P-regular nonlinear optimization

A. A. Tret'yakov*

*Computing Center RAS, tret@ap.siedlce.pl

Applications of the p-regularity theory, successfully developing for the last years, to optimization problems are represented. The main result of this theory gives a detailed description of the structure of the zero set of an irregular nonlinear mapping. Among the applications, the construction of p-factor-operator is used to construct numerical methods for solving degenerate optimization problems and p-order necessary and sufficient optimality conditions are formulated to solve degenerate optimization problems with equality constraints, in which the Lagrange multiplier associated with the objective function might be equal to zero. The p-factor-approach is applied to singular calculus of variations problems, i.e. when constraints are nonregular at the solution point. All results are illustrated by classical examples of optimization problems [1]–[3].

P-order Extremal Principle. Necessary and Sufficiens Conditions for Optimality in Irregular Equality-Constrained Problems. Consider the following equality constraints problem

$$\underset{x\in X}{\text{minimize}}\ \varphi(x) \tag{1}$$

$$\text{subject to } F(x)=0, \tag{2}$$

where $\varphi : X \to \mathcal{R}$ is a sufficiently smooth function in a Banach space X and F is a sufficiently smooth mapping from a Banach space X to a Banach space Y. For the sake of simplicity consider the completely degenerate case, i.e. $F^{(n)}(x^*) \equiv 0$, $n=\overline{1,p-1}$, and X, Y is finite dimensional spaces. Introduce so-called p-factor Lagrange function

$$\begin{aligned} & \mathcal{L}p(x,h,\lambda_0(h),y(h)) = \\ = \quad & \lambda_0(h)\varphi(x) + \langle y(h), F^{(p-1)}(x)[h]^{p-1}\rangle. \end{aligned}$$

Theorem 1 *Let* $\varphi \in C^2(X)$, $F \in C^{p+1}(X)$, $F(x^*)=0$ *and* $\operatorname{Im} F^{(p)}(x^*)[h]^{p-1} = Y$ *for any* $h \in \operatorname{Ker}^p F^{(p)}(x^*)\backslash\{0\}$ *(i.e.* $F^{(p)}(x^*)[h]^p = 0$*). If there exists* $\alpha > 0$ *and* $y(h) \in Y^*$ *such that*

$$\mathcal{L}p_x(x^*,h,1,y(h)) = 0$$

and

$$\mathcal{L}p_{xx}(x^*,h,1,\frac{2}{p(p+1)}\,y(h))[h]^2 \geq \alpha\|h\|^2 \quad \forall\ h \in \operatorname{Ker}^p F^{(p)}(x^*),$$

then x^* *is a strict local minimizer to problem* (1)–(2).

The p-Factor Method for Solving Singular Unconditional Optimization Problems. Consider the following optimization problem

$$\underset{x\in\mathcal{R}^n}{\text{minimize}}\ \varphi(x),$$

where $\varphi \in C^{p+1}(\mathcal{R}^n)$ and at the solution point x^*, $\varphi''(x^*)$ is singular. For example, $\varphi(x) = x_1^4 + x_2^4$ or $\varphi(x) = x_1^2 + x_1^2 x_2 + x_2^4$ and $x^* = (0,0)^\top$. The principal scheme of p-factor method in completely degenerate case is following:

$$\begin{aligned} x^{k+1} \ &= \ x^k - \left\{\varphi^{(p)}(x^k[h]^{p-2}\right\}^{-1} \times \\ &\times\ \left[\varphi'(x^k) + \varphi^{(p-1)}(x^k)[h]^{p-2}\right] \end{aligned} \tag{3}$$

for $k=0,1,\ldots$.

The following result holds

Theorem 2 *Suppose that there exists* $\left\{\varphi^{(p)}(x^*)[h]^{p-2}\right\}^{-1}$ *for some* h, $\|h\|=1$. *Then for* $\varepsilon > 0$ *sufficiently small and for any* $x_0 \in U_\varepsilon(x^*)$ *the sequence* (3) *converges to* x^* *and*

$$\|x^{k+1} - x^*\| \leq c\,\|x^k - x^*\|^2, \quad k=0,1,\ldots,$$

where $c>0$ *is independent constant.*

For $\varphi(x) = x_1^4 + x_2^4$ application of (3) gives as

$$x^{k+1} = -\frac{1}{6}((x_1^k)^3, (x_2^k)^3)^\top$$

or

$$\|x^{k+1} - x^*\| \leq \frac{1}{6}\|x^k - x^*\|^2.$$

Using Newton method we obtain

$$x_N^{k+1} = x_N^k - [\varphi''(x_N^k)]^{-1}\varphi'(x_N^k) = \frac{2}{3}\,x_N^k,$$

i.e. there is not exist quadratic convergence rate of the classical Newton's method.

For function $\varphi(x) = x_1^2 + x_1^2 x_2 + x_2^4$, $x^* = (0,0)^\top$ we could not quarantee convergence in general because at the initial points $x_1^0 = x_2^0\sqrt{1+(x_2^0)^2}$ does't exist $\{\varphi''(x^0)\}^{-1}$.

