韓國 軍事運營分析 學會誌
第13券 第1號, 1987.6.

A Least Squares Iterative Method For Solving Nonlinear Programming Problems With Equality Constraints

Sok, Yong U.

Abstract

This paper deals with an algorithm for solving nonlinear programming problems with equality constraints. Nonlinear programming problems are transformed into a square sums of nonlinear functions by the Lagrangian multiplier method. And an iteration method minimizing this square sums is suggested and then an algorithm is proposed. Also theoretical basis of the algorithm is presented.

1. Introduction

The least square method for nonlinear programming with constraints is an important tool for solving problems in some fields of engineering and basic sciences. This paper deals with algorithms to find optimal solutions of nonlinear programming problems with equality constraints.

To find solutions by least square method, Morrison(6) suggested an algorithm, in which the nonlinear programming problems with equality constraints were transformed into those with unconstraints. This algorithm makes to be approximated the lower bound of objective function to the minimum value of that.

Jacoby(4) had studied for applicable problems with equality constraints.

Transforming nonlinear functions with equality constraints into unconstrained nonlinear functions by using the Lagrangian multiplier method, we do these into the form of square sum by Morrison's problems.

The conditions for local minimum value of problems being expressed by square sums of nonlinear functions, will be found and the algorithm to find the optimal solution under these conditions will be suggested and then it will be shown that pointwise sequences obtained by the suggested algorithm converge to the local optimal solutions under the above conditions.

2. Problems with Equality Constraints

Consider the nonlinear programming problems of the following form:

Minimize $f(x)$ (2.1)

subject to $g(x) = 0$,

where $f : R^n \rightarrow R^1$, $g: R^n \rightarrow R^m$, $x \in R^n$.

To find the local minimum value of the problems (2.1), use the Lagrangian function:

$$L(x, \lambda) = f(x) + \lambda^t g(x) \quad (2.2)$$

where $\lambda \geqslant 0$ is an m-dimensional column vector denoting the Lagrangian multiplier and λ^t is a transpose of λ.

The gradients of $L(x, \lambda)$ for x and λ are, respectively,

$L_x(x, \lambda) = \nabla f(x) + \nabla g^t(x) \lambda$ and

$L_\lambda(x, \lambda) = g(x)$, where $g(x)$ is a Jacobian nxn matrix with components $\partial g_i / \partial x_j$ for $i = 1, \ldots, m; j=1, \ldots, n$, and $\nabla f(x)$ is an n-dimensional column vector denothing the gradient of f for x.

Hence, to find the local minumum value of problem (2.1) is to find x and λ satisfying m + n equations:

$L_x(x, \lambda) = 0$, and $L(x, \lambda) = 0$.

Instead of solving m + n equations directly, the square sum of the following nonlinear function will be minimized.

3. Derivation of Least Square Method

Define $M_1(x, \lambda)$ as follows:

$$M_1(x, \lambda) = \| L_x(x, \lambda) \|^2 + \| L_\lambda(x, \lambda) \|^2. \quad (3.1)$$

Clearly $M_1(x, \lambda) = 0$ implies $L_x(x, \lambda) =$ and $L_\lambda(x, \lambda) = 0$.

Therefore, we can find the point (x, λ) such that $L_x(x, \lambda) = 0$ and $L_\lambda(x, \lambda) = 0$ by minimizing $M_1(x, \lambda)$.

Let f and g_i (i=1, 2, ..., m) be triple continuously differentiable functions, $\frac{\partial g(x)}{\partial x}$ n x n matrix with (i, j) components $\partial g_i(x)/\partial x_j$, and $L_{x,x}$ Hesisan matrix with (j, k) components $\partial^2 L/\partial x_j \partial x_k$.

The following theorem (3) is a well known method to find the local minimum value in Problem (2.1).

