J Glob Optim (2008) 42:1–21
DOI 10.1007/s10898-007-9191-8

# Reference variable methods of solving min–max optimization problems

**Baiquan Lu · Yuan Cao · Min jie Yuan · Jianzhen Zhou**

Received: 28 January 2007 / Accepted: 6 June 2007 / Published online: 12 July 2007


**Abstract** In this paper, reference variable methods are proposed for solving nonlinear Minmax optimization problems with unconstraint or constraints for the first time, it uses reference decision vectors to improve the methods in Vincent and Goh (J Optim Theory Appl 75:501–519, 1992) such that its algorithm is convergent. In addition, a new method based on KKT conditions of min or max constrained optimization problems is also given for solving the constrained minmax optimization problems, it makes the constrained minmax optimization problems a problem of solving nonlinear equations by a complementarily function. For getting all minmax optimization solutions, the cost function $f(x, y)$ can be constrained as $M_1 < f(x, y) < M_2$ by using different real numbers $M_1$ and $M_2$. To show effectiveness of the proposed methods, some examples are taken to compare with results in the literature, and it is easy to find that the proposed methods can get all minmax optimization solutions of minmax problems with constraints by using different $M_1$ and $M_2$, this implies that the proposed methods has superiority over the methods in the literature (that is based on different initial values to get other minmax optimization solutions).

**Keywords** Minmax optimization · Topologic construction · Convergence problem · Reference vectors · KKT conditions and complementarily function · Hybrid method

## 1 Introduction

Considering the following minmax problems with unconstraint or constraints,

$$\min_x \max_y f(x, y) \tag{1}$$

B. Lu (✉)
Shanghai key laboratory of power station, School of Mechatronics and Automation,
Shangai University, Shanghai, China
e-mail: ibq123188@yahoo.com.cn

Y. Cao · M. j. Yuan · J. Zhou
Department of Automation,
Shanghai University,
Shanghai 200083, China

where $f(x, y)$ is continuously differentiable at point $(x, y)$, $x \in R^s$ and $y \in R^r$ are s and r dimensional decision vectors, respectively.

$$\begin{cases} \min_x \max_y f(x, y) \\ h_i(x, y) = 0 \qquad i = 1, 2, \ldots, m1, \quad j = 1, 2, \ldots, m2 \\ g_j(x, y) \geq 0 \end{cases} \tag{2}$$

where $x = (x_1, x_2, \ldots, x_{n1})^T \in R^{n1}$, $y = (y_1, y_2, \ldots, y_{n2})^T \in R^{n2}$, $f(x, y) \in R$, $h_i(x, y) \in R$ and $g_j(x, y) \in R$ are continuously differentiable.

Because $h_i(x, y) = 0 \Leftrightarrow h_i(x, y) \geq 0, -h_i(x, y) \geq 0$, problem (2) can be rewritten as

$$\begin{cases} \min_x \max_y f(x, y) \\ g_j(x, y) \geq 0 \end{cases} \qquad j = 1, 2, \ldots, m \tag{3}$$

In this paper, we consider minmax optimization solutions of problems (1) and (3), which are defined by the following definitions,

**Definition 1** If there exist $\delta$ and a point $(x^*, y^*)$ such that $f(x^*, y) \leq f(x^*, y^*) \leq f(x, y^*)$ holds for all points $(x, y)$ satisfying $||x - x^*|| \leq \delta$, $||y - y^*|| \leq \delta$, then point $(x^*, y^*)$ is a local optimization solution of problem (1).

**Definition 2** If there exist $\delta$ and a point $(x^*, y^*)$ such that $f(x^*, y) \leq f(x^*, y^*) \leq f(x, y^*)$ holds for all points $(x, y)$ satisfying $||x - x^*|| \leq \delta$, $||y - y^*|| \leq \delta$ in set $S = \{|g_j(x, y) \geq 0, j = 1, \ldots, m\}$, then point $(x^*, y^*)$ is a local optimization solution of problem (3).

The minmax optimization problem with constraints or unconstraint has yet not been satisfactorily solved in the literature. The first issue is whether the optimization solution of Eq. 1 exists, particularly for problem (3). Many promising results for minmax optimization problems with unconstraint were reported [1–7], which are difficult for calculation, such as methods based on topological theory in Refs. [2,6]; But for minmax optimization problems with constraints, reports about this issue are few if any. The second issue is how to obtain minmax optimization solutions of problem (1) or (3), there exist several different methods for solving problem (1) or (3), for example, when the cost function $f(x, y)$ is discontinuous for arbitrary $y$, a variety of regularization approaches has been used to obtain smooth approximations to problem (1), but sometimes the smooth approximating problems become significantly ill-conditioned, to this end, feedback precision-adjustment rule is given in Ref. [8–12]; when the cost function $f(x, y)$ is continuously differentiable for $x$ and $y$, there exists some methods such as T.L.Vincent's method [13] to solve problem (1), though T.L.Vincent's method given the convergence condition on minmax optimization solutions of problem (1) or problem (3) with bounds on $x$ and $y$, for some examples, such as Example 3 in Ref. [13], this condition is not satisfied, in other words, it is impossible that minmax optimization solutions of the kind of problems are got by this method, the reason of generating this case is that the topologic construction of this method is not convergent near its minmax optimization solutions. To overcome this, this paper gives new methods to improve its topologic construction, which adds reference variables into T.L.Vincent's method in order to ensure that its topologic construction is convergent near its minmax optimization solutions. In addition, in order to solve problem (3) with general constraints, a new method based on solving nonlinear equations is also given in this paper, which applies a nonlinear complementarily function [14] to necessary conditions of problem (3) like KKT conditions and gets nonlinear equations, thus the minimax optimization problems with constraints becomes a problem of solving nonlinear equations by current methods [15], such as inexact Newton method and so on, its key technique is Newton iteration method, usual Newton iteration method is locally convergent, if homotopy method or Trust region method is added into this method, then a hybrid method with global convergence is got. Because the topologic construction of every method is different from each other, it leads to that their convergence trajectories are different from each other, thus considering convergence speed and global convergence of each method, hybrid methods of current methods are proposed. Because sometimes there exist many minmax optimization solutions in problem (3), for getting all solutions, in this paper, the cost function $f(x, y)$ is constrained as $M_1 < f(x, y) < M_2$ where $M_1$ and $M_2$ are real numbers, thus all minmax optimization solutions of problem (3) can be got by using different $M_1$ and $M_2$.

Based on the above, aim of this paper is to get both new convergence algorithms based on reference variables for solving problem (1) or problem (3) with bounded constraints on variables $x$, $y$ and a new method based on solving nonlinear equations for solving problem (3). At the same time, to show efficiency of the proposed method, six examples are used to compare with results in the literature, as a result, we find out that (1) the proposed method is convergent near its minmax optimization solution for Examples 1 and 2, but T.L.Vincent's method is not convergent; (2) the proposed method is also convergent near its minmax optimization solution for Examples 3 and 4 as with T.L.Vincent's method; (3) all minmax optimization solutions of problem (3) are got by the proposed method for Examples 5 and 6, but this is difficult to be get by methods in the literature, this indicates that the proposed method is very useful.

