Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 10, No. 1, pp. 151–162, January 2014

SOME NEW IDENTIFICATION FUNCTIONS FOR ACTIVE CONSTRAINTS IN NONLINEAR PROGRAMMING

Zhou-Hong Wang*, Xiaodong Ding† and Yang Yang

Abstract: In this paper we consider the problem of identifying active set for nonlinear programs with inequality constraints. Such an identification is important from both a theoretical and a practical point of view. Based on the work of Facchinei, Fischer and Kanzow (1998), we construct several new identification functions by using a class of NCP functions. The correctness is proved and some numerical results are also presented.

Key words: *constrained optimization, identification of active constraints, degeneracy, identification function*

Mathematics Subject Classification: *90C30, 90C31, 65K05*

---

1 Introduction

Consider the following nonlinear program:

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & c_i(\mathbf{x}) \le 0,\, i \in I = \{1, \ldots, m\} \end{array} \tag{1.1}$$

where $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ and $c_i : \mathbb{R}^n \to \mathbb{R}$ $(i \in I)$ are at least continuously differentiable functions. In this paper we will discuss the correct identification of active constraints near the optimal solution of problem (1.1). Without loss of generality, we omit the equality constraints in (1.1).

The correct identification of active constraints is important in nonlinear programming from both a theoretical and a practical point of view. The locally superlinear convergence properties of algorithms are usually established on the assumption that the active constraints are correctly identified, see, e.g., [1, 8, 16]. Such an identification can also be used to improve local convergence behavior of algorithms. Techniques for identifying active inequality constraints have been studied extensively. Burke and Moré [4, 5] take a geometric approach under certain conditions, where active set identification corresponds to the identification of the face which contains the solution x^*. They also show the application of their results in gradient projection and sequential quadratic programming (SQP)

*The author is supported by NSFC(#10831006) and the Fundamental Research Funds for the Central Universities(No. 2013JBZ005)

†The author is supported by NSFC(#11201425)

algorithms. Facchinei, Fisher and Kanzow [7] present a technique based on identification function which uses estimates of the Lagrange multipliers along with the current x, where neither the strict complementary slackness nor the uniqueness of the multipliers is required. The book [3, Chapter 12] describes a trust region algorithm for convex constrained problems, where active sets are identified when iterates enter some neighborhood of the optimal solution x^* for nondegenerate cases. Based on the work of [7], Oberlin and Wright [14] propose some identification schemes which do not require good estimates of Lagrange multipliers. Then Lewis and Wright [12] generalize the method to composite nonsmooth minimization which includes the classical nonlinear programming as a special case, and provide a simple conceptual framework for general active-set techniques.

The identification of active constraints has also played an important role in interior point methods for linear programming. Ye and Mehrotra [18, 13] propose strategies for identifying active set near the primal-dual interior optimal solution, and prove that interior point methods also possess the finite termination property just as the simplex methods for linear programming. El-Bakry, Tapia and Zhang [6] discuss various indicators for identifying active constraints in interior algorithms for linear programming.

In this paper we focus on the construction of identification functions. The identification function technique for identifying active constraints in nonlinear programming was first proposed by [7], and two identification functions were discussed there. Based on a class of NCP functions, we construct several new identification functions. The correctness of the construction method is proved in detail. Some numerical results are also presented to show the efficiency of these new identification functions.

We conclude this section by introducing notations and some background material used in this paper. Following the usual terminology in constrained programming, the Lagrangian for (1.1) is:

$$L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i c_i(x), \tag{1.2}$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)^T \in \mathbb{R}^m$ are Lagrangian multipliers. Under some constraint qualification, the first order necessary conditions for x^* to be a solution of (1.1) are that there exist multipliers $\lambda^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_m^*)^T$ such that the following KKT (Karush-Kuhn-Tucker) system is satisfied:

$$\begin{cases} \nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum\limits_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) = 0 \\ \lambda_i^* \geq 0, c_i(x^*) \geq 0, i = 1, \ldots, m \\ \lambda_i^* c_i(x^*) = 0, i = 1, \ldots, m \end{cases} \tag{1.3}$$

The pair (x^*, λ^*) is called a *KKT point* of problem (1.1), and $x^* \in \mathbb{R}^n$ is called a *stationary point* of problem (1.1).

Denote

$$B_\varepsilon = \{x \in \mathbb{R}^n \mid \|x\| < \varepsilon\}, \quad B_\varepsilon(x^*) = \{x \in \mathbb{R}^n \mid \|x - x^*\| < \varepsilon\},$$

where $\|\cdot\|$ is the Euclidean norm. In the sequel x^* will always denote a fixed, isolated stationary point, i.e., there exists some $\varepsilon > 0$ such that x^* is the only stationary point in its neighborhood $B_\varepsilon(x^*)$. The "dual" solution set Λ (which means the Lagrange multipliers set) corresponding to x^* and the primal-dual solution set $\mathcal{K}$ are defined as follows:

$$\Lambda = \{\lambda | (x^*, \lambda) \text{ is a KKT point of (1.1)}\}, \quad \mathcal{K} = \{(x^*, \lambda) | \lambda \in \Lambda\}. \tag{1.4}$$

We say that the Mangasarian-Fromovitz constraint qualification (MFCQ) holds at x^* if there exists a vector $d \in \mathbb{R}^n$ such that

$$d^T \nabla c_i(x^*) > 0, \forall i \in I(x^*). \tag{1.5}$$

It is well known that Λ is a compact convex set if and only if MFCQ is satisfied at x^* (see [9]).

