: R \rightarrow R$ with the properties

(a) ϕ is twice continuously differentiable and strictly convex on R

(b) $\phi(0) = 0$, $\phi'(0) = 0$ and $\phi''(0) = 1$

(c) $\lim_{t \rightarrow -\infty} \lim_{t \rightarrow -\infty} \phi'(t) = -\infty$ and $\lim_{t \rightarrow \infty} \lim_{t \rightarrow \infty} \phi'(t) = \infty$.

Examples of the function $\phi(t)$ are:

$$\phi(t) = \frac{t^2}{2} \quad [2],$$

$$\phi(t) = \cosh(t) - 1 \quad [2],$$

$$\phi(t) = \frac{\ln(t+1) + \frac{1}{9}((t+1)^{-9} - 1)}{9} \quad [4],$$

$$\phi(t) = t - \ln(t+1) \quad [5].$$

When χ is very large it can be shown that the corresponding vector constraint

$$\mathbf{h}(\mathbf{x}) = \mathbf{0}$$

($\mathbf{h} = [h_1, h_2, \dots, h_m]^T$) is nearly satisfied and, generally, numerical difficulties are encountered when trying to solve Problem (I). For example, a near incompressible problem often leads to numerical difficulties when a finite element displacement solution is sought. One way to overcome these difficulties is to formulate Problem (I) in an alternative form [3] as given below.

Problem (II): Minimize

$$f(x) + \epsilon \sum_{i=1}^m \phi(g_i(\mathbf{p}))$$

subject to the constrained

$$\hat{\mathbf{h}}(\mathbf{x}) = \epsilon \hat{\mathbf{g}}(\mathbf{p}),$$

where $[\hat{g}_1, \hat{g}_2, \dots, \hat{g}_m]^T = \hat{\mathbf{g}}$, $\hat{\mathbf{h}} = \mathbf{s}(\mathbf{h})$, \mathbf{s} is an invertible function,

$$\epsilon \hat{\mathbf{g}}(\mathbf{p}) = \mathbf{s}(\epsilon(\mathbf{g}(\mathbf{p}))), [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_m]^T = \mathbf{g},$$

and \mathbf{p} is an m -vector variable. We note in several practical problems, the function $\hat{\mathbf{h}}$ is often approximated [1]. An engineering example of Problem (II) can be found in Shariff and Parker [1].

Using the method of Lagrange multiplier, the solution of Problem (II) can be obtained from the following problem.

Problem (III): Find \mathbf{x} , \mathbf{p} and \mathbf{q} such that

$$\nabla_x f(\mathbf{x}) + (\nabla_x \hat{\mathbf{h}}(\mathbf{x}))^T \mathbf{q} = \mathbf{0} \quad (1)$$

$$\hat{\mathbf{h}}(\mathbf{x}) - \epsilon \hat{\mathbf{g}}(\mathbf{p}) = \mathbf{0} \quad (2)$$

$$\epsilon (\nabla_p \mathbf{g}(\mathbf{p}))^T \phi'(\mathbf{g}(\mathbf{p})) - \epsilon (\nabla_p \hat{\mathbf{g}}(\mathbf{p}))^T \mathbf{q} = \mathbf{0}. \quad (3)$$

We note that Problem (II) can be solved using the standard augmented Lagrangian method. In the case when ϵ is very close to zero, it can be easily seen from Equation (3) that the numerical values of \mathbf{p} are not reliable due to computations in non-exact arithmetic [3]. In addition to this, using the standard augmented Lagrangian method introduces an unnecessary variable \mathbf{q} in the formulation. To reduce the number of variables from three to two, we propose a modified augmented Lagrangian method to solve an equivalent Problem (IV) given below. In order to formulate Problem (IV), we assume that $(\nabla_p \hat{\mathbf{g}}(\mathbf{p}))^T$ is nonsingular and hence from Equation (3) we see that the Lagrange multiplier

$$\mathbf{q}(\mathbf{p}) = ((\nabla_p \hat{\mathbf{g}}(\mathbf{p}))^T)^{-1} (\nabla_p \mathbf{g}(\mathbf{p}))^T \phi'(\mathbf{g}(\mathbf{p})), \quad (4)$$

where

$$\phi'(\mathbf{g}) = \begin{pmatrix} \phi'(g_1) \\ \phi'(g_2) \\ \dots \\ \phi'(g_m) \end{pmatrix}. \quad (5)$$