## 2 New algorithms for solving minmax optimization problems with unconstraint or bounded constraints on variables

In this section, first of all, we introduce trajectory-following methods given in Ref. [13] since it is very robust for solving minmax optimization problems with unconstraint or bounded constraints on variables, but for some examples, such as Example 3 in Ref. [13] or problem (3), this is difficult to be get by methods in the literature, to this end, based on trajectory-following method, we give new methods with reference decision vectors for solving minmax optimization problems with unconstraint or bounded constraints on variables.

T.L.Vincent's method for solving problem (1) with unconstraint can be written as

$$\Delta x = -\eta \frac{\partial f(x, y)}{\partial x}, \quad \Delta y = \eta \frac{\partial f(x, y)}{\partial y} \tag{4}$$

where $\eta$ is a small positive number.

T.L.Vincent's method for solving $\min_x \max_y f(x, y)$ with constraints $\Omega = \{(x, y)|x^L \leq x \leq x^H, y^L \leq y \leq y^H\}$ is given by

$$\Delta x = \begin{cases} 0 & \text{if } x = x^L, \ \frac{\partial f(x,y)}{\partial x} \leq 0 \\ 0 & \text{if } x = x^H, \ \frac{\partial f(x,y)}{\partial x} \geq 0 \\ -\eta \frac{\partial f(x,y)}{\partial x} & \text{otherwise} \end{cases}$$

$$\Delta y = \begin{cases} 0 & \text{if } y = y^L, \ \frac{\partial f(x,y)}{\partial y} \leq 0 \\ 0 & \text{if } y = y^H, \ \frac{\partial f(x,y)}{\partial y} \geq 0 \\ -\eta \frac{\partial f(x,y)}{\partial y} & \text{otherwise} \end{cases} \tag{5}$$

where $\eta$ is a small positive number.

To ensure convergence of Eqs. 4 and 5, T.L.Vincent's method gives the following sufficient condition.

The sufficient condition of minmax optimization solution: For a given continuously differentiable function $f(x, y)$, trajectories of Eqs. 4 or 5 converge to the equilibrium point $(x^*, y^*)$ from any initial point $(x, y) \in \Omega$ if $-(x-x^*)^T\ (\partial f(x, y)/\partial x)+(y-y^*)^T\ (\partial f(x, y)/\partial y) < 0$ holds, where point $(x^*, y^*)$ is a minmax optimization solution of $f(x, y)$.

Because point $(x^*, y^*)$ satisfying the above sufficient condition is unknown, it is difficult to ensure that trajectories of Eqs. 4 or 5 converge to minmax optimization solution $(x^*, y^*)$ when initial values of variables $x$ and $y$ are far from point $(x^*, y^*)$. In fact, for some examples, the trajectories of Eqs. 4 or 5 are also not convergent near its minmax optimization solution even if $\eta$ is very small, such as example $f(x, y) = xy$ in Ref. [13]. From the following theorem 1, it is easy to see that the reason of generating this case is that the topologic construction of Eqs. 4 or 5 is not convergent near its minmax optimization solution for some examples.

**Theorem 1** *For a given continuously differentiable function $f(x, y)$, assuming that point $(x^*, y^*)$ is a minmax optimization solution of $f(x, y)$, if $|Z| < 1$ for all solutions $Z$ of $|A - ZI| = 0$ near point $(x^*, y^*)$, then trajectories of Eq. 4 are convergent; if $|Z| > 1$ for all solutions $Z$ of $|A - ZI| = 0$ near point $(x^*, y^*)$, then trajectories of Eq. 4 are not convergent; (for proof, see Appendix), where $\eta$ is a small positive number (see Eq. 4) and*

$$A = \begin{bmatrix} I - \eta\frac{\partial^2 f(x^*,y^*)}{\partial^2 x} & \eta\frac{\partial^2 f(x^*,y^*)}{\partial xy} \\ \eta\frac{\partial f(x^*,y^*)}{\partial xy} & -I - \eta\frac{\partial^2 f(x^*,y^*)}{\partial^2 y} \end{bmatrix}.$$

Now we apply Theorem 1 to example $f(x, y) = (x^2 - 1)(y^2 - 1)$. By Definition 1, points (1, 1), (1, −1), (−1, 1) and (−1, −1) are minmax optimization solutions of $f(x, y)$, near any minmax optimization solution, such as point (1, 1), we have $A = \begin{bmatrix} 1 & 4\eta \\ 4\eta & -1 \end{bmatrix}$. From $|A - ZI| = 0$, it is easy to see that $|Z| > 1$ for arbitrary positive number $\eta$, in other words, the topologic construction of Eq. 4 near point (1, 1) is not convergent. This implies that it is difficult that minmax optimization solution (1, 1) is got by using Eq. 4 even if initial values are set to number close to point (1, 1), similar to Eq. 4, there exists also this case for Eq. 5. To overcome this case, we propose two methods for changing the topologic construction of Eqs. 4 or 5: simplest method is to add moment items into Eqs. 4 or 5, the other method is to add reference variables to Eqs. 4 or 5.

### 2.1 Adding moment item methods for solving minmax optimization problems with unconstraint or bounded constraints on variables

To improve the topologic construction of Eq. 4, an adding moment item method is given by

$$\Delta x = -\eta_1\frac{\partial f(x, y)}{\partial x} + \eta_2(x(k-1) - x(k-2)),$$
$$\Delta y = \eta_1\frac{\partial f(x, y)}{\partial y} - \eta_2(y(k-1) - y(k-2)) \qquad (6)$$

where $\eta_1$ and $\eta_2$ are small positive numbers.

T improve the topologic construction of Eq. 5, another adding moment item method is given by

$$\Delta x = \begin{cases} 0 & \text{if } x = x^L,\ \frac{\partial f(x,y)}{\partial x} \le 0 \\ 0 & \text{if } x = x^H,\ \frac{\partial f(x,y)}{\partial x} \ge 0 \\ -\eta_1 \frac{\partial f(x,y)}{\partial x} + \eta_2(x(k-1) - x(k-2)) & \text{otherwise} \end{cases}$$
$$\Delta y = \begin{cases} 0 & \text{if } y = y^L,\ \frac{\partial f(x,y)}{\partial y} \le 0 \\ 0 & \text{if } y = y^H,\ \frac{\partial f(x,y)}{\partial y} \ge 0 \\ -\eta_1 \frac{\partial f(x,y)}{\partial y} + \eta_2(y(k-1) - y(k-2)) & \text{otherwise} \end{cases} \tag{7}$$

where $\eta_1$ and $\eta_2$ are small positive numbers.

Because $x^L \le x \le x^H, y^L \le y \le y^H$, there exist vectors $U = \{u_1, \ldots, u_{n1}\}$ and $V = \{V_1, \ldots, V_{n2}\}$ such that $x_i = x_i^L + 2 * (x_i^H - x_i^L)/(1 + e^{-u_i})$ and $y_i = y_i^L + 2 * (y_i^H - y_i^L)/(1 + e^{-v_i})$, substituting these into Eq. 7 yields

$$\Delta U = -\eta_1 \frac{\partial f(U, V)}{\partial U} + \eta_2(U(k-1) - U(k-2)),$$
$$\Delta V = -\eta_1 \frac{\partial f(U, V)}{\partial V} + \eta_2(V(k-1) - V(k-2))$$

To ensure that the moment items improve the topologic construction of Eqs. 4 and 5, the following Theorem 2 is given.