Theorem 3.1. Under the following conditions;

(1) there is at least one solution $(\bar{x}, \bar{\lambda})$ such that $L_x(\bar{x}, \bar{\lambda}) = 0$ and $L_\lambda(\bar{x}, \bar{\lambda}) = 0$, where $\bar{x} \in D$, $\bar{\lambda} \in R^m$ and $D = [x: g_i(x) = 0, i=1, ..., m]$.

(2) the Hessian matrix $L_{x,x}(\bar{x}, \bar{\lambda})$ is positive definite, and

(3) if rank $(\partial g(\bar{x})/\partial x) = m$, we have $f(\bar{x}) < f(x)$ for any $x \in W(x)-\{x\}$, where $W(x)$ is some neighborhood of x.

Define (n + m)–dimensional vector $y_1(x, \lambda)$ and (n + m) x (n + m) matrix $A_1(x, \lambda)$ as

$$Y_1(x, \lambda) = (L_x(x, \lambda), L_\lambda(x, \lambda)) \text{ and} \tag{3.2}$$

$$A_1(x, \lambda) = \begin{bmatrix} L_{x,x}(x, \lambda) & \left(\frac{\partial g(x)}{\partial x}\right)^t \\ \frac{\partial g(x)}{\partial x} & 0 \end{bmatrix} \text{, respectively.}$$

Let us denote an (n + m)–dimensional vector (x, λ) as ν, and the point $(\bar{x}, \bar{\lambda})$, satisfying (1)-(3) as $\bar{\nu}$. Especially, $A_1(x, \lambda)$ is a symmetric matrix.

Let us consider the problem to minimize the least square sum

$$M_1(\nu) = \| L_x(\nu) \|^2 + \| L_\lambda(\nu) \|^2$$

of nonlinear equations.

To find the minimum value of $M_1(\nu)$, an iteration method is suggested as follows:

$$\nu^{(k+1)} = \nu^{(k)} - \alpha_1 \left[\frac{1}{\|A_1((\nu)^{(k)})\|_E^2}\right] Y_1(\nu^{(k)}) A_1(\nu^{(k)}), \tag{3.3}$$

where α_1 is a constant in interval (0, 2) and $\|A_1(\nu)\|_E$ is an Euclidean norm of $A_1(\nu)$.

By using the iteration method (3.3) an algorithm minimizing $M_1(\nu)$ is proposed as the followings to find an optimal solution $\bar{\nu}$.

Algorithm

Step 1. Take an initial value $\nu^{(0)}$.

Let k=0, and determine $\epsilon > 0$ and α_1 in (0, 2).

Step 2. Calculate $\nu^{(k+1)}$ by using (3.3).

Step 3. If $\max_j |x_j^{(k+1)} - x_j^{(k)}| < \varepsilon$ or $\| x_j^{(k+1)} - x_j^{(k)} \| < \varepsilon$,

then terminate: otherwise, let k = k + 1 and go to step 2.

4. Theoretical Basis of the Algorithm

The iteration method (3.3) can be proved by the following theorem.

Theorem 4.1. Under the same condition as Theorem (3.1), for any initial value $\nu^{(0)}$, there is a neighborhood $U_1(\bar{\nu})$ such that the sequence $[\nu^{(k)}]$ is in $U_1(\bar{\nu})$ and converges to the mininum value within a finite number k, and $\nu^{(0)} \in U_1(\bar{\nu})$.

The proof of Theorem (4.1) is induced by Lemma 4.1, 4.2, and 4.3.

Lemma 4.1. Under the same conditions as Theorem (3.1), we have grad $M_1(\nu) = 0$.

Proof. from the condition (1) in Theorem (3.1), we obtain

$$\frac{\partial M_1(\bar{\nu})}{\partial X_j} = 2 L_X(\bar{\nu}) \left(\partial L(\bar{\nu}) / \partial x_j \right)'_X + 2L\lambda (\bar{\nu}) \left(\partial L(\bar{\nu}) / \partial x_j \right)$$

$$= 0 \quad (j = 1, \cdots, n)$$

and

$$\frac{\partial M_1(\bar{\nu})}{\partial X_i} = 2 L_X(\bar{\nu}) \left(g_i(X) \right)'_X = 0 \quad (j = 1, \cdots, m).$$

Let F_j and G_i be Hessian matrices of $\frac{\partial L(\nu)}{\partial x_j}$ (j=1, ..., n) and $\frac{\partial L(\nu)}{\partial \lambda_i}$ (i=1, ..., m), respectively.