Denote

$$\begin{gathered} I(x^*) = \{i | c_i(x^*) = 0, i \in I\} \\ I^+(x^*) = \{i \in I(x^*) \mid \text{ there exists some } \lambda_i^* \in \Lambda \text{ such that } \lambda_i^* > 0\} \\ I^0(x^*) = I(x^*) \backslash I^+(x^*) = \{i \in I(x^*) \mid \lambda_i^* = 0, \forall \lambda_i^* \in \Lambda\} \\ \mathcal{G}(x^*) = \{d \mid d^T \nabla c_i(x^*) = 0, \forall i \in I^+(x^*); d^T \nabla c_i(x^*) \geq 0, \forall i \in I^0(x^*)\}, \end{gathered}$$

where $I(x^*)$ is the index set of *active inequality constraints*, $I^+(x^*)$ is said to be the index set of *strongly active inequality constraints* and $I^0(x^*)$ is called as the index set of *weakly active inequality constraints*. We say that second-order sufficient condition (SOSC) holds at x^* if $f(x)$ and $c_i(x)$ $(i = 1, \ldots, m)$ are all twice continuously differentiable and there exists a positive value $\gamma > 0$ such that

$$v^T \nabla_{xx} L(x^*, \lambda^*) v \geq \gamma \|v\|^2 \quad \text{for all } v \in \mathcal{G}(x^*) \text{ and for all } \lambda^* \in \Lambda. \tag{1.6}$$

$\boxed{2}$ Construction of Identification Functions

In this section we consider how to construct identification functions for identifying active set by a primal-dual point (x, λ) near the KKT point set $\mathcal{K}$ which is defined in formula (1.4). The identification function technique is proposed by [7], where a nonnegative valued function $\rho : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}_+$ is defined, and the predicted active constraints index set is defined as follows:

$$A(x, \lambda) = \{i \in I | c_i(x) \leq \rho(x, \lambda)\}. \tag{2.1}$$

Definition 2.1. Let $\rho(x, \lambda) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}_+$ be a nonnegative valued function. If there exists a positive constant $\varepsilon > 0$ such that

$$A(x, \lambda) = I(x^*) \quad \text{for all } (x, \lambda) \in \mathcal{K} + B_\varepsilon, \tag{2.2}$$

we call the function $\rho(x, \lambda)$ as an *identification function* for $\mathcal{K}$.

The above definition is different from the one given in [7]. However, they are in essence the same for identifying active set, as the following lemma indicates.

Lemma 2.2. *Let $\rho(x, \lambda) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}_+$ be a nonnegative valued function, and* $\operatorname{dist}(x, S)$ *be the usual distance function between the point x and the set S, i.e.*

$$\operatorname{dist}(x, S) = \inf_{y \in S} \|x - y\|. \tag{2.3}$$

If the function $\rho(x, \lambda)$ satisfies the following conditions:

(a) *If* $\operatorname{dist}[(x, \lambda), \mathcal{K}] = 0$*, then* $\rho(x, \lambda) = 0$.

(b) *The function* $\rho(x, \lambda) \to 0$ *as* $\operatorname{dist}[(x, \lambda), \mathcal{K}] \to 0$*, i.e. for any* $\varepsilon > 0$*, there exists a real* $\delta > 0$ *such that for all* (x, λ) *with* $0 < \operatorname{dist}[(x, \lambda), \mathcal{K}] < \delta$*, we have* $\rho(x, \lambda) < \varepsilon$.

(c) *For all* $(x, \lambda) \notin \mathcal{K}$*, we have* $\frac{\rho(x,\lambda)}{\mathrm{dist}[(x,\lambda),\mathcal{K}]} \to +\infty$ *as* $\mathrm{dist}[(x,\lambda),\mathcal{K}] \to 0$.

Then it is an identification function.

The proof procedure of the above Lemma is the same as the one presented in [7, Theorem 2.2], and we omit it here.

Denote

$$\begin{aligned} \psi_1(a,b) &= \psi_{min}(a,b) = \min\{a,b\} \\ \Phi_1(x,\lambda) &= \begin{pmatrix} \nabla_x L(x,\lambda) \\ \psi_1(c(x),\lambda) \end{pmatrix} \end{aligned} \tag{2.4}$$

where $\psi_1(c(x),\lambda)$ is taken componentwise, i.e. $\psi_1(c(x),\lambda) = (\psi_1(c_1(x),\lambda_1), \psi_1(c_2(x),\lambda_2),\ldots,\psi_1(c_m(x),\lambda_m))^T$. The following result has been established in earlier works, see, e.g. [7, Theorem 3.6] and [17, Theorem A.1].