**Theorem 2** *For a given continuously differentiable function $f(x, y)$, assuming that point $(x^*, y^*)$ is a minmax optimization solution of $f(x, y)$, if $|Z| < 1$ for all solutions $Z$ of $|A(Z)| = 0$ near point $(x^*, y^*)$, then Eq. 6 is convergent; if $|Z| > 1$ for all solutions $Z$ of $|A(Z)| = 0$ near point $(x^*, y^*)$, then Eq. 6 is not convergent; (for proof, see Appendix), where*

$$A = \begin{bmatrix} -Z^2 I + (1 + \eta_2 - a)Z + \eta_2 I & bZ \\ bZ & -Z^2 I + (1 + \eta_2 - c)Z + \eta_2 I \end{bmatrix},$$
$$a = \eta_1 \frac{\partial^2 f(x^*, y^*)}{\partial^2 x}, b = \eta_1 \frac{\partial^2 f(x^*, y^*)}{\partial xy} \text{ and } c = \eta_1 \frac{\partial^2 f(x^*, y^*)}{\partial^2 y},$$

*and $\eta_1$ and $\eta_2$ are small positive numbers.*

Now we apply this theorem to example $f(x, y) = (x^2 - 1)(y^2 - 1)$, near any minmax optimization solution, such as point (1, 1), we have $A = \begin{bmatrix} -Z^2 + (1 + \eta_2)Z + \eta_2 & 4Z\eta_1 \\ 4Z\eta_1 & -Z^2 + (1 + \eta_2)Z + \eta_2 \end{bmatrix}$. In theory, it follows that $|Z| < 1$ for selecting $\eta_1$ and $\eta_2$ (see Appendix), in other words, Eq. 6 with momentum items is convergent by selecting $\eta_1$ and $\eta_2$, though this implies that momentum items can change the topologic construction of Eq. 4, it is difficult to select $\eta_1, \eta_2$ such that $|Z| < 1$. To overcome this case, we propose a new method based on reference decision vectors for improving the topologic construction of Eqs. 4 and 5, which is easy to select learning rates $\eta_1,\ \eta_2,\ \eta_3$ such that $|Z| < 1$.

## 2.2 Reference variable methods of solving min–max optimization problems with unconstraint or bounded constraints on vectors $x$ and $y$

To improve the topologic construction of Eq. 4, a reference variable method is given by

$$\begin{aligned} x(k) &= x(k-1) - \eta_1 \frac{\partial f(x,y)}{\partial x} - \eta_2(x(k) - x_r(k-1)) \\ y(k) &= y(k-1) + \eta_1 \frac{\partial f(x,y)}{\partial y} - \eta_2(y(k-1) - y_r(k-1)) \\ x_r(k) &= x_r(k-1) + \eta_3(x(k-1) - x_r(k-1)) \\ y_r(k) &= y_r(k-1) + \eta_3(y(k-1) - y_r(k-1)) \end{aligned} \tag{8}$$

where initial values of reference vectors $x_r, y_r$ are different from that of vectors $x, y, \eta_1, \eta_2$ and $\eta_3$ are small positive numbers.

T improve the topologic construction of Eq. 5, another reference variable method is given by

$$\Delta x = \begin{cases} 0 & \text{if } x = x^L,\ \frac{\partial f(x,y)}{\partial x} \geq 0 \\ 0 & \text{if } x = x^H,\ \frac{\partial f(x,y)}{\partial x} \leq 0 \\ -\eta \frac{\partial f(x,y)}{\partial x} - \eta_2(x - x_r) & \text{otherwise} \end{cases}$$

$$\Delta y = \begin{cases} 0 & \text{if } y = y^L,\ \frac{\partial f(x,y)}{\partial y} \leq 0 \\ 0 & \text{if } y = y^H,\ \frac{\partial f(x,y)}{\partial y} \geq 0 \\ \eta_1 \frac{\partial f(x,y)}{\partial y} - \eta_2(y - y_r) & \text{otherwise} \end{cases}$$

$$\Delta x_r(k) = \eta_3(x(k-1) - x_r(k-1)),\ \Delta y_r(k) = \eta_3(y(k-1) - y_r(k-1)) \tag{9}$$

where initial values of reference vectors $x_r, y_r$ are different from that of vectors $x, y$, with $x^L \leq x \leq x^H, x^L \leq x_r \leq x^H, y^L \leq y \leq y^H, y^L \leq y_r \leq y^H, \eta_1, \eta_2$ and $\eta_3$ are small positive numbers.

To ensure that the moment items improve the topologic construction of Eqs. 4 and 5, the following Theorems 3 and 4 are given.

**Theorem 3** *Considering* $\min_x \max_y f(x, y)$ *with unconstraint, for Eq. 8, if* $\left\| \frac{\partial^2 f(x,y)}{\partial^2 x} \right\|$ *and* $\left\| \frac{\partial^2 f(x,y)}{\partial^2 y} \right\|$ *are bounded, then vectors x and y are convergent by selecting* $\eta_1, \eta_2$ *and* $\eta_3$ *near any minmax optimization solution of* $\min_x \max_y f(x, y)$ (*for proof see Appendix*)

From this theorem, it is easy to see that reference variables can change the topologic construction of Eq. 4 and make Eq. 8 converge by suitable choice of $\eta_1, \eta_2$ and $\eta_3$. Though this theorem is a sufficient condition, from the proof of this theorem, we see that if only $\eta_2$ and $\eta_3$ are chosen as numbers close to 1 and $\eta_1$ is very small positive number (which is easy to carry out), then there always exists an equilibrium point (that may be unknown point) such that Eq. 8 converge to it even if initial values of variables $x$ and $y$ are far from this point, this indicates that the proposed method is feasible and useful. Similar to Theorem 3, for minmax optimization problems with bounded constraints on vectors $x$ and $y$, we have the following Theorem 4.