Application p-regularity theory to degenerate nonlinear programming problems. Consider the nonlinear optimization problem

$$\min_{x \in A} \varphi(x). \tag{4}$$

Here, the feasible set is $A = \{x \in \mathcal{R}^n \mid g(x) \leq 0_m\}$, where 0_m is a zero vector in $\mathcal{R}^m$, $(g(x))^\top = (g_1(x), g_2(x), \ldots, g_m(x))$ is a row vector function, and the function $\varphi(x)$ and $g(x)$ map $\mathcal{R}^n$ to $\mathcal{R}$.

The Lagrange function for problem (4) is given by $\mathcal{L}(x,v) = \varphi(x) + v^\top g(x)$, where $v \in \mathcal{R}_+^m$ is a Lagrange multiplier vector. Assuming that $\varphi(x)$ and $g(x)$ are twice continuously differentiable, the gradient and Hessian of the Lagrange function are defined as

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

It is assumed that the solution set $X^* \subset \mathcal{R}^n$ of problem (4) 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 $g_i(x^*) > 0$ for $i = 1, 2, \ldots, m$.

Consider the non-standard version of the MLF method in which the modified Lagrange function has the form

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

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 $w \in \mathcal{R}^{n+m}$. 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

$$G(w) = \begin{bmatrix} \nabla\varphi(x) + \dfrac{1}{2}\displaystyle\sum_{i=1}^m \lambda_i^2 \nabla g_i(x) \\ D(\lambda) g(x) \end{bmatrix} = 0_{m+n}. \tag{5}$$

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 (5) can generally have an infinite set of solutions even in the neighborhood of w^*. Let $\nabla g^\top(x)$ be the Jacobi matrix of the mapping $g(x)$. For system (5), the Jacobi matrix is given by

$$G'(w) = \left[\begin{array}{c|c} \nabla^2\varphi(x) + \dfrac{1}{2}\displaystyle\sum_{i=1}^m \lambda_i^2 \nabla^2 g_i(x) & \nabla g(x) D(\lambda) \\ \hline D(\lambda)\nabla g^\top(x) & D(g(x)) \end{array}\right]. \tag{6}$$

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

$$z^\top \nabla_{xx}^2 \mathcal{L}_E(x^*, \lambda^*) z \geq \nu \|z\|^2 \tag{7}$$

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

Assume that the SC condition is does't 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 (6) becomes singular at the point w^* and, consequently, system (5) cannot be solved by Newton-type methods.

Consider the system of nonlinear equations (5). Let the mapping G be nonregular at the point w^*, in other words, the Jacobi matrix (6) is singular and $\text{rank}(G'(w^*)) = r < n + m$. In this case, w^* is called a degenerate solution to system (5).

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

$$G'(w^*)h = 0_{m+n}. \tag{8}$$

Obviously, for such a vector h, the solution to system (5) also solves the modified system

$$\Phi(w) = G(w) + G'(w)h = 0_{m+n}. \qquad (9)$$

For a singular matrix $G'(w^*)$, the matrix $\Phi'(w^*) = G'(w^*) + G''(w^*)h$ is nonsingular and, consequently, the solution w^* to system (9) 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 $G'(w) + G''(w)h$, $h \in \mathcal{R}^{n+m}$, $\|h\| \neq 0$, where the vector h satisfies the condition

$$\operatorname{rank}[G'(w^*) + G''(w^*)h] = n + m. \qquad (10)$$

A particular form of h depends on the specific features of system (5). Note that the 2-factor-operator can be defined in different manners. 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 $h \in \mathcal{R}^{n+m}$ if condition* (10) *is satisfied.*

Consider an iterative process for solving system (5), which is called the 2-factor-method:

$$w^{k+1} = w^k - [G'(w^k) + G''(w^k)h]^{-1}[G(w^k) + G'(w^k)h], \qquad (11)$$

where $k = 0, 1, \ldots$, and w^0 is an initial approximation in a sufficiently small neighborhood of w^*.

Theorem 3 *Let w^* be a solution to system* (5)*, $U_\varepsilon(w^*)$ be a sufficiently small neighborhood of w^*, and the mapping $G \in C^3(\mathcal{R}^{n+m} \to \mathcal{R}^{n+m})$ be 2-regular at w^* with respect to some nonzero element $h \in \mathcal{R}^{n+m}$ satisfying* (8).

Then the sequence defined by (11) *converges to w^* and satisfies*

$$\|w^{k+1} - w^*\| \leq \alpha \|w^k - w^*\|^2, \qquad (12)$$

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

Acknowledgements. The work is partially supported by the Russian Foundation for Basic Research, Grants No. 08-01-00619 and No. 11-01-00786-a and by the Leading Scientific Schools, Grant No. 4096.2010.1.

This work was carried out jointly with University of Siedlce, Department of Natural Sciences, Institute of Mathematics and Physics, Siedlce, Poland, and System Research Institute PAS, Warsaw, Poland.

References

[1] A. A. Tret'yakov. Necessary and Sufficient Conditions for Optimality of p-th Order // *Comput. Math. and Math. Phys.*, **24** (1984).

[2] O. A. Brezhneva and A. A. Tret'yakov. Optimality conditions to degenerate extremum problems with equality constraints // *SIAM J. Control and Optim.*, **42** (2003).

[3] A. Prusinska and A. Tret'yakov. P-order necessary and sufficient conditions for optimality in singular calculs of variations // *Discussiones Mathematicue, Differential Inclusions, Control and Optimization*, **30** (2010).