ISSN 1064-5624,Doklady Mathematics, 2006. Vol. 73. No. 3, pp.384–387.

MATHEMATICS

DOI: 10.1134/S1064562406030185

The 2-Factor-Method with a Modified Lagrange Function for Degenerate Constrained Optimization Problems

O.A. Brezhneva[a], **Corresponding Member of the RAS Yu.G. Evtushenko**[b], **and A.A. Tret'yakov**[b,c,d]

Received February 28, 2006

[a] *Department of Mathematics and Statistics, 123 Bachelor Hall, Miami University, Oxford, Ohio, 45056 USA*
e-mail: brezhnoa@muohio.edu
[b] *Dorodnicyn Computing Center, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 119991 Russia*
e-mail: evt@ccas.ru
[c] *System Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland*
[d] *University of Podlasie in Siedlce 3 Maja 54, 08-110 Siedlce, Poland*

A special modification of the Lagrange function is used to reduce a nonlinear optimization problem to a system of nonlinear equations that is singular in the general case. By applying the p-factor transformation, this system is reduced to a new nondegenerate system with the same solution, which can be solved by Newton-type methods. Thus, the method suggested combines a version of the modified Lagrange function (MLF) method proposed by Evtushenko in [3] and Tret'yakov's 2-factor-method (its description can be found, for example, in [1]).

Traditionally, nonlinear systems are divided into two classes: regular and nonregular. Regular systems are those to which the classical implicit function theorem can be applied, and nonregular systems are those to which this theorem is inapplicable. This paper deals with nonregular systems.

Consider the nonlinear optimization problem

$$\min_{x\in X} f(x). \tag{1}$$

Here, the feasible set is $X = \{x \in \mathbb{R}^n \mid g(x) \le 0_m\}$, where 0_m is a zero vector in $\mathbb{R}^m$, $(g(x))^\top = (g_1(x), g_2(x), \dots, g_m(x))$ is a row vector function, and the functions $f(x)$ and $g_j(x)$ map $\mathbb{R}^n$ to $\mathbb{R}$.

The Lagrange function for problem (1) is given by $\mathcal{L}(x, v) = f(x) + v^\top g(x)$, where $v \in \mathbb{R}^m_+$ is a Lagrange multiplier vector. Assuming that $f(x)$ and $g(x)$ are twice continuously differentiable, the gradient and Hessian of the Lagrange function are defined as

$$\begin{aligned} \nabla_x\mathcal{L}(x,v) &= \nabla f(x) + \sum_{i=1}^m v_i \nabla g_i(x), \\ \nabla_{xx}\mathcal{L}(x,v) &= \nabla^2 f(x) + \sum_{i=1}^m v_i \nabla^2 g_i(x). \end{aligned}$$

It is assumed that the solution set $X^* \subset \mathbb{R}^n$ of problem (1) is not empty. In what follows, we also assume that the constraint regularity condition (CRC) is satisfied, in other words, the gradients of the active constraints $\nabla g_i(x^*)$ are linearly independent. This condition guarantees that each $x^* \in X^*$ is associated with a unique Lagrange multiplier vector $v^* \in V^*$ that satisfies $\nabla_x \mathcal{L}(x^*, v^*) = 0_n$ and $v_i^* = 0$ if $g_i(x^*) > 0$, $i = 1, 2, \ldots, m$.

Consider the nonstandard version of the MLF method proposed in [3], in which the modified Lagrange function has the form

$$\mathcal{L}_E(x, \lambda) = f(x) + \frac{1}{2} \sum_{i=1}^{m} \lambda_i^2 g_i(x), \tag{2}$$

where $\lambda^\top = (\lambda_1, \lambda_2, \ldots, \lambda_m)$.

Obviously, the ith component of the Lagrange multiplier vector v is expressed in terms of the ith component of the new vector λ by the formula $v_i = \dfrac{(\lambda_i)^2}{2}$. Thus, the use of λ automatically ensures that the corresponding Lagrange multiplier vector v is nonnegative.

A solution $x^* \in X^*$ is associated with a vector λ^* with components $\lambda_i^* = \pm\sqrt{2v_i^*}$. The vectors x and λ are jointly denoted by the single symbol $w \in \mathbb{R}^{n+m}$. Similarly, the pair $[x^*, \lambda^*]$ is denoted by w^*. Therefore, $\mathcal{L}_E(x, \lambda) = \mathcal{L}_E(w)$. According to the Kuhn–Tucker theorem, w^* satisfies the system

$$G(w) = \begin{bmatrix} \nabla f(x) + \frac{1}{2} \sum_{i=1}^{m} \lambda_i^2 \nabla g_i(x) \\ D(\lambda) g(x) \end{bmatrix} = 0_{m+n}. \tag{3}$$