We note that when $\mathbf{g} = \hat{\mathbf{g}}$, we have

$$\mathbf{q} = \phi'(\hat{\mathbf{g}}). \quad (6)$$

On replacing the Lagrange multiplier by a function of \mathbf{p} given in Equation (4) in the Lagrangian function we obtain an equivalent statement:

(IV) Find \mathbf{x} and \mathbf{p} such that $\delta L_0 = 0$, where

$$L_0 = f(x) + \epsilon \sum_{i=1}^m \phi(g_i(\mathbf{p})) + \mathbf{q}^T(\mathbf{p})(\hat{\mathbf{h}}(\mathbf{x}) - \epsilon \hat{\mathbf{g}}(\mathbf{p})), \quad (7)$$

where $\mathbf{q}(\mathbf{p})$ is given by Equation (4).

3. A Modified Augmented Lagrangian Method

Problem (IV) can be solved using a large scale iterative method designed for an indefinite system. However, the convergence properties of most large scale iterative methods for indefinite systems are generally not as good as those iterative methods for symmetric positive definite (SPD) systems. The augmented Lagrangian method, however, solves a SPD system in each iteration and generally the solution is obtained in only a few iterations. Here, we modify this method to solve problem (IV). The modified algorithm solves a SPD system in each iteration and is given by:

1. Set $i = 0$, select \mathbf{p}_i and a penalty parameter $\beta \geq c_i > 0$.
2. Set $\mathbf{q}_i = \mathbf{q}(\mathbf{p}_i)$
3. $\mathbf{x}_i = \min_x L_{c_i}(\mathbf{x}, \mathbf{p}_i)$
4. If $|\hat{\mathbf{h}}(\mathbf{x}_i) - \epsilon \hat{\mathbf{g}}(\mathbf{p}_i)| < tolerance$: stop
5. $\mathbf{q}_{i+1} = \mathbf{q}_i + \phi'(c_i(\hat{\mathbf{h}}(\mathbf{x}_i) - \epsilon \hat{\mathbf{g}}(\mathbf{p}_i)))$
6. $\mathbf{p}_{i+1} = \mathbf{q}^{-1}(\mathbf{q}_{i+1})$
7. Set $i = i + 1$; go to 2

The function

$$L_c(\mathbf{x}, \mathbf{p}) = f(x) + \epsilon \sum_{i=1}^m \phi(g_i(\mathbf{p})) + \mathbf{q}^T(\mathbf{p})(\hat{\mathbf{h}}(\mathbf{x}) - \epsilon \hat{\mathbf{g}}(\mathbf{p})) + \frac{\sum_{i=1}^m \phi(c(\hat{\mathbf{h}}(\mathbf{x}_i) - \epsilon \hat{\mathbf{g}}(\mathbf{p}_i)))}{c}, \quad (8)$$

In step 6, if $\hat{\mathbf{g}} = \mathbf{g}$ then numerical value of \mathbf{p}_{i+1} is generally easier to obtain numerically than when $\mathbf{g} \neq \hat{\mathbf{g}}$. For example, when $\phi(t) = \cosh(t) - 1$, we have $\mathbf{p}_{i+1} = \sin h^{-1}(q_{i+1})$. In the special case when $\phi(t) = t^2/2$ we simply have, for $\hat{\mathbf{g}} = \mathbf{g}$, $\mathbf{p}_{i+1} = \mathbf{q}_{i+1}$.

4. Convergence

In this section we show in Proposition 1, that under a certain conditions the sequences $\{x_i\}$ and $\{p_i\}$ converge to their appropriate values.