**Theorem 4** *Considering* $\min_x \max_y f(x, y)$ *with bounded constraints on vectors x and y, for Eq. 9, if* $\left\| \frac{\partial^2 f(x,y)}{\partial^2 x} \right\|$ *and* $\left\| \frac{\partial^2 f(x,y)}{\partial^2 y} \right\|$ *are bounded, then vectors x and y are convergent by selecting* $\eta_1, \eta_2$ *and* $\eta_3$ *near any minmax optimization solution of* $\min_x \max_y f(x, y)$

*with bounded constraints on vectors x and y (the proof of this theorem is somewhat similar to Theorem 3, omit here). Because* $x^L \le x \le x^H$, $y^L \le y \le y^H$, *there exist vectors* $U = \{u_1, \ldots, u_{n1}\}$ *and* $V = \{V_1, \ldots, V_{n2}\}$ *such that* $x_i = x_i^L + 2 * (x_i^H - x_i^L)/(1 + e^{-u_i})$ *and* $y_i = y_i^L + 2 * (y_i^H - y_i^L)/(1 + e^{-v_i})$, *substituting these into Eq. 9 yields*

$$\Delta U = -\eta \frac{\partial f(U, V)}{\partial U} - \eta_2 (U(k-1) - U_r(k-1)), \quad \Delta V = -\eta \frac{\partial f(U, V)}{\partial V} + \eta_2 (V(k-1) - V_r(k-1))$$

$$\text{and } \Delta U_r(k) = \eta_3 (U(k-1) - U_r(k-1)), \ \Delta V_r(k) = \eta_3 (V(k-1) - V_r(k-1)).$$

From this theorem, we see that reference variables can change the topologic construction of Eq. 5, and make Eq. 9 converge, in other words, reference variable method can solve minmax optimization problems with bounds on vectors $x$ and $y$ or unconstrained minmax optimization problems, but it is difficult to be applied to problem (3). To this end, we try to find a new method to solve all solutions of problem (3) in the next section, which makes the constrained minmax optimization problems a problem of solving nonlinear equations by a complementarily function, for getting all minmax optimization solutions, the cost function $f(x, y)$ is constrained as $M_1 < f(x, y) < M_2$ where $M_1$ and $M_2$ are real numbers.

## 3 New methods for solving minmax optimization problems with constraints

It is well known that KKT conditions are necessary condition of min or max optimization problems with constraints, though the KKT conditions are only necessary condition, it given a method for solving min or max optimization problems with constraints using nonlinear equations. Applying this idea to problem (3), that how to use this necessary condition of problem (3) to get its all solutions will be discussed in this section.

### 3.1 The necessary condition of Minmax optimization problems with constraints

By reference to methods of getting KKT conditions of min or max optimization problems with constraints, KKT necessary condition of Eq. 2 is given by the following Theorems 5 and 6.

**Theorem 5** *Assume that point* $(x^*, y^*)$ *is a minmax optimization solution of problem (2),* $f(x, y)$ *and* $g_j$ *are continuously differentiable at point* $(x^*, y^*)$ *as* $j \in J$ *where* $J = \{j | g_j(x, y) = 0, j = 1, \ldots, m\}$, *then there exists no vectors* $P1$ *and* $P2$ *such that* $\begin{cases} \nabla_x f(x^*, y^*) P1 < 0 \\ \nabla_x g_j(x^*, y^*) P1 > 0 \end{cases}$ *and* $\begin{cases} \nabla_y f(x^*, y^*) P2 > 0 \\ \nabla_y g_j(x^*, y^*) P2 > 0 \end{cases}$ *hold (for the proof, see appendix).*

**Theorem 6** *Assume that point* $(x^*, y^*)$ *is a minmax optimization solution of problem (2), and the set* $\{\nabla g_i | i \in I\}$ *with* $(I = \{i | g_i = 0\})$ *are linearly independent at point* $(x^*, y^*)$, *then there exists point* $(x^*, y^*)$ *such that*

$$\begin{aligned}
&(1)\ \gamma_i g_i(x^*, y^*) = 0,\ \lambda_i g_i(x^*, y^*) = 0,\ \lambda_i, \gamma_i \ge 0,\ g_i(x^*, y^*) \ge 0, \quad i = 1, 2, \ldots, m, \\
&(2)\ \nabla_x f(x^*, y^*) - \sum_i^m \lambda_i \nabla_x g_i(x^*, y^*) = 0, \\
&(3)\ \nabla_y f(x^*, y^*) + \sum_i^n \gamma_i \nabla_y g_i(x^*, y^*) = 0, \quad j = 1, 2, \ldots, m. \text{ (The proof, see appendix)}
\end{aligned} \tag{10}$$

It should be noted that there exists the case of $\lambda_i = 0$, $g_i(x^*, y^*) < 0$ for $\lambda_i g_i(Xx^*, y^*) = 0$, which is not solution of Eq. 10, To make $\lambda_i g_i(\bar{X}) = 0$, $\lambda_i \geq 0$, $g_i(\bar{X}) \geq 0$ hold, a nonlinear complementarily function [14] NCF $\Phi$: $R^2 \rightarrow R$ can be used to limit feasible region of solutions of Eq. 10 where $\Phi(U, V) = 0 \Leftrightarrow U \geq 0, V \geq 0, UV = 0$, generally $\Phi(U, V)$ is selected as $\sqrt{U^2 + V^2} - U - V$. Because $\Phi(U, V)$ is not differentiable at $U = V = 0$, $\Phi^2(U, V)$ can be used to limit feasible region of Eq. 10 since $\Phi^2(U, V)$ is differentiable at $U = V = 0$. Thus, Eq. 10 can be written as

$$\begin{aligned} &(1)\ \Phi(\lambda_i, g_i) = 0, \quad i = 1, 2, \ldots, \mathrm{m}, \\ &(2)\ \nabla_x f(x^*, y^*) - \sum_i^n \lambda_i \nabla_x g_i(x^*, y^*) = 0, \\ &(3)\ \nabla_y f(x^*, y^*) + \sum_i^n \gamma_i \nabla_y g_i(x^*, y^*) = 0, \quad j = 1, 2, \ldots, \mathrm{m} \end{aligned} \tag{11}$$

where $\Phi(u, v) = \sqrt{u^2 + v^2} - u - v$.

Because Theorem 6 is a necessary condition of problem 2, it is possible that there exist many solutions in Eq. 11. In order to get all feasible solutions, in this paper, as new constraint, inequality $M_1 + \varepsilon \leq f(x, y) \leq M_2 - \varepsilon$ is used to limit feasible region of solutions in Eq. 11 such that all feasible solutions of Eq. 11 are solved continuously by using different $M_1$ and $M_2$, where $M_1$ and $M_2$ are real numbers, $\varepsilon$ is a given small positive number such as 0.1.

Though if there exists a solution in Eq. 11, then we always can find a method of solving nonlinear equations to get this solution, it should be noted that some methods of solving nonlinear equations are locally convergent, if initial values are far from the solutions of Eq. 11, then their trajectories are not convergent, this does not imply that there exist no solution in this equations. For this, global convergence methods are used to solve Eq. 11 as much as possible, such as Continuation method.

### 3.2 Hybrid methods for solving nonlinear equations

To solve Eq. 11, there are many methods, the key technique of which is Newton iteration method, usual Newton iteration method is locally convergent, if homotopy method or Trust region method is added into this method, then a hybrid method with global convergence is got. Because the topologic construction of every method is different from each other, it leads to that their convergence trajectories are different from each other, thus considering convergence speed and global convergence of each method, hybrid methods of the following various methods are necessary.

(1) Methods from optimization toolbox of MATLAB, for example, Hessian method, Jacobean method, Large-Scale method, LevenbergMarquardt method, LineSearchType method and so on, whose Maximum iteration number and terminate tolerance on the function value can be set. If terminate tolerance on the function value is satisfied, a solution of nonlinear equations is got; If the maximum iteration number is reached, this requires changing initial values or using other methods to solve it again.