Theorem 2.3. *Suppose that* x^* *is an isolated stationary point of problem (1.1), and the MFCQ (1.5) and the SOSC (1.6) are satisfied at* x^**. Then there are constants* $\varepsilon > 0$, $\kappa_1 > 0$ *and* $\kappa_2 > 0$ *such that*

$$\kappa_1 \mathrm{dist}[(x,\lambda),\mathcal{K}] \le \|\Phi_1(x,\lambda)\| \le \kappa_2 \mathrm{dist}[(x,\lambda),\mathcal{K}], \quad \forall (x,\lambda) \in \mathcal{K} + B_\varepsilon$$

Define

$$\rho_1(x,\lambda) = \|\Phi_1(x,\lambda)\|^\sigma, \sigma \in (0,1). \tag{2.5}$$

Based on Lemma 2.2 and Theorem 2.3, we have the following result which has been given in [7, Theorem 3.7] and [14, Theorem 2.2].

Corollary 2.4. *Let* x^* *be an isolated stationary point of problem (1.1). The MFCQ (1.5) and the SOSC (1.6) are satisfied at* x^**. Then the function* $\rho_1(x,\lambda)$ *defined in (2.5) is an identification function for* $\mathcal{K}$*.*

In the definition of the identification function $\rho_1(x,\lambda)$, the NCP-function $\psi_1(a,b) = \min\{a,b\}$ is used. A function $\psi(a,b) : \mathbb{R}^2 \to \mathbb{R}$ is called a NCP-function (see, e.g. [2, 15, 11]) if and only if

$$\psi(a,b) = 0 \Leftrightarrow a \ge 0, b \ge 0, ab = 0.$$

In the following, we will consider using other NCP-functions to construct identification functions (see, e.g. [2, 15, 11]), and compare their numerical behaviors. The following NCP-functions are considered in this paper:

$$\begin{aligned} &\psi_2(a,b) = \psi_{FB}(a,b) = \sqrt{a^2+b^2} - (a+b), \\ &\psi_3(a,b) = \psi_{CCK}(a,b) = \psi_{FB}(a,b) - \alpha a_+ b_+, \ \alpha > 0, \\ &\psi_4(a,b) = \sqrt{[\psi_2(a,b)]^2 + \alpha(a_+ b_+)^2}, \ \alpha > 0, \\ &\psi_5(a,b) = \psi_{YF}(a,b) = \sqrt{[\psi_2(a,b)]^2 + \alpha[(ab)_+]^4}, \ \alpha > 0, \\ &\psi_6(a,b) = \sqrt{[\psi_2(a,b)]^2 + \alpha[(ab)_+]^2}, \ \alpha > 0. \end{aligned} \tag{2.6}$$

where for $\forall v \in R$, $v_+ = \max\{v, 0\}$. The following result has been established in [15, Lemma 1].

Lemma 2.5. *All* $\psi_j, j \in \{2,3,4,5,6\}$ *are strongly semismooth functions and all* $(\psi_j)^2, j \in \{2,3,4,5,6\}$ *are continuously differentiable.*

For $j = 2, 3, \dots, 6$, define

$$\Phi_j(x, \lambda) = \begin{pmatrix} \nabla_x L(x, \lambda) \\ \psi_j(c(x), \lambda) \end{pmatrix}, \tag{2.7}$$

$$\rho_j(x, \lambda) = \|\Phi_j(x, \lambda)\|^\sigma, \quad \sigma \in (0, 1), \tag{2.8}$$

where the functions $\psi_j(c(x), \lambda)$ $(j = 2, 3, \dots, 6)$ are all taken componentwise, i.e. $\psi_j(c(x), \lambda) = (\psi_j(c_1(x), \lambda_1), \psi_j(c_2(x), \lambda_2), \dots, \psi_j(c_m(x), \lambda_m))^T$, where $\psi_j(\cdot, \cdot)$ $(j = 2, \dots, 6)$ are defined in formula (2.6). Below we will prove that the functions $\rho_j(x, \lambda)$ $(j = 2, 3, \dots, 6)$ are all identification functions.

Lemma 2.6. *Let functions $\psi_j(a, b)$ $(j = 2, 3, \dots, 6)$ be defined in formula (2.6). Define $D_M = \{(a, b) \in R^2 \mid |a| \le M, |b| \le M$ and $|ab| \le 1\}$. Then for all $(a, b) \in D_M$, there exist positive constants κ_1, κ_2 with $1 \ge \kappa_1 > 0$ and $\kappa_2 > 1$ such that for all $j \in \{2, 3, \dots, 6\}$, we have*

$$\kappa_1 |\psi_1(a, b)| \le |\psi_j(a, b)| \le \kappa_2 |\psi_1(a, b)|. \tag{2.9}$$

Proof. First consider the function $\psi_2 = \sqrt{a^2 + b^2} - (a + b)$. There are two cases to consider, i.e. $ab \le 0$ and $ab > 0$. In case of $ab \le 0$, we may assume that $a \ge 0, b \le 0$ without loss of generality. Then we get

$$|\psi_2| = |\sqrt{a^2 + b^2} - (a + b)| = (\sqrt{a^2 + b^2} - a) - b = (\sqrt{a^2 + b^2} - a) + |b|. \tag{2.10}$$

Since $0 \le \sqrt{a^2 + b^2} - a \le |b|$ for $a \ge 0$, by (2.10) we get

$$|\psi_1| = |b| \le |\psi_2| = (\sqrt{a^2 + b^2} - a) + |b| \le 2|b| = 2|\psi_1|. \tag{2.11}$$

In case of $ab > 0$, we have the following two cases to consider.