Here, $D(\lambda)$ is a diagonal matrix whose dimension is determined by the dimension of λ and its ith diagonal element is λ_i. Note that system (3) can generally have an infinite set of solutions even in the neighborhood of w^*. Let $\nabla g(x)^\top$ be the Jacobi matrix of the mapping $g(x)$. For system (3), the Jacobi matrix is given by

$$G'(w) = \left[\begin{array}{c|c} \nabla^2 f(x) + \frac{1}{2} \sum_{i=1}^{m} \lambda_i^2 \nabla^2 g_i(x) & \nabla g(x) D(\lambda) \\ \hline D(\lambda) \nabla g^\top(x) & D(g(x)) \end{array}\right]. \tag{4}$$

For the pair $[x^*, \lambda^*]$, we define the set of active constraints as $I(x^*)$, the set of weakly active constraints as $I_0(x^*)$, and the set of strongly active constraints as $I_+(x^*)$

$$\begin{aligned} I(x^*) &= \{j = 1, 2, \ldots, m \mid g_j(x^*) = 0\}, \\ I_0(x^*) &= \{j = 1, 2, \ldots, m \mid \lambda_j^* = 0,\ \ g_j(x^*) = 0\}, \\ I_+(x^*) &= \{j = 1, 2, \ldots, m \mid \lambda_j^* \neq 0,\ \ g_j(x^*) = 0\}. \end{aligned}$$

When the MLF method is substantiated and analyzed, the CRC condition is usually supplemented with the following conditions:

(i) Strict complementarity (SC) condition; i.e., $\lambda_i^* g_i(x^*) = 0$ for $i = 1, 2, \ldots, m$ and, if $g_i(x^*) = 0$, then $\lambda_i^* \neq 0$ for all $i = 1, 2, \ldots, m$.

(ii) Second-order sufficient optimality conditions: there is a number $\nu > 0$ such that

$$z^\top \nabla_{xx}^2 \mathcal{L}_E(x^*, \lambda^*) z \geq \nu \|z\|^2 \tag{5}$$

for all $z \in \mathbb{R}^n$ satisfying $\nabla g_j(x^*)^\top z \leq 0$, $j \in I(x^*)$.

Assume that the SC condition is fulfilled at the point x^*. Then both $\lambda_i^* = 0$ and $g_i(x^*) = 0$ hold for some index i. Therefore, $I_0(x^*)$ is not empty. In this case, matrix (4) becomes singular at the point w^* and, consequently, system (3) cannot be solved by Newton-type methods. The main goal of this paper is to show that the p-regularity theory can be effectively used in this situation (for the basic principles of the theory, see [4, 5]).

Consider the system of nonlinear equations (3). Let the mapping G be nonregular at the point w^*, in other words, the Jacobi matrix (4) is singular and $\text{rank}(G'(w^*)) = r < n+m$. In this case, w^* is called a degenerate solution to system (3).

The singularity of the matrix $G'(w^*)$ means that there is at least one nonzero vector h such that

$$G'(w^*)h = 0_{m+n}. \tag{6}$$

Obviously, for such a vector h, the solution to system (3) also solves the modified system

$$\Phi(w) = G(w) + G'(w)h = 0_{m+n}. \tag{7}$$

For a singular matrix $G'(w^*)$, Lemma 2 below shows that the matrix $\Phi'(w^*) = G'(w^*) + G''(w^*)h$ is nonsingular and, consequently, the solution w^* to system (7) is locally unique. The nonsingularity of $\Phi'(w^*)$ underlies the construction of the 2-factor-method for solving degenerate systems of nonlinear equations.

Consider the 2-factor-operator $G'(w) + G''(w)h$, $h \in \mathbb{R}^{n+m}$, $\|h\| \neq 0$, where the vector h satisfies the condition

$$\text{rank}\,[G'(w^*) + G''(w^*)h] = n+m. \tag{8}$$

A particular form of h depends on the specific features of system (3). Note that the 2-factor-operator can be defined in different manners (see, e.g., [4, 5]). In this paper, we use the most convenient form.

Definition 1. *The mapping G is called 2-regular at the point w^* with respect to some vector $h \in \mathbb{R}^{n+m}$ if condition* (8) *is satisfied.*

Consider an iterative process for solving system (3), which is called the 2-factor-method:

$$w^{k+1} = w^k - [G'(w^k) + G''(w^k)h]^{-1}[G(w^k) + G'(w^k)h], \qquad k = 0, 1, \ldots, \tag{9}$$

where w^0 is an initial approximation in a sufficiently small neighborhood of w^*.