(2) Continuation method [16] with convergence in large region (that is got by homotopy method and Newton iteration).
For a given equations $F(X) = 0$, if $X_0$ is initial value, the iteration formula is given by $\frac{dX(t)}{dt} = \left[F'(X) + \alpha(1 - t^3)I\right]^{-1} (3\alpha t^2(X - X_0) - F(X_0))$, $t \in [0, 1]$ where $\alpha$ is an large number enough so that the inverse of the above matrix exists. This equation can be

written by using integration method with mean point as

$$X_1 = X_0 - (1/N)[J(X_0) + \alpha_1 I]^{-1} F(X_0),\ X_{k+1/2} = X_k + (X_k - X_{k-1})/2,$$
$$t_{k+1/2} = (k + 1/2)/N$$
$$X_{k+1} = X_k - (1/N)[J(X_{k+1/2}) + \alpha_2(1 - t^3_{k+1/2})I]^{-1}[F(X_0) - 3\alpha_2 t^2_{k+1/2}(X_{k+1/2} - X_0)]$$

where $J$ is the Jacobian matrix of $F(X)$, $1/N$ is integration steplength.
Because this algorithm is convergent in large region, it can be used to give initial values to a local convergence algorithm, such as LargeScale method from MATLAB optimization toolbox. If only $N$ is an large number enough so that its trajectory finally enters into convergence region of the local convergence algorithm that can be a method from MATLAB optimization toolbox or a hybrid algorithm (that is composed of some methods from MATLAB optimization toolbox, their sequence is generated by a rule, and iteration results of former method are used as initial value of the next method), then its trajectory gradually approaches to a solution of $F(X) = 0$ via each iteration. In addition, because computing time of solving higher-dimensional cost function problem is longer than that of solving low- dimensional cost function, iteration number $N$ is selected as smaller number as possible for higher-dimensional cost function problem.

(3) Reference variable method with convergence in large region (Lu, Submitted)

$$X_k = X_{k-1} + K_1 f(X_k) - K_2(X_{k-1} - Y_{k-1})$$
$$Y_k = Y_{k-1} + K_3(X_{k-1} - Y_{k-1}),\ f(X) = [f_1, f_2, \cdots, f_n]$$
$$K_1 = \mathrm{diag}\left(-K\,\mathrm{sign}\left(\frac{\partial f_1}{\partial x_1}\right), -K\,\mathrm{sign}\left(\frac{\partial f_2}{\partial x_2}\right), \ldots, -K\,\mathrm{sign}\left(\frac{\partial f_n}{\partial x_n}\right)\right)$$
$$X_k = [x_1, x_2, \ldots, x_n]$$

where $X_k$ and $Y_k$ are solution vector of equations $f(X) = 0$ and corresponding reference vector, respectively, $0 < K < 1$ and $0 < K_2, K_3 < 2$. This algorithm can be used to set an initial value for some local convergence algorithms.

(4) Other methods such as inexact Newton method for solving equations $F(X) = 0$ in Ref. [17–21].
Assume that the steplength is 1, for a given $X_0$, starting from iteration number $K = 0$, the following is executed until the iteration is convergent.
Choose $\eta_k \in [0, 1)$ and $S_k$ properly such that $\left\|F'(X_k)S_k + F(X_k)\right\| \le \eta_k \|F(X_k)\|$, $X_{k+1} = X_k + S_k$ hold, where $\eta_k$ often depend on $S_k$.
This algorithm has been proved to be local convergence. If combining inexact Newton method with inexact linear search method or Trust region algorithm, it is globally convergent.

Considering Eq. 11 and hybrid method of the above methods, a new algorithm is got for solving minmax optimization problem with constraints.

### 3.3 A new algorithm for solving minmax optimization problems with constraints

In Subsect. 3.1, minmax optimization problems are changed to a problem of solving nonlinear Eq. 11 continuously, while in Subsect. 3.2, methods for solving nonlinear Eq. 11 has been discussed, combining Subsect. 3.1 with Subsect. 3.2, the following new algorithm is got for solving minimax optimization problems with constraints.

**Algorithm A**

*Step 1:* For given two numbers $M_0$ and $M_1(M_0 < M_1)$ such that $f(X_0) = M_0$ and $f(X_1) = M_1$, where $X_0$ and $X_1$ are not minimax optimization solutions of problem 2, then construct the set of constraints S $= \{f(x) \le f(X_0) - \varepsilon$ or $f(X_0) + \varepsilon \le f(x) \le f(X_1) - \varepsilon$ or $f(X_1) + \varepsilon \le f(x)\}$, where $\varepsilon$ is a given small positive number such as 0.1.

*Step 2:* For ith constraint $f(X_i) < f(x) < f(X_{i+1})$ in set S, first, add it to Eq. 11 by using nonlinear complementarily function, then solve this equations by using a hybrid method given in Subsect. 3.2, if the solution exists (that means error of each equation is smaller than $10^{-14}$), then the solution $X_i^{\text{new}}$ is got for $i$th constraint in set S, otherwise in order to test and verify whether the equations have other solutions, several different initial values are chosen to solve Eq. 11 again, if its solution still dose not exist, then $\varepsilon = \varepsilon/10$, solve the equations again until the resulting Eq. 11 for $i$th constraint is not solvable any more, this implies that there exists no solution for problem 2 as $f(X_i) < f(x) < f(X_{i+1})$, thus this constraint $f(X_i) < f(x) < f(X_{i+1})$ will be removed from set S, when all constraints are removed, the algorithm is over, otherwise go to step 3.

*Step 3:* Adding constraints $f(X_i) < f(x) < f(X_i^{\text{new}})$ and $f(X_i^{\text{new}}) < f(x) < f(X_{i+1})$ for all $i$ $(i = 1, 2, \ldots,)$ into constraint set S, rearrange order of $X$ according to value of $f(X)$, then go to Step 2.

Because it is difficult to determine whether a nonlinear equations has solution by current methods, it is necessary that several different initial values are chosen to test and verify whether the equations has other solution.

The algorithm is not only suitable to solve minmax optimization problems with separated minmax optimization solutions whose number is finite, but is also suitable to solve minmax optimization problems with continuous optimization solutions since it can get the distribution of the optimization solutions for this problems.

As stated above, the proposed method makes the constraint minmax optimization problem a problem of solving nonlinear equations, and can get all minmax optimization solutions, this is very important to practical problems. From this, we can claim that the proposed method is superior to existing methods for solving nonlinear minmax optimization problems in the literature.

## 4 Examples

In this section, some simulations are carried out in order to verify the above ideas in three respects: The first is that for Examples 1 and 2, their minmax optimization solutions are got by the proposed reference variable method, but it can't be got by T.L.Vincent's methods. The second is that for Examples 3 and 4, their minmax optimization solutions are got by the proposed reference variable method as with T.L.Vincent's methods. The third is to verify the proposed algorithm may find all minmax optimization solutions of problem (2) with constraints, such as Examples 5 and 6, which can't be solved by T.L.Vincent's methods. A set of Matlab programs is written on 1.5G PC Windows XP platform to implement the tests. For Examples 5 and 6, eight different initial values are chosen to test and verify whether Eq. 11 have solution that makes the absolute value of the error of each equation smaller than $10^{-14}$.