1. If $a < 0, b < 0$, we have $|\psi_2| = \sqrt{a^2 + b^2} - a - b = \sqrt{a^2 + b^2} + |a| + |b|$ and $\max\{|a|, |b|\} = |\min\{a, b\}| = |\psi_1|$. So we get

$$|\psi_2| = \sqrt{a^2 + b^2} + |a| + |b| \le (2 + \sqrt{2}) \max\{|a|, |b|\}| = (2 + \sqrt{2})|\psi_1|, \tag{2.12}$$

and

$$|\psi_2| = (\sqrt{a^2 + b^2}) + (|a| + |b|) \ge 2 \max\{|a|, |b|\}| = 2|\psi_1|. \tag{2.13}$$

2. If $a > 0, b > 0$, we may assume $a \ge b$ without loss of generality. Then we have

$$|\psi_2| = (a + b) - \sqrt{a^2 + b^2} = \frac{2ab}{\sqrt{a^2 + b^2} + a + b} \ge \frac{2ab}{(2 + \sqrt{2})a} = \frac{2}{2 + \sqrt{2}} |\psi_1|, \tag{2.14}$$

and

$$|\psi_2| = b + (a - \sqrt{a^2 + b^2}) \le b = |\psi_1|, \tag{2.15}$$

Hence by (2.11)–(2.15), for all cases we have

$$\frac{2}{2 + \sqrt{2}} |\psi_1| \le |\psi_2| \le (2 + \sqrt{2})|\psi_1|. \tag{2.16}$$

Next we consider the function $\psi_3(a, b) = \psi_2(a, b) - \alpha a_+ b_+ (\alpha > 0)$. If $a \le 0$ or $b \le 0$, we have $|\psi_3| = |\psi_2|$ and the result follows from (2.16). Otherwise if $a > 0, b > 0$, we have

$$|\psi_3| = |\sqrt{a^2 + b^2} - (a + b) - \alpha ab| = \alpha ab + a + b - \sqrt{a^2 + b^2} \ge |\psi_2(a, b)|, \tag{2.17}$$

Since $|a| \le M, |b| \le M$, we have $\sqrt{a^2+b^2}+a+b \le (2+\sqrt{2})M$. So we get

$$|\psi_2| = |(a+b) - \sqrt{a^2+b^2}| = |\frac{2ab}{\sqrt{a^2+b^2}+a+b}| \ge \frac{2}{(2+\sqrt{2})M}|ab|. \tag{2.18}$$

By (2.18) we obtain

$$|\psi_3| \le \alpha|ab| + |\psi_2| \le \left[1 + \frac{(2+\sqrt{2})M\alpha}{2}\right]|\psi_2|. \tag{2.19}$$

By (2.17)–(2.19) and (2.16) we have (2.9) for $j = 3$.

Finally we consider the functions ψ_4, ψ_5 and ψ_6. According to their definitions in (2.6), we have $|\psi_j| \ge |\psi_2|$ for $j = 4, 5, 6$. Since $|ab| \le 1$, we have $[(ab)_+]^4 \le [(ab)_+]^2 \le |ab|^2$ and $(a_+b_+)^2 \le |ab|^2$. Then by (2.18) it is easy to see that there exists a constant $\kappa_3 > 1$ such that $|\psi_j| \le \kappa_3|\psi_2|$ for $j = 4, 5, 6$. Therefore by (2.16) we get (2.9) for $j = 4, 5, 6$. □

Based on Theorem 2.3 and Lemma 2.6, we can get the following result.

Theorem 2.7. *Let x^* be an isolated stationary point of problem (1.1), and $\Lambda, \mathcal{K}$ be defined in (1.4). If the MFCQ (1.5) and the SOSC (1.6) are satisfied at x^*, then the functions $\rho_j(x, \lambda), j = 2, 3 \ldots, 6$ defined in (2.8) are all identification functions for $\mathcal{K}$.*

Proof. By the properties of NCP-functions (see, e.g. [15] and references therein), we know that $\Phi_j(x, \lambda) = 0 \Leftrightarrow (x, \lambda) \in \mathcal{K}$. Since the MFCQ (1.5) is satisfied, we know that the set Λ is a compact convex set by the Theorem in [9], and so is the set $\mathcal{K}$ by the isolation of x^*. Hence there exist positive constants $\varepsilon_1 > 0$ and $M > 0$ such that for all $(x, \lambda) \in \mathcal{K} + B_{\varepsilon_1}$, we have $c_i(x) \le M, \lambda_i \le M$ and $\lambda_i c_i(x) \le 1$ $(i = 1, \ldots, m)$. Then by Lemma 2.6 there exists positive constants κ_1, κ_2 with $0 < \kappa_1 \le 1, \kappa_2 > 1$ such that for all $i = 1, \ldots, m$ and all $j = 2, 3, \ldots, 6$, we have

$$\kappa_1|\psi_1(c_i(x), \lambda_i)| \le |\psi_j(c_i(x), \lambda_i)| \le \kappa_2|\psi_1(c_i(x), \lambda_i)|. \tag{2.20}$$