Define (n + m) x (n + m)–matrix $C_1(\nu)$ and (n + m) x (n + m)–matrix $H_1(\nu)$ as

$$C_1(\nu) = \sum_{j=1}^{n} \left(\frac{\partial L(\nu)}{\partial x_j} \right) F_j + \sum_{i=1}^{m} \left(\frac{\partial L(\nu)}{\partial \lambda_i} \right) G_i \quad \text{and}$$

$$H_1(\nu) = A_1(\nu)' A_1(\nu) + C_1(\nu), \text{ respectively.}$$

Then, from these definitions, we have the following lemma.

Lemma 4.2. Under the same condition as Theorem (3.1), the following inequality holds:

$$\min_{\|\rho\|=1} \|\rho A_1(\bar{\nu})\|^2 > \| C_1(\bar{\nu}) \|_E.$$

Proof. consider the following conditions:

(4) the n x n matrix $L_{xx}(\nu)$ is positive definite,

(5) the m x n matrix $\frac{\partial g(x)}{\partial x}$ has a maximum rank.

By conditions (4) and (5), for (n + m) x (n + m) matrix $A_1(\nu)$ in (3.2), we have rank $A_1(\nu)$ = n+m. So, we obtain also

$\min_{\|\rho\|=1} \|\rho A_1(\bar{\nu})\|^2 > 0.$ While, the conditions (1) means $\|C_1(\nu)\|_E = 0$, which completes the proof.

If we define d_1 as $d_1 = \min_{\|\rho\|=1} \|\rho A_1(\bar{\nu})\|^2 - \|C_1(\bar{\nu})\|_E$, then the following lemma holds.

Lemma 4.3. For any fixed ν, the following inequality holds:

$$d_1 - \min_{\|\rho\|=1} (\rho, \rho H_1(\nu)) \le \| H_1(\nu) - H_1(\bar{\nu}) \|_E.$$

Proof. For any fixed ν, $(\rho, \rho H_1(\nu))$ is continuous on the compact set $\{\rho : \|\rho\| = 1\}$. So, there exists $\tilde{\rho}$ such that $(\tilde{\rho}, \tilde{\rho} H_1(\nu)) = \min_{\|\rho\|=1} (\rho, \rho H_1(\nu))$.

Hence $d_1 \leq \min_{\|\rho\|=1} \|\rho A_1(\bar{\nu})\|^2 + (\tilde{\rho}, \tilde{\rho} C_1(\bar{\nu})) = \|\tilde{\rho} A_1(\bar{\nu})\|^2 + (\tilde{\rho}, \tilde{\rho} C_1(\nu)) = (\tilde{\rho}, \tilde{\rho} H_1(\bar{\nu}))$.

Therefore, by the Schwarz inequality, we obtain the following inequality:

$$d1 - \min_{\|\rho\|=1} (\rho, \rho H_1(\nu)) \leq (\tilde{\rho}, \tilde{\rho} H_1(\bar{\nu})) - (\tilde{\rho}, \tilde{\rho} H_1(\nu))$$

$$\leq \|\rho\|^2 \| H_1(\tilde{\nu}) - H_1(\nu) \|_E$$

$$= \|H_1(\nu) - H_1(\bar{\nu})\|_E,$$ which completes the proof.

Now Theorem (4.1) will be proved by the previous statements.