*Example 1* $\min\limits_{x} \max\limits_{y} f(x, y) = (x^2 - 1)(y^2 - 1)$ (Ref. [13]),

From $(\partial f(x, y)/\partial y) = 0, (\partial f(x, y)/\partial x) = 0$, we get solutions $(0, 0), (-1, -1), (1, 1), (-1, 1)$ and $(1, -1)$. By Definition 1, we see that Minmax optimization solutions of $f(x, y)$

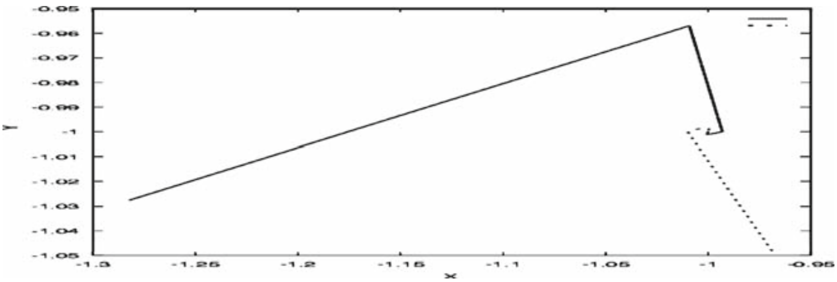


**Fig. 1** Trajectories starting from initial value [(−2.1,−3.1),(−2.1,−1.1)] for variables $x$, $y$ and reference variables $x_r$, $y_r$ for Example 1 (*Solid line* and *dot line* means trajectory of variables $x$, $y$ and trajectory of reference variables $x_r$, $y_r$, respectively)

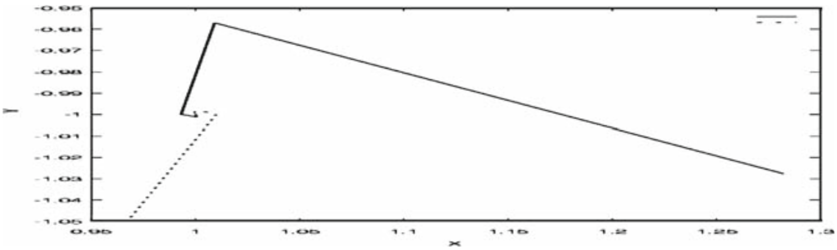


**Fig. 2** Trajectories starting from initial value [(2.1, −3.1),(2.1, −1.1)] for variables $x$, $y$ and reference variables $x_r$, $y_r$ for Example 1 (*Solid line* and *dot line* means trajectory of variables $x$, $y$ and trajectory of reference variables $x_r$, $y_r$, respectively)

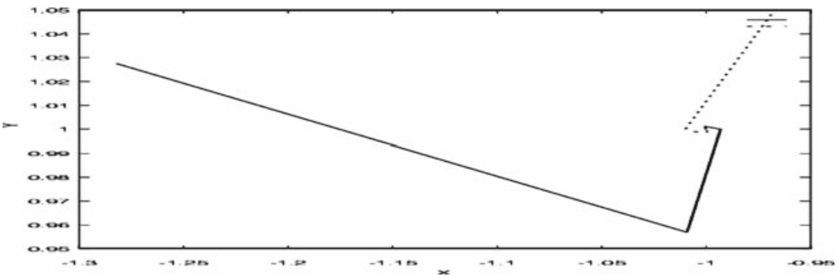


**Fig. 3** Trajectories starting from initial value [(−2.1, 3.1),(−2.1, 1.1)] for variables $x$, $y$ and reference variables $x_r$, $y_r$ for Example 1 (*solid line* and *dot line* means trajectory of variables $x$, $y$ and trajectory of reference variables $x_r$, $y_r$, respectively)

are points (−1, −1), (1, 1), (−1, 1) and (1, −1). In the simulations, near each minmax optimization solution, the initial values of variables $x$, $y$ and reference variables $x_r$, $y_r$ are taken as [(−2.1, −3.1),(−2.0, −1.1)], [(2.1, −3.1),(2.0, −1.1)], [(−2.1, 3.1),(−2.0, 1.1)], [(2.1, 3.1), (2.0, 1.1)], respectively, $\eta_1$, $\eta_2$ and $\eta_3$ are 0.001, 0.002, 0.001, respectively, trajectories starting from different initial values converge to points (−1, −1), (1, −1), (−1, 1) and (1, 1), respectively as shown in Figs. 1–4. As a result, the proposed reference variable method is convergent near minmax optimization solutions for this example, but T.L.Vincent's method is not convergent even if initial values of variables $x$ and $y$ are numbers close to Minmax optimization solutions, this indicates that the proposed method is very useful.

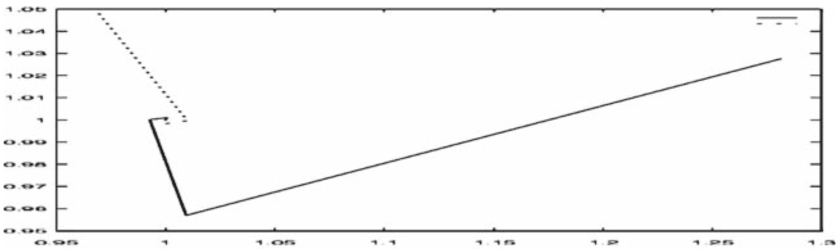


**Fig. 4** Trajectories starting from initial value [(2.1, 3.1),(2.1, 1.1)] for variables $x$, $y$ and reference variables $x_r$, $y_r$ for Example 1 (*Solid line* and *dot line* means trajectory of variables $x$, $y$ and trajectory of reference variables $x_r$, $y_r$, respectively)

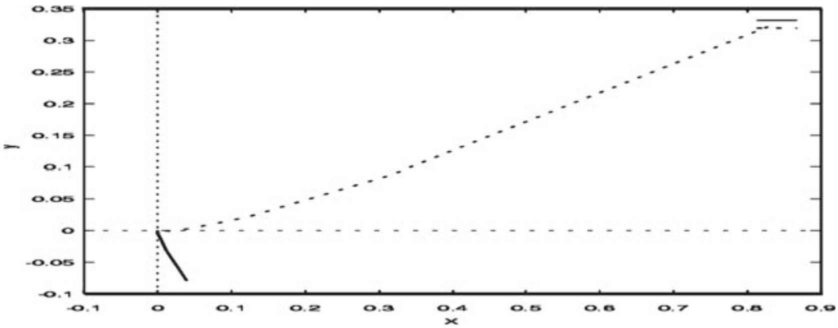


**Fig. 5** Trajectories starting from initial value [(2.1, 3.1),(2.1, 1.1)] for variables $x$, $y$ and reference variables $x_r$, $y_r$ for Example 2 (*Solid line* and *dot line* means trajectory of variables $x$, $y$ and trajectory of reference variables $x_r$, $y_r$, respectively)

*Example 2* $\min_x \max_y f(x, y) = xy$ (Ref. [13]),

From $(\partial f(x, y)/\partial y) = 0$, $(\partial f(x, y)/\partial x) = 0$ and definition 1, we see that point (0, 0) is the minmax optimization solution of $f(x, y)$. In the simulations, near minmax optimization solution, the initial values of variables $x$, $y$ and reference variables $x_r$, $y_r$ are taken as [(2.1, 3.1), (2.0, 1.1)], $\eta_1$, $\eta_2$ and $\eta_3$ are 0.01, 0.001, 0.0001, respectively, trajectories converge to point (0, 0) as shown in Fig. 5. As a result, the proposed reference variable method is convergent near minmax optimization solutions for this example, but T.L.Vincent's method is not convergent even if initial values of variables $x$ and $y$ are numbers close to Minmax optimization solutions, this indicates that the proposed method is very useful again.