By (2.7) for $j = 2, 3, \ldots, 6$ we have

$$\|\Phi_j(x, \lambda)\|^2 = \|\nabla_x L(x, \lambda)\|^2 + \sum_{i=1}^{m} |\psi_j(c_i(x), \lambda_i)|^2. \tag{2.21}$$

Note that $0 < \kappa_1 \le 1$ and $\kappa_2 > 1$, by (2.20) and (2.21) for all $(x, \lambda) \in \mathcal{K} + B_{\varepsilon_1}$ we have

$$\kappa_1\|\Phi_1(x, \lambda)\| \le \|\Phi_j(x, \lambda)\| \le \kappa_2\|\Phi_1(x, \lambda)\|, \; j = 2, 3, \ldots, 6. \tag{2.22}$$

Then by Theorem 2.3 and the above formula, there exist constants $\varepsilon_2, \kappa_3, \kappa_4$ with $\varepsilon_1 \ge \varepsilon_2 > 0$, $1 > \kappa_3 > 0$ and $\kappa_4 > 1$ such that for all $(x, \lambda) \in \mathcal{K} + B_{\varepsilon_2}$ and all $j = 2, 3, \ldots, 6$, we have

$$\kappa_3 \mathrm{dist}[(x, \lambda), \mathcal{K}] \le \|\Phi_j(x, \lambda)\| \le \kappa_4 \mathrm{dist}[(x, \lambda), \mathcal{K}] \tag{2.23}$$

Then by Lemma 2.2 we get the result.

□

Since $\psi_1^2(a, b)$ is not continuously differentiable everywhere, nor is the identification function $\rho_1(x, \lambda)$. According to the Lemma 2.5 and other results presented in [15], we know that identification functions $\rho_j(x, \lambda), j = 2, 3 \ldots, 6$ defined in (2.8) all are continuous differentiable. Hence identification functions $\rho_j(x, \lambda), j = 2, 3 \ldots, 6$ possess better smooth property. In the following section we will consider the numerical behaviors of the six identification functions discussed above.

3 Some Numerical Results

The aim of this section is to illustrate and compare the effectiveness of the above six identification functions by numerical tests. Here we use the three test problems presented in [7, Section 4], where the (approximately) optimal solution x^* and the corresponding sets $\Lambda, \mathcal{K}$ are all known. For each problem, random points (x, λ) at different distances from the solution set $\mathcal{K}$ were generated in the following ways. Let r be the parameter controlling the distance, s_j $(j = 1, \dots, n)$ and w_i $(i = 1, \dots, m)$ be the $n+m$ random variables drawn from the uniform distribution on $(-1, 1)$, the perturbed point (x, λ) is defined as follows:

$$\begin{aligned} x_j &= x_j^* + (\frac{s_j}{|s_j|} + 0.2s_j)r, j = 1, \dots, n, \\ \lambda_i &= \lambda_i^* + (\frac{w_i}{|w_i|} + 0.2w_i)r, i = 1, \dots, m \end{aligned} \tag{3.1}$$

For each $r \in \{0.001, 0.01, 0.05, 0.1, 0.5, 1, 10\}$ and each test example, we generated 300 random vectors (x, λ) in such a way. Then we use identification functions ρ_1–ρ_6 which are defined in (2.5) and (2.8) to identify the active constraints at these points. In all tests, we set the parameter $\sigma = \frac{1}{2}$ for ρ_1–ρ_6 * and the parameter $\alpha = 1$ for ρ_3–ρ_6, and compare the true active set $I(x^*)$ with the predicted active set $\mathcal{A}(x, \lambda)$ at each randomly generated point (x, λ). If $\mathcal{A}(x, \lambda) = I(x^*)$, we say that the whole active set is correctly identified at point (x, λ). Unlike the numerical results listed in [7, Section 4], we only recorded the sum of the correct identifications of *the whole active set* in order to compare the performance of the six identification functions ρ_1–ρ_6. For each identification function ρ_i $(i = 1, \dots 6)$ and each distance controlling parameter $r \in \{0.001, 0.01, 0.05, 0.1, 0.5, 1, 10\}$, we listed the sum of correct identifications of *the whole active set* over the 300 randomly generated points (x, λ) in Table 1–3 for Test example 1–3. Then we drew them in Figure 1–3.