Let us define an (n + m)–dimensional vector $h_1^{(k)}$ as $h_1^{(k)} = \nu^{(k)} - \nu$ and let $\beta_1^{(k)} = 1/\|A_1(\nu^{(k)})\|_E^2$

Then, since grad $M_1(\nu) = 2y_1(\upsilon) A_1(\upsilon)$, the iteration method can be written as

$$h_1^{(k+1)} = h_1^{(k)} - (\tfrac{1}{2})\alpha_1 \beta_1^{(k)} \{\text{grad } M_1(\nu^{(k)}) - \text{grad } M_1(\bar{\nu})\} \tag{4.1}$$

for k=1, ..., by (3.3) and lemma 4.1 and $2H_1(\nu)$ is a Hessian matrix of $M_1(\nu)$.

By the Taylor's expansion, we can write (4.1) as

$$h_1^{(k+1)} = h_1^{(k)} (I - J_1^{(k)}), \tag{4.2}$$

where I is an (n + m) x (n + m)–identity matrix and

$$J_1^{(k)} = \alpha_1 \beta_1^{(k)} \int_0^1 H_1(\bar{\nu} + t(\nu^{(k)} - \bar{\nu}))\, dt. \tag{4.3}$$

Let ε be any fixed constant such that

$$0 < \varepsilon < \min \{ d_1, e_1(\alpha_1) \}. \tag{4.4}$$

Then, by the condition (1), we have

$$d_1 = \min_{\|\rho\|=1} \|A_1(\bar{\nu})\|^2 - \|C_1(\bar{\nu})\|_E^2$$

$$= \min_{\|\rho\|=1} \|\rho A_1(\bar{\nu})\|^2$$

and $e_1(\alpha_1) = [(2-\alpha_1)/2(\alpha_1+1)] \|A_1(\bar{\nu})\|_E^2$

Hence $d_1 > 0$ by Lamma (4.2).

By the schwarz inequality, we obtain

$$d_1 = \|\tilde{\rho} A_1(\bar{\nu})\|^2 \leq \|\tilde{\rho}\|^2 \|A_1(\bar{\nu})\|_E^2 = \|A_1(\bar{\nu})\|_E^2$$

Hence, for a constant α_1, $0 < \alpha_1 < 2$, we have

$$e_1(\alpha_1) \geq [(2-\alpha_1)/2(\alpha_1+1)]\, d_1 > 0.$$

For ε satisfying (4.4), there exists a positive constant ∂ such that

$\nu \in U_1(\bar{\nu}) = \{ \nu : \|\nu - \bar{\nu}\| < \delta \}$.

This implies that the following inequalities hold.

$$\| C_1(\nu) \|_E \leq \varepsilon \tag{4.5}$$

$$\| A_1(\bar{\nu}) \|_E^2 - \varepsilon \leq \| A_1(\nu) \|_E^2 \leq \| A_1(\bar{\nu}) \|_E^2 + \varepsilon \tag{4.6}$$

$$\| H_1(\nu) - H_1(\bar{\nu}) \|_E \leq \varepsilon . \tag{4.7}$$

By Lamma (4.3) and (4.7), we have

$$d_1 - \varepsilon \leq \min_{\|\rho\|=1} (\rho, \rho H_1(\nu)). \tag{4.8}$$

If $\nu^{(k)} \in U_1(\bar{\nu})$, then by (4.8),

$$\begin{aligned} d_1 - \varepsilon &\leq \int_0^1 \min_{\|\rho\|=1} (\rho, \rho H_1(\bar{\nu} + t(\nu^{(k)} - \bar{\nu}))\, dt \\ &\leq \int_0^1 (\rho, \rho H_1(\bar{\nu} + t(\nu^{(k)} - \bar{\nu}))\, dt \\ &= (\rho, \rho \int_0^1 H_1(\bar{\nu} + t(\nu^{(k)} - \bar{\nu}))\, dt. \end{aligned}$$

Now from (4.3), we have $\alpha_1 \beta_1^{(k)} (d_1 - \varepsilon) \leq (\rho, \rho J_1^{(k)})$ (4.9)
for $\nu^{(k)} \in U_1(\bar{\nu})$.
Hence $\sigma_1 \equiv \alpha_1 (d_1 - \varepsilon) / (\| A_1(\bar{\nu}) \|_E^2 + \varepsilon)$