*Example 3* $\min_x \max_y f(x, y) = x^2 - 8xy + 3x - 9y^2 + 7y - 1$ (Ref. [13]),

From $(\partial f(x^*, y)/\partial y) = 0$, $(\partial f(x, y^*)/\partial x) = 0$ and definition 1, we see that point (1/50, 19/50) is the minmax optimization solution of $f(x, y)$. In the simulations, initial values of variables $x$, $y$ and reference variables $x_r$, $y_r$ are generated from uniform distributions $[-2, 2]$, $\eta_1$, $\eta_2$ and $\eta_3$ are 0.02, 0.02, 0.1, respectively, trajectories converge to point (1/50, 19/50) as shown in Fig. 6. As a result, the proposed reference variable method is also convergent as with T.L.Vincent's method for this example.

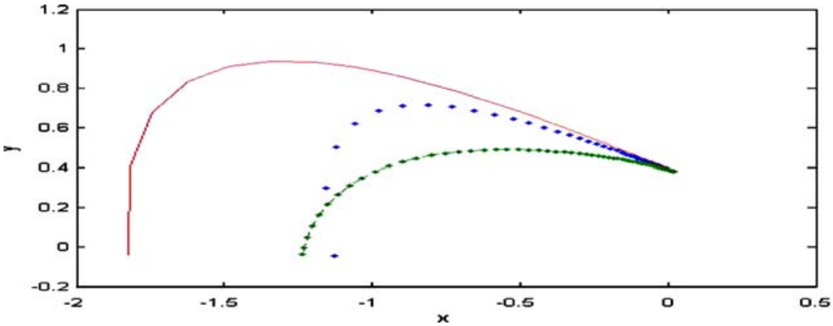


**Fig. 6** Trajectories converge to point (1/50, 19/50) for Example 3 (*solid line*, *dot line* and *dash-dot line* mean trajectory got by T.L.Vincent's method, trajectory of variables $x$, $y$, trajectory of reference variables $x_r$,$y_r$, respectively)

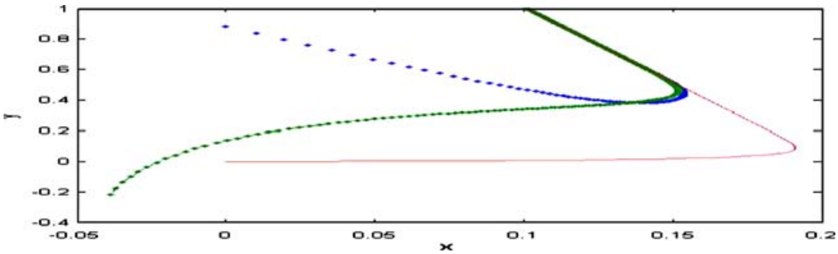


**Fig. 7** Trajectories converge to point (0.1, 1) for $g_1$ of Example 4 (*solid line*, *dot line* and *dash-dot line* mean trajectory got by T.L.Vincent's method, trajectory of variables $x$, $y$, trajectory of reference variables $x_r$, $y_r$, respectively)

*Example 4* That two firms sell substitutable products and seek to maximize their profits through advertising is given by (Ref. [13]),

$$g_1(x, y) = x + 5(x^2 - x)/(3 + y), g_2(x, y) \\ = x + 3(x^2 - x)/(2 + 2y) \text{ where } 0 \leq x, y \leq 1.$$

Applying the above methods to $g_1$ with $x$ as the minimizer and $y$ as the maximizer subject to the above inequalities, we see that point (0.1, 1) is the minmax optimization solution of $g_1$, initial values of variables $x$, $y$ and reference variables $x_r$, $y_r$ are generated from uniform distributions $[-1, 1]$, $\eta_1$, $\eta_2$ and $\eta_3$ are 0.04, trajectories converge to point (0.1, 1) as shown in Fig. 7 for min–max of $g_1$. Similarly, the trajectories starting from three different initial values are shown in Figs. 8–10 for min–max of $g_2$ with $x$ as the minimizer and $y$ as the maximizer subject to the above inequalities, $\eta_1$, $\eta_2$ and $\eta_3$ are 0.01. In the cases, we see that the trajectories converge to $y \in [0.5, 1]$ and $x = 0$ as with the results in Ref. [13].

*Example 5* $\min_x \max_y = \sin^2 x - x \cos y + 2 \sin x - \cos^2 y + y - 1$
Subject to $x^2 + y^2 \geq 25; \quad x, y \in [-5, 5]$,

In order to obtain all feasible solutions of Example 5, simulations with two cases: (1) Large-scale method. (2) A hybrid method of Largescale method and Hessian method (Hessian method uses results of Largescale method as initial values to run), are carried out starting at $M_1 < f(x, y) < M_2$ where $M_1 = -1{,}000{,}000$ and $M_2 = 30$. In Eq. 11, the initial values of variables are generated from uniform distributions [0, 4] for the two cases. Simulation results

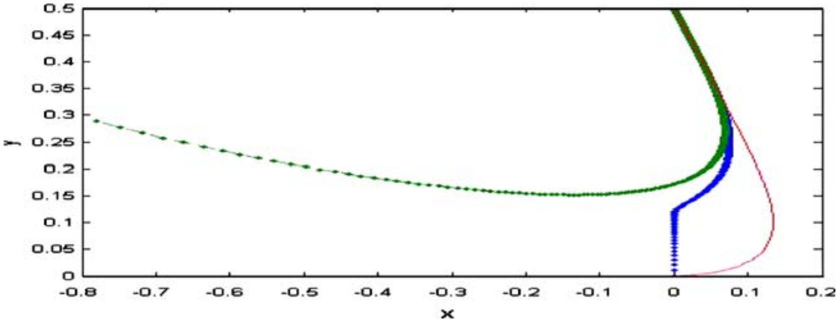


**Fig. 8** Trajectories starting from initial values with uniform distributions [−2, 2], converge to point (0, 0.5) for $g_2$ of Example 4 (*solid line*, *dot line* and *dash-dot line* mean trajectory got by T.L.Vincent's method, trajectory of variables $x$, $y$, trajectory of reference variables $x_r$, $y_r$, respectively)

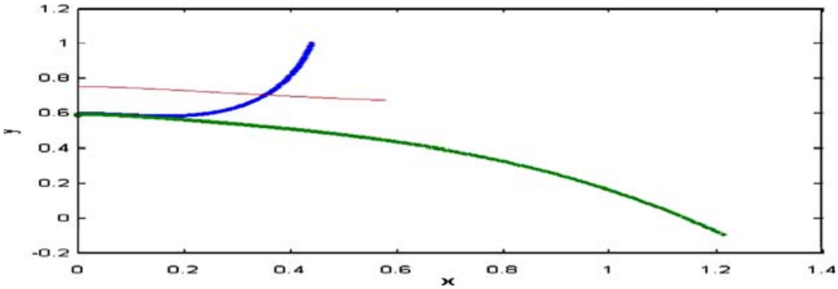


**Fig. 9** Trajectories starting from initial values with uniform distributions [−0.4, 1.6] for $g_2$ of Example 4, converge to point (0, 0.5803) and point (0, 0.7013), respectively by our method and T.L.Vincent's method (*solid line*, *dot line* and *dash-dot line* mean trajectory got by T.L.Vincent's method, trajectory of variables $x$, $y$, trajectory of reference variables $x_r$, $y_r$, respectively)

of the two cases are shown in Table 1, where constraints $x, y \in [-5, 5]$ are equivalent to $x^2 \leq 25$ and $y^2 \leq 25$.