Theorem 1. *Let w^* be a solution to system* (3), *$U_\varepsilon(w^*)$ be a sufficiently small neighborhood of w^*, and the mapping $G \in C^3(\mathbb{R}^{n+m} \to \mathbb{R}^{n+m})$ be 2-regular at w^* with respect to some nonzero element $h \in \mathbb{R}^{n+m}$ satisfying* (6).

Then the sequence defined by (9) *converges to w^* and satisfies*

$$\|w^{k+1} - w^*\| \leq \alpha \|w^k - w^*\|^2, \tag{10}$$

where $\alpha > 0$ is an independent constant and $w^0 \in U_\varepsilon(w^*)$.

Lemma 2 given below implies that, under the sufficient optimality conditions for problem (1), the mapping G defined by (3) is 2-regular at w^* with respect to some element h. Consequently, the system $G(w) = 0_{m+n}$ can be solved by applying the 2-factor-method. By Theorem 1, the method has a quadratic convergence rate.

Assume that the SC condition may be violated at x^*. Without loss of generality, we also assume that the set $I_0(x^*)$ consists of the first s indices; i.e., $I_0(x^*) = \{1, 2, \ldots, s\}$. The set $I_0(x^*)$ can be numerically determined by applying the zero-element identification procedure [2]. Additionally, we assume that $I_+ = \{s+1, s+2, \ldots, p\}$ and introduce the notation $\ell = m - p$. Since $\lambda_j^* = 0$ and $g_j(x^*) = 0$ for all $j = 1, 2, \ldots, s$, the rows and columns of $G'(w^*)$ with the indices from the $(n+1)$th to $(n+s)$th are all zero. The vector $h \in \mathbb{R}^{n+m}$ is defined as $h^\top = (0_n^\top, 1_s^\top, 0_{m-s}^\top)$, where $1_s^\top$ is an s-dimensional all-one row vector. Consider the mapping

$$\Phi(w) = G(w) + G'(w)h. \tag{11}$$

Lemma 1. *Let V be an $n \times n$ matrix and Q be an $n \times p$ matrix such that the columns of Q are linearly independent and $\langle Vx, x\rangle > 0$ for all $x \in \in \{\ker Q^\top\}\backslash\{0\}$. Suppose that G_N is an $\ell \times \ell$ diagonal matrix of full rank.*

Then the matrix

$$\bar{A} = \begin{pmatrix} V & Q & 0 \\ Q^\top & 0 & 0 \\ 0 & 0 & G_N \end{pmatrix} \tag{12}$$

is nonsingular.

Lemma 2. *Let $f,\ g_i \in C^3(\mathbb{R}^n)$ $(i = 1, 2, \dots, m)$; the mapping Φ be defined by formula* (11)*; and the CRC condition and the second-order sufficient optimality conditions* (5) *be satisfied.*

Then the 2-factor-operator $\Phi'(w) = G'(w) + G''(w)h$ is not degenerate at the point w^*.

The proof follows from Lemma 1 if we set $V = \nabla^2_{xx}\mathcal{L}_E(x^*, \lambda^*)$; $G_N = D(g_N(x^*))$, where $g_N(x) = (g_{p+1}(x), g_{p+2}(x), \dots, g_m(x))^\top$; and $Q = [\nabla g_1(x^*), \dots, \nabla g_s(x^*), \lambda^*_{s+1}\nabla g_{s+1}(x^*), \dots, \dots, \lambda^*_p \nabla g_p(x^*)]$. Then $\Phi'(w^*) = \bar{A}$.

Lemma 2 implies that the 2-factor-method (9) can be used to solve system (3). Applying Theorem 1 to the solution of the original problem (1), we derive the following result.

Theorem 2. *Assume that x^* is a solution of problem* (1)*. Let $f,\ g_i \in C^3(\mathbb{R}^n)$, $i = = 1, 2, \dots, m$; and let the CRC condition and the second-order sufficient optimality conditions* (5) *be satisfied.*

Then there exists a sufficiently small neighborhood $U_\varepsilon(w^*)$ of the Kuhn–Tucker point $w^* = = [x^*, \lambda^*]$ such that estimate (10) *holds for method* (9)*.*

Note that the singularity of $G'(w^*)$ makes it possible to construct a whole class of 2-factor-methods by using the condition $PG'(w^*)h = 0$, where h is a nonzero vector from $\mathbb{R}^{n+m}$ and P is the orthoprojector onto the orthogonal complement of the image of $G'(w^*)$. In this case, the 2-regularity condition reduces to $\operatorname{rank}[G'(w^*) + PG''(w^*)h] = n + m$ and the scheme for the 2-factor-method is written as

$$w^{k+1} = w^k - [G'(w^k) + PG''(w^k)h]^{-1} \times [G(w^k) + PG'(w^k)h]. \tag{13}$$

The convergence theorem also holds for method (13). The only difference is that the vector $h \in \operatorname{Ker} G'(w^*)$ does not need to be calculated. However, an additional matrix P has to be specified in this case.