Proposition 1: Assume that the sequence $\{x_i\}$ converges to a vector x^* , the sequence $\{q_i\}$ is bounded and $\beta \geq c_{i+1} \geq c_i$. Then the sequence

$$\{q_{i+1} = q_i + \phi'(c_i(\hat{h}(x_i) - \varepsilon \hat{g}(p_i)))\}$$

converges to a vector $q^* = q(p^*)$ which satisfies

$$\nabla_x f(x^*) + \nabla_x \hat{h}(x^*)^T q^* = 0 \tag{9}$$

$$\hat{h}(x^*) - \varepsilon \hat{g}(p^*) = 0 \tag{10}$$

$$q^* = ((\nabla_p \hat{g}(p^*))^T)^{-1} (\nabla_p g(p^*))^T \phi'(g(p^*)). \tag{11}$$

Proof: We only consider exact arithmetic calculations. For a given p_i , x_i minimizes L_{c_i} and hence from Equation (8) and step 3 we have

$$\begin{aligned} & \nabla_x L_{c_i}(x_i, p_i) \\ &= \nabla_x f(x_i) + (\nabla_x \hat{h}(x_i))^T [q_i + \phi'(c_i(\hat{h}(x_i) - \varepsilon \hat{g}(p_i)))] \tag{12} \\ &= \nabla_x L_0(x_i, p_i) = 0. \end{aligned}$$

Here, we assume $\nabla_x h(x)$ [2] has a full row rank. Hence, in view of step 5, the above equation can be rearranged to give

$$q_{i+1} = -(\mathbf{B}_i \mathbf{B}_i^T)^{-1} \mathbf{B}_i \nabla_x f(x_i), \tag{13}$$

where $\mathbf{B}_i = \nabla_x \hat{h}(x_i)$. As $\{x_i\} \rightarrow x^*$ we have $q_{k+1} \rightarrow q^*$. Hence from Equation (13) we get the relation

$$\nabla_x f(x^*) + (\nabla_x \hat{h}(x^*))^T q^* = 0. \tag{14}$$

Since $\{q_i\}$ is bounded and

$$\{q_i + \phi'(c_i(\hat{h}(x_i) - \varepsilon \hat{g}(p_i)))\} \rightarrow q^*,$$

it follows that $\phi'(c_i(\hat{h}(x_i) - \varepsilon \hat{g}(p_i)))$ is bounded. From Step 5 in the proposed algorithm and taking into account that $\{q_i\}$ is a subsequence (except for q_0) of $\{q_{i+1}\}$, we have $\phi(c_i(\hat{h}(x_i) - \varepsilon \hat{g}(p_i))) \rightarrow 0$. In view of the properties of ϕ , we have $\hat{h}(x^*) - \varepsilon \hat{g}(p^*) = 0$. From step 6 and Equation (4) we obtain the relation given in Equation (11).

5. Numerical Test

For the reader to have confidence in the proposed algorithm, a simple numerical test problem is given. Consider the simple test problems:

Test 1: Minimize:

$$F = (x-2)^4 + (x-2y)^2 + \chi \frac{(x^2-y)^2}{2}$$

For this problem we have

$$\phi(t) = t^2/2, h(x) = \hat{h}(x) = x^2 - y \text{ and } g(p) = \hat{g}(p) = p.$$

We reformulate this problem below using the method given in Section 2:

Find x, y and p such that $\delta L_0 = 0$, where

$$L_0 = (x-2)^4 + (x-2y)^2 - \varepsilon \frac{p^2}{2} + p(x^2 - y).$$

The function $L_c(x, y, p)$ given in Equation (8) takes the form

$$L_c = L_0 + c \frac{(x^2 - y - \varepsilon p)^2}{2}.$$

In the third step of the proposed algorithm, the corresponding values of x and y are obtained using a nonlinear conjugate gradient method [6]. The tolerance for the conjugate gradient method is set to 10^{-6} and the tolerance for the proposed algorithm is set to 10^{-6} . We let $c_k = 500$ and the starting value for x, y and p is zero.

The results are tabulated below for several values of (Tables 1-4)

Test 2: Minimize:

$$F = (x-2)^4 + (x-2y)^2 + \varepsilon(\cosh(\chi(x^2 - y)) - 1).$$

For this problem we have $\phi(t) = \cosh(t) - 1$, $h(x) = \hat{h}(x) = x^2 - y$, $g(p) = \hat{g}(p) = p$ and $q(p) = \sinh(p)$.