$\leq (\rho, \rho J_1^{(k)})$ (by (4.6), and (4.7))

$\leq \alpha_1 (\| A_1(\bar{\nu}) \|_E^2 + 2\varepsilon) / (\| A(\bar{\nu}) \|_E^2 - \varepsilon) \equiv \sigma_2$

(by (4.5), and (4.6))

where $0 < \sigma_1 < \sigma_2 < 2$ for $\nu^{(k)} \in U_1(\bar{\nu})$, and let $q_1 = \max(|1-\sigma_1|, |1-\sigma_2|)$
then $0 < q_1 < 1$ holds. And clearly

$$1 - \sigma_2 \leq (\rho, \rho(I - J_1^{(k)}) \leq 1 - \sigma_1. \tag{4.10}$$

Since $I-J_1^{(k)}$ is a symmetric matrix, we have the following inequality by results of matrices theorem (cf. Varga (10, P11)):

$$\max_{\|\rho\|=1} \| \rho(I - J_1^{(k)}) \| = \max_{\|\rho\|=1} | (\rho, \rho(I - J_1^{(k)}) | \tag{4.11}$$

Clearly, by (4.10) and (4.11), we obtain

$$\max \| \rho(I - J_1^{(k)}) \| \leq q_1 \tag{4.12}$$

If we let $\xi_1^{(k)} = h_1^{(k)} / \| h_1^{(k)} \|$, then the following inequality holds;

$$\| h_1^{(k+1)} \| = \| h_1^{(k)} \| \; \| \xi_1^{(k)} (I - J_1^{(k)}) \| \qquad \text{(by (4.2))}$$

$$\leq \| h_1^{(k)} \| \; (\max_{\| \rho \| = 1} \| \rho (I - J_1^{(k)}) \|)$$

$$\leq q_1 \; \| h_1^{(k)} \|. \qquad \text{(by (4.12))}$$

Hence, $\| \nu^{(k+1)} - \bar{\nu} \| \leq q_1 \; \| \nu^{(k)} - \bar{\nu} \|$ for $k=1, \cdots,$ where $0 < q_1 < 1$.

This implies that, for any $\nu^{(o)}$ in $U_1(\bar{\nu})$, Theorem (4.1) is statisfied.

5. Conclusion

In this paper, nonlinear pgroamming problems (NLP) with equality constraints are transformed into those with unconstraints by the Lagrangian multiplier method. Then instead of direct finding solutions for nonlinear equations obtained from Lagrangian functions, a system of equations is transformed into the form of square sums.

And then an iteration method minimizing the square sums of nonlinear functions is suggested and an algorithm for NLP is proposed. also the theoretical basis of the algorithm is presented. This algorithm indicated good performance on the basis of our computational experiences. (c.f. Appendix) Also, NLP with equality and inequality constraints can be treated in a similar fashion by transforming objective function of nonlinear least square sums under Kunn-Tucker conditions.

References

1. Fiacco A.V. and McCormick G.P., "Computational Algorithm for the Sequential Unconstrained Minimization Technique for Nonlinear Programming, ' 'Management Sci., Vol. 10, pp. 601-617 (1964).
2. Hadley G., Nonlinear and Dynamic Programming, Addison Wesley, Reading, Mass. (1964).
3. Hestenes M.R., "Multiplier and Gradient Method," J. Opt. Th. Appl., Vol. 4, pp. 303-320 (1969).
4 Jacoby S.L.S., Kowalik J.S. and Pizzo J.T., Iterative Methods for Nonlinear Optimization Problems, Prentice-Hall, Inc., Englewood Cliffs, New Jersey(1972).
5. Luenberger D.G., "An Approach to Nonlinear Programming, "J. Opt. Th. Appl., Vol. 11, pp. 219-227 (1973).
6. Morrison D.D., "Optimization by Least Squares," SIAM J., Numer Anal., Vol. 5, pp. 83-88 (1986).
7. Powell M.J.D., A Method for Nonlinear Constraints in Minimization Problems, in Optimization ed. by Fletcher R., Academic Press, New York (1969).
8. Rockafellow R.T. "A Dual Approach to Solving Nonlinear Programming Problem by Unconstrained Optimization," Math. Prog., Vol. 5, pp. 354-373 (1973).