Test example 1 is the problem 113 of [10]. It is a convex optimization problem with $n = 10$ variables and $m = 8$ inequality constraints. Its objective function f and the constraints $c_i(x)$ $(i = 1, \dots, 8)$ are defined as follows:

$$\begin{aligned} f(x) &= x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2 + 4(x_4 - 5)^2 + (x_5 - 3)^2 \\ &\quad + 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 - 10)^2 + (x_{10} - 7)^2 + 45 \\ c_1(x) &= -4x_1 - 5x_2 + 3x_7 - 9x_8 + 105 \geq 0 \\ c_2(x) &= -10x_1 + 8x_2 + 17x_7 - 2x_8 \geq 0 \\ c_3(x) &= 8x_1 - 2x_2 - 5x_9 + 2x_{10} + 12 \geq 0 \\ c_4(x) &= -3(x_1 - 2)^2 - 4(x_2 - 3)^2 - 2x_3^2 + 7x_4 + 120 \geq 0 \\ c_5(x) &= -5x_1^2 - 8x_2 - (x_3 - 6)^2 + 2x_4 + 40 \geq 0 \\ c_6(x) &= -0.5(x_1 - 8)^2 - 2(x_2 - 4)^2 - 3x_5^2 + x_6 + 30 \geq 0 \\ c_7(x) &= -x_1^2 - 2(x_2 - 2)^2 + 2x_1x_2 - 14x_5 + 6x_6 \geq 0 \\ c_8(x) &= 3x_1 - 6x_2 - 12(x_9 - 8)^2 + 7x_{10} \geq 0 \end{aligned}$$

†* For each identification function, each test example and each distance $r \in \{0.001, 0.01, 0.05, 0.1, 0.5, 1, 10\}$, the numerical performance are usually the best for $\sigma = \frac{1}{2}$ among values of $\sigma \in (0, 1)$, e.g., $\sigma = 0.1, 0.2, \dots, 0.9$. Hence we only give numerical results for $\sigma = \frac{1}{2}$ here for brevity

Its solution and corresponding active set are

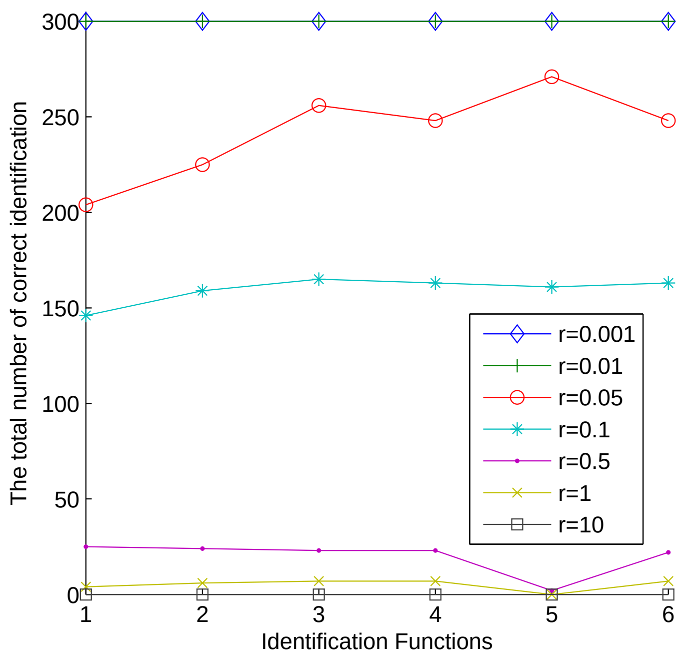
$$x^* \approx (2.17, 2.36, 8.77, 5.10, 0.99, 1.43, 1.32, 9.83, 8.28, 8.38)^T,$$
$$\lambda^* \approx (1.72, 0.48, 1.38, 0.02, 0.31, 0, 0.29, 0)^T,$$
$$I(x^*) = \{1, 2, 3, 4, 5, 7\}.$$

Our results are summarized in Table 1 and Figure 1. Note the numbers in Table 1 and Figure 1 are the sum of correct identifications of *the whole active set* over the 300 randomly generated vectors (x, λ).

Table 1: Numerical results for Example 1

ρ	$r = 0.001$	$r = 0.01$	$r = 0.05$	$r = 0.1$	$r = 0.5$	$r = 1$	$r = 10$
ρ_1	300	300	204	146	25	4	0
ρ_2	300	300	225	159	24	6	0
ρ_3	300	300	256	165	23	7	0
ρ_4	300	300	248	163	23	7	0
ρ_5	300	300	271	161	2	0	0
ρ_6	300	300	248	163	22	7	0

Figure 1: Numerical results for Example 1



Test example 2 is modified from the problem 46 of [10]. Its objective function f and the constraints $c_i(x)$ $(i = 1, 2, 3)$ are defined in [7, Section 4] as follows:

$$\begin{aligned} f(x) &= (x_1 - x_2)^2 + (x_3 - 1)^2 + (x_4 - 1)^4 + (x_5 - 1)^6 \\ c_1(x) &= x_1^2 x_4 + \sin(x_4 - x_5) - 1 \geq 0 \\ c_2(x) &= x_2 + x_3^4 x_4^2 - 2 \geq 0 \\ c_3(x) &= 1 - x_2 \geq 0 \end{aligned}$$

Thus we have $m = 3$ and $n = 5$. The optimal solution is $x^* = (1, 1, 1, 1, 1)^T$ and the corresponding Lagrange multiplier vector is $\lambda^* = (0, 0, 0)^T$. It is easy to check that all constraints are active at x^*, and (x^*, λ^*) is totally degenerate. For each $r \in \{0.001, 0.01, 0.05, 0.1, 0.5, 1, 10\}$ and each identification function ρ_i $(i = 1, \ldots, 6)$, the sum of the correct identifications of *the whole active set* is recorded in Table 2 and drawn in Figure 2 over 300 randomly generated points (x, λ).