The method described is illustrated by the following example.

Example. Consider the problem

$$\min_{x \in \mathbb{R}^2} (x_1^2 + x_2^2 + 4x_1x_2) \tag{14}$$

subject to $x_1 \geq 0$ and $x_2 \geq 0$.

It is easy to verify that the point $x^* = (0,0)^\top$ is the solution to this problem with the corresponding Lagrange multiplier $v^* = (0,0)^\top$. In this example, $I_0(x^*) = \{1,2\}$ and the modified Lagrange function is $\mathcal{L}_E(x,\lambda) = x_1^2 + x_2^2 + 4x_1x_2 - \frac{1}{2}\lambda_1^2 x_1 - \frac{1}{2}\lambda_2^2 x_2$. Let $h = (0,0,1,1)^\top$. Then system (3) can be written as

$$G(w) = \begin{bmatrix} 2x_1 + 4x_2 - \frac{1}{2}\lambda_1^2 \\ 2x_2 + 4x_1 - \frac{1}{2}\lambda_2^2 \\ -\lambda_1 x_1 \\ -\lambda_2 x_2 \end{bmatrix} = 0_4. \tag{15}$$

The Jacobi matrix for system (15) is

$$G'(w) = \begin{bmatrix} 2 & 4 & -\lambda_1 & 0 \\ 4 & 2 & 0 & -\lambda_2 \\ -\lambda_1 & 0 & -x_1 & 0 \\ 0 & -\lambda_2 & 0 & -x_2 \end{bmatrix}.$$

This matrix is singular at the point $w^* = (0,0,0,0)^\top$. The mapping G is 2-regular at w^* with respect to the introduced element h, and the scheme for the 2-factor-method is written as

$$\begin{bmatrix} 2 & 4 & -\lambda_1 - 1 & 0 \\ 4 & 2 & 0 & -\lambda_2 - 1 \\ -\lambda_1 - 1 & 0 & -x_1 & 0 \\ 0 & -\lambda_2 - 1 & 0 & -x_2 \end{bmatrix} \begin{bmatrix} x_1^{k+1} - x_1 \\ x_2^{k+1} - x_2 \\ \lambda_1^{k+1} - \lambda_1 \\ \lambda_2^{k+1} - \lambda_2 \end{bmatrix} = - \begin{bmatrix} 2x_1 + 4x_2 - \frac{1}{2}\lambda_1^2 - \lambda_1 \\ 2x_2 + 4x_1 - \frac{1}{2}\lambda_2^2 - \lambda_2 \\ -\lambda_1 x_1 - x_1 \\ -\lambda_2 x_2 - x_2 \end{bmatrix},$$

where $k = 0, 1, \dots$ and $(x_1, x_2, \lambda_1, \lambda_2)^\top = (x_1^k, x_2^k, \lambda_1^k, \lambda_2^k)^\top$. ■

In this example, system (7) has a nonunique solution. Therefore, the method fails to converge globally. However, we can use another version of the method and solve the system $\Psi(w) = G'(w)h = 0_4$, which is linear with respect to w. Here, $h \in \operatorname{Ker} G'(w^*)$. Since G is 2-regular with respect to the element h at the point w^*, the matrix $\Psi'(w^*)$ is nonsingular. Consequently, the problem has the unique solution $w^* = 0_{n+m}$.

ACKNOWLEDGMENTS

This work was based on the joint studies performed at the Dorodnicyn Computing Center, Russian Academy of Sciences (Moscow, Russia), the University of Podlasie (Siedlce, Poland), and the System Research Institute, Polish Academy of Sciences (Warsaw, Poland). This work was supported by the program "Leading Scientific Schools", project no. NSh-2240.2006.1.

REFERENCES

1. K.N. Belash and A.A. Tret'yakov, Zh. Vychisl. Mat. Mat. Fiz. **28**, 1097–1102 (1988).
2. O.A. Brezhneva and A.A. Tret'yakov, *Methods for Solving Essentially Nonlinear Problems* (Vychisl. Tsentr Ross. Akad. Nauk, Moscow, 2000) [in Russian].
3. Yu.G. Evtushenko, J. Optim. Theory Appl. **21** (2), 121–135 (1977).
4. A.F. Izmailov and A.A. Tret'yakov, *Factor Analysis of Nonlinear Mappings* (Nauka, Moscow, 1994) [in Russian].
5. A.A. Tret'yakov and J.E. Marsden, Commun. Pure Appl. Anal. **2**, 425–445 (2003).