9. Rockafellow R.T., "The Multiplier Method of Hesteness and Powell Applied to Convex Programming," J. Opt. Th. Appl., Vol. 12, pp. 555-562 (1973).

10. Varga R.S., Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, New Jersey (1962).

Appendix (Computational Experiences)

Experiments with the new LSM algorithm were applied to several test problems. All problems were solved by the algorithm under different α_1 values for given initial values. The Euclidean norm between the previous and the current vectors was used for the termination condition. The limit of maximum number of variables is 10, and all computer programs were coded in FORTRAN language and run on the VAX-11/785 system at Air Force Academy.

Problem 1. Min $f(x) = x_1^2 + x_2^2 + x_3^3$,
s.t. $g_1(x) = 0$ where $g_1(x) = x_1 + x_2 + 3x_3 - 1$

Initial Value ($\nu^{(o)}$)	Termination Scalar (e)	α_1 Values				
		0.1	0.55	1.0	1.45	1.9
	0.01		87	54	45	36
	0.001	175	161	99	72	57
(5, 10, 15, -5)	0.0001		237	140	100	78
	0.00001		332	191	135	105
	0.01	31	16	17	16	15
	0.001	162	85	57	44	36
(1, 1, 1, -1)	0.0001		161	98	72	57
	0.00001		256	150	107	83
	0.01		6	6	6	5
	0.001	54	34	29	25	21
(.2, .3, .4, -.5)	0.0001	293	109	70	53	42
	0.00001		204	122	88	69

(.333, .333, .333, -.667): Optimal Solution

Problem 2. Min $f(x) = x_1x_2 + x_2x_3 + x_3x_1$
s.t $g_1(x) = 0$ where $g_1(x) = x_1 + x_2 + x_3 - 3$

Initial Values ($\nu^{(o)}$)	Termination Scalar (e)	α_1 Values				
		0.1	0.55	1.0	1.45	1.9
	0.01	21	16	16	14	13
(.5, .5, .5, -1.5)	0.001	155	65	43	32	26
	0.0001	436	115	70	51	40
	0.00001		178	104	73	57
	0.01	8	28	22	19	16
(.5, 1.5, 1.0,	0.001	219	77	49	36	29
-2.5)	0.0001	500	127	76	55	43
	0.00001		189	110	77	60
	0.01		5	4	2	4
	0.001	36	31	24	2	17
(.9, .9, .9, -1.9)	0.0001	244	81	51	38	30
	0.00001		144	85	61	48

(1, 1, 1, -2): Optimal Solution

Problem 3. Min $f(x) = x_1^2 + x_2^2 + x_3^2$

s.t $g_1(x) = 0$ where $g_1(x) = x_1 + x_2 + 3x_3 - 2$ and

$g_x(x) = 0$ where $g_2(x) = 5x_1 + 2x_2 + x_3 - 5$

Initial Value (ν(o))	Termination Scalar (e)	α_1 Values				
		0.1	0.55	1.0	1.45	1.9
(.5, .5, .3, -.08, -.3)	0.01		10	9	9	10
	0.001	80	75	55	44	37
	0.0001		176	111	82	66
	0.00001		306	182	131	103
(1, .5, .5, -.5, -.5)	0.01	2	11	8	7	6
	0.001	79	31	22	18	15
	0.0001	223	92	65	51	42
	0.00001		223	137	101	80
(.8, .36, .3, -.086, -.3)	0.01					
	0.001		6	7	7	7
	0.0001	63	49	40	34	29
	0.00001		182	116	87	70

(.81, .35, .28, -.0867, -.3067): Optimal Solution