Table 2: Numerical results for Example 2

ρ	$r = 0.001$	$r = 0.01$	$r = 0.05$	$r = 0.1$	$r = 0.5$	$r = 1$	$r = 10$
ρ_1	300	300	300	285	196	177	0
ρ_2	300	300	300	287	196	177	0
ρ_3	300	300	300	288	196	177	0
ρ_4	300	300	300	288	196	177	0
ρ_5	300	300	300	287	196	228	150
ρ_6	300	300	300	288	196	177	0

Test example 3 is modified from the problem 43 of [10], which is defined in [7, Section 4] as follows:

$$\begin{array}{ll} \min & f(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4 \\ \text{s.t.} & c_1(x) = -x_1^2 - x_2^2 - x_3^2 - x_4^2 - x_1 + x_2 - x_3 + x_4 + 8 \geq 0 \\ & c_2(x) = -x_1^2 - 2x_2^2 - x_3^2 - 2x_4^2 + x_1 + x_4 + 10 \geq 0 \\ & c_3(x) = -2x_1^2 - x_2^2 - x_3^2 - 2x_1 + x_2 + x_4 + 5 \geq 0 \\ & c_4(x) = x_1^3 + 2x_2^2 + x_4^2 + x_1 - 3x_2 - x_3 + 4x_4 + 7 \geq 0 \end{array}$$

The solution is $x^* = (0, 1, 2, -1)^T$ and the corresponding Lagrange multipliers are

$$\Lambda = \left\{ (3 - t, 0, t, t - 2)^T | t \in [2, 3] \right\}.$$

The constraints c_1, c_3 and c_4 are active at x^* and

$$\nabla c_1(x^*) - \nabla c_3(x^*) - \nabla c_4(x^*) = 0.$$

Hence the linear independence constraint qualification (LICQ) is violated. Furthermore strict complementarity is violated when $t = 2$ or $t = 3$.

For each $r \in \{0.01, 0.05, 0.1, 0.5, 1, 2, 10\}$, we first took a random number $t \in [2, 3]$ and got the corresponding "dual solution" $\lambda^* = (3 - t, 0, t, t - 2)$ since the Lagrange multipliers is not unique. Then we generated a random vector (x, λ) according to formula (3.1). We generated 300 random vectors (x, λ) in this way to illustrate and compare the effectiveness of