We reformulate this problem below using the method given in Section 2:

Find x, y and p such that $\delta L_0 = 0$, where

$$L_0 = (x-2)^4 + (x-2y)^2 + \varepsilon(\cos h(p) - 1) + \sin h(p)(x^2 - y - \varepsilon p).$$

Table 1. $\varepsilon = 0$. Number of iterations = 8.

x	0.9456
y	0.8942
p	3.371

Table 2. $\varepsilon = 10^{-6}$. Number of iterations = 8.

x	0.9456
y	0.8942
p	3.371

Table 3. $\varepsilon = 10^{-3}$. Number of iterations = 11.

x	0.9466
y	0.8927
p	3.355

Table 4. $\varepsilon = 10^{-1}$. Number of iterations = 23.

x	1.024
y	0.8103
p	2.390

Table 5. $\varepsilon = 0$. Number of iterations = 3.

x	0.9455
y	0.8940
p	1.9301
q	3.371

Table 6. $\varepsilon = 10^{-6}$. Number of iterations = 5.

x	0.9456
y	0.8941
p	1.930
q	3.371

Table 7. $\varepsilon = 10^{-3}$. Number of iterations = 3.

x	0.9462
y	0.8933
p	1.927
q	3.362

Table 8. $\varepsilon = 10^{-1}$. Number of iterations = 32.

x	1.001
y	0.832
p	1.702
q	2.653

The function $L_c(x, y, p)$ given in Equation (8) takes the form

$$L_c = L_0 + \frac{\cos h(c(x^2 - y - \varepsilon p)) - 1}{c}.$$

The tolerance for the conjugate gradient method is set to 10^{-6} and the tolerance for the proposed algorithm is set to 10^{-6} . We let $c_k = 100$ and the starting value for x , y and p is zero. The results are tabulated below for several values of ε (**Tables 5-8**)

From the above tables we see that the proposed algorithm can also produce solutions for values of ε that are not small. The solutions change smoothly (stable) as

we vary the values of ε . As expected, the rate of convergence for $\varepsilon = 0.1$ is not as good as those for smaller values of ε . We must emphasize that the proposed algorithm is developed primarily for $\varepsilon \ll 1$.

6. Conclusions

To overcome the numerical difficulties for problems associated with high penalty values, we reformulated the problems via the use of Lagrange multipliers. A modified augmented Lagrangian method is proposed to solve the reformulated problem. The proposed algorithm is stable and its solution is shown to converge to its appropriate value. The numerical test done seems to support the theoretical analysis.

REFERENCES

- [1] M. H. B. M. Shariff and D. F. Parker, "An Extension of Key's Principle to Nonlinear Elasticity," *Journal of Engineering Mathematics*, Vol. 37, No. 1-3, 2000, pp. 171-190. [doi:10.1023/A:1004734311626](https://doi.org/10.1023/A:1004734311626)
- [2] D. P. Bertsekas, "Constrained Optimization and Lagrangian Multiplier Methods," *Athena Scientific*, Massachusetts, USA, 1981.
- [3] M. H. B. M. Shariff, "A modified augmented Lagrangian method for a class of constrained problems," *Journal of Computational and Applied Mathematics*, Vol. 151, No. 2, 2003, pp. 257-270. [doi:10.1016/S0377-0427\(02\)00813-0](https://doi.org/10.1016/S0377-0427(02)00813-0)
- [4] R. W. Ogden, "Volume Changes Associated with the Deformation of Rubber-like Solids," *Journal of Mechanics and Physics Solids*, Vol. 24, No. 6, 1976, pp. 323-338. [doi:10.1016/0022-5096\(76\)90007-7](https://doi.org/10.1016/0022-5096(76)90007-7)
- [5] P. J. Blatz, "Finite Elastic Deformation of a Plane Strain Wedge-Shaped Radial Crack in a Compressible Cylinder," *Second-Order Effects in Elasticity, Plasticity and Fluid Dynamics*, Pergamon Press, Oxford, 1964, pp. 145-161.
- [6] M. H. B. M. Shariff, "Parallel Finite Element Software for Nonlinear Stress Problems," *Parallel Computing and Transputer Applications*, IOS Press, Amsterdam, 1992, pp. 1261-1270.