Figure 2: Numerical results for Example 2

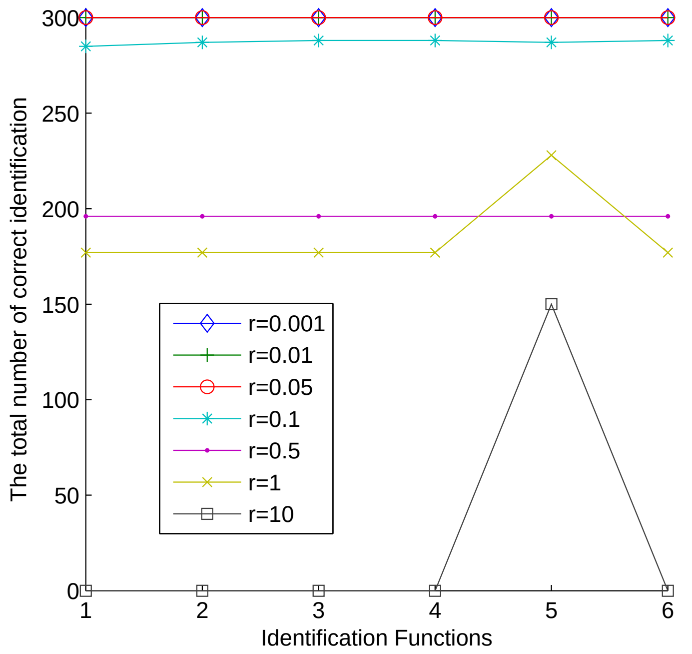


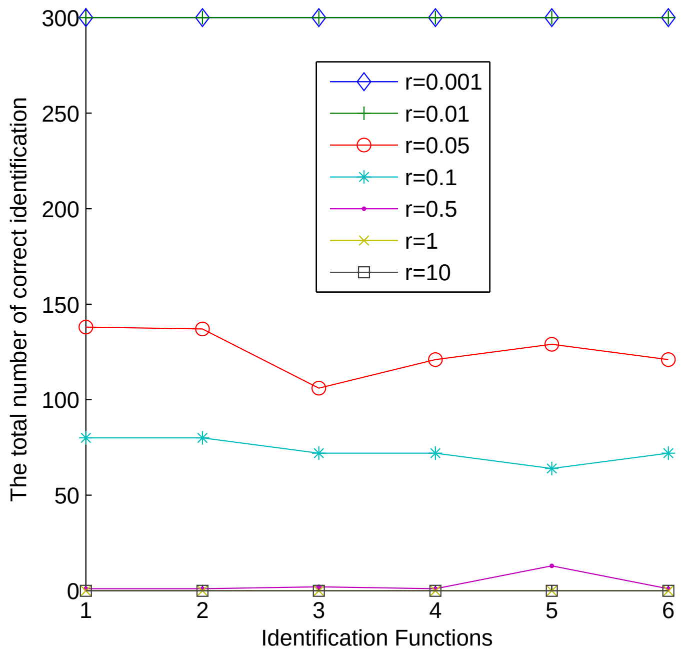
Table 3: Numerical results for Example 3

ρ	$r = 0.001$	$r = 0.01$	$r = 0.05$	$r = 0.1$	$r = 0.5$	$r = 1$	$r = 10$
ρ_1	300	300	138	80	1	0	0
ρ_2	300	300	137	80	1	0	0
ρ_3	300	300	106	72	2	0	0
ρ_4	300	300	121	72	1	0	0
ρ_5	300	300	129	64	13	0	0
ρ_6	300	300	121	72	1	0	0

identification functions ρ_1–ρ_6 which are defined in (2.5) and (2.8). The results are presented in Table 3 and Figure 3.

From the above numerical results, we can see that for $r \leq 0.01$, the whole active sets of the three test examples are all correctly identified by the six identification functions ρ_1–ρ_6, which show the effectiveness of identification function method. For $r > 0.01$, the identification function ρ_5 seems to be the most robust one, and ρ_1, ρ_2 should be the next two. However, the difference among the performances of the six identification functions does not seem to be significant. From the above numerical results, we can see that smooth property does not play an important role in the designing of identification functions. In further studies we will show how to use these identification functions effectively with practical algorithms and do more numerical tests to compare their performance.

Figure 3: Numerical results for Example 3



References

[1] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[2] B. Chen, X. Chen and C. Kanzow. A penalized Fischer-Burmeister NCP-function, *Mathematical Programming* 88 (2000) 211–216.

[3] A.R. Conn, N.I.M. Gould, and Ph. L. Toint, *Trust-Region Methods*, SIAM Publications, Philadelphia, 2000.

[4] J.V. Burke and J.J. Moré, On the identification of active constraints, *SIAM J. Numer. Anal.* 25 (1988) 1197–1211.

[5] J.V. Burke and J.J. Moré, Exposing constraints, *SIAM J. Optim.* 4 (1994) 573–595.

[6] A.S. El-Bakry, R.A. Tapia and Y. Zhang, A study of indicators for identifying zero variables in interior-point methods, *SIAM Rev.* 36 (1994) 45–72.

[7] F. Facchinei, A. Fischer and C. Kanzow. On the accurate identification of active constraints, *SIAM J. Optim.* 9 (1998) 14–32.

[8] R. Fletcher, *Practical Methods of Optimization*, John Wiley, New York, 1987.

[9] J. Gauvin, A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming, *Math. Programming* 12 (1977) 136–138.

[10] W. Hock and K. Schittkowski, *Test Examples for Nonlinear Programming Codes*, Lecture Notes in Economics and Math. Systems 187, Springer-Verlag, Berlin, 1981.

[11] J. Han, N. Xiu and H. Qi, *Theory and Algorithms for Nonlinear compelmentary Problems (in Chinese)*, Shanghai Science and Technology Press, 2006.

[12] A.S. Lewis and S.J. Wright, Identifying activity, *SIAM J. Optim.* 21 (2011) 597–614.

[13] S. Mehrotra and Y. Ye, On finding the optimal facet of linear programs, *Mathematical Programming* 62 (1993) 497–515.

[14] C. Oberlin and S.J. Wright, Active set identification in nonlinear programming, *SIAM J. Optim.* 17 (2006) 577–605.

[15] D. Sun and L. Qi, On NCP-Functions, *Computational pptimization and applications* 13 (1999) 201–220.

[16] W. Sun and Y. Yuan, *Optimization Theory and Methods: Nonlinear Programming*, Springer Science+Business Media, New York, 2006.

[17] S.J. Wright, Modifying SQP for degenerate problems, *SIAM J. Optim.* 13 (2002), 470–497.

[18] Y. Ye, *On the finite convergence of interior-point algorithms for linear programming*, *Mathematical Programming* 57 (1992) 325–335.

Manuscript received 21 March 2013
revised 2 September 2013
accepted for publication 13 October 2013

Zhou-Hong Wang
School of Science, Beijing Jiaotong University, Beijing 100044, P. R. China
E-mail address: `wangzhh@bjtu.edu.cn`

Xiaodong Ding
Department of Applied Mathematics, Zhejiang University of Technology, Hangzhou 310023, PR China
E-mail address: `dxd@lsec.cc.ac.cn`

Yang Yang
School of Science, Beijing Jiaotong University, Beijing 100044, P. R. China
E-mail address: `11121763@bjtu.edu.cn`