In Table 1, symbol "–" means that no minmax optimization solution is found even if different initial values are used. Largescale 20+Hessian 70 means Hessian method runs 70 times after Largescale method runs 20 times. From this table, we see that computing time taken by using hybrid methods is shorter than that using single method, where hybrid method is only used for the cases in which those computing time taken by using Largescale method is longer. In addition, points (−0.88734, 5) and (3.6689, 3.6796) are minmax optimization solutions of Example 5 that are got by using different $M_1$ and $M_2$.

*Example 6* $\max_X \min_Y f(x, y) = x_1^2+x_2^2+x_1x_2-14x_1-16x_2+(y_6-10)^2+4(y_7-5)^2+(y_8-3)^2+2(y_1-1)^2+5y_2^2+6.5(y_3-11)^2+2(y_4-10)^2+(y_5-7)^2+45$

$$\text{s.t} \quad g_1 = 800 - 4x_1 - 5x_2 + 3y_2 - 9y_3 \geq 0,$$
$$g_2 = 800 - x_1^2 - 2(x_2-2)^2 + 2x_1x_2 - 14y_8 + 6y_1 \geq 0,$$
$$g_3 = 800 - 3(x_1-2)^2 - 4(x_2-3)^2 - 2y_6^2 + 7y_7 \geq 0,$$
$$g_4 = 800 - 10x_1 + 8x_2 + 17y_2 - 2y_3 \geq 0,$$
$$g_5 = 1000 + 8x_1 - 2x_2 - 5y_4 + 2y_5 \geq 0,$$

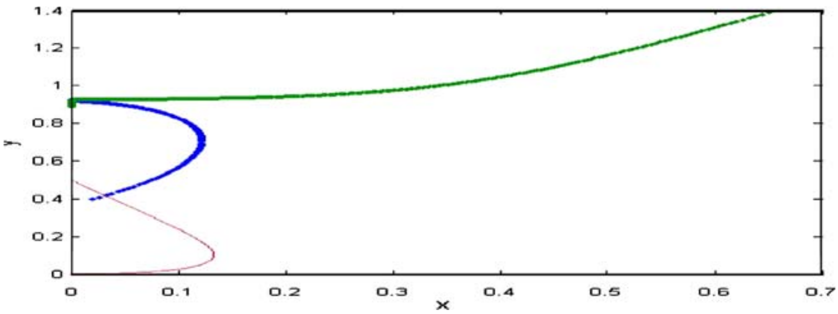


**Fig. 10** Trajectories starting from initial values with uniform distributions [−0.4, 1.6] for $g_2$ of Example 4, converge to point (0, 0.8874) and point (0, 0.5), respectively by our method and T.L.Vincent's method (*solid line*, *dot line* and *dash-dot line* mean trajectory got by T.L.Vincent's method, trajectory of variables $x$, $y$, trajectory of reference variables $x_r$, $y_r$, respectively)

**Table 1** The simulation results of Example 5

| Iteration method | Constraints of $f(x)$ | Minmax feasible solutions $(x_1, x_2)$ | $(\lambda_1, \lambda_2, \ldots, \lambda_5)$ | $f(X)$ | $\varepsilon$ | CPU time (s) |
|---|---|---|---|---|---|---|
| Largescale | [−1,000,000, 30] | (−0.88734, 5) | (0, 0, 0, 0, 0.13069) | 3.2217 | 0.1 | 2.3930 |
| | [3.2217, 30] | (3.6689, 3.6796) | (0., 0,…,0) | 4.3396 | | 3.1150 |
| | [−1,000,000, 3.2217] | – | – | – | | – |
| | [3.2217, 30] | | | | | |
| | [30, 1,000,000] | | | | | |
| Largescale 30+ Hessian 70 | [−1,000,000, 30] | (−0.88734, 5) | (0, 0, 0, 0, 0.13069) | 3.2217 | 0.1 | 0.4920 |
| Largescale 20+ Hessian 70 | [3.2217, 30] | (3.6689 3.6796) | (0., 0,…,0) | 4.3396 | | 0.19 |
| Largescale 20+ Hessian 70 | [−1,000,000, 3.2217] | – | – | – | | – |
| | [3.2217, 30] | | | | | |
| | [30, 1,000,000] | | | | | |

$$
\begin{aligned}
& g_6 = 800 - 5x_1^2 - 8x_2 - (y_6 - 6)^2 + 2y_7 \geq 0, \\
& g_7 = 800 + 3x_1 - 6x_2 - 12\,(y_4 - 8)^2 + 7y_5 \geq 0, \\
& g_8 = 800 - 0.5\,(x_1 - 8)^2 - 2\,(x_2 - 4)^2 - 3y_8^2 + y_1 \geq 0, \\
& -\sqrt{1{,}000} \leq x_i \leq \sqrt{1{,}000}, \quad -\sqrt{1{,}000} \leq y_j \leq \sqrt{1{,}000}, \quad i = 1, 2, \ \ j = 1, \ldots, 8.
\end{aligned}
$$

To solve Eq. 11, fsolve from the Matlab Optimization toolbox with the Largescale option is used. Simulations are carried out starting at $f(x, y) < M_2$ where $M_1 = 2{,}000$. In the simulations, the initial values of variables are generated from uniform distributions [0, 5]. Simulation results are shown in Table 2, where constraints $x_i, y_j \in \left\lfloor -\sqrt{1{,}000}, \sqrt{1{,}000} \right\rfloor$ are equivalent to $x_i^2 \leq 1{,}000$ and $y_j^2 \leq 1{,}000$.

In Table 2, symbol "–" means that no minmax optimization solution is found even if different initial values are used. Largescale 300 means that Largescale method runs 30 times. From this table, we see that point (4, 6, 1, 0, 11, 10.7, 10, 5, 3) is minmax optimization solution of Example 6, it taken 2.219 s.

**Table 2** The simulation results of Example 6

| Iteration method | Constraints of $f(x)$ | Minmax feasible solutions $(x_1, x_2, y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8)$ | $f(x, y)$ | $\varepsilon$ | CPU time (s) |
|---|---|---|---|---|---|
| Largescale 300 | $[-\infty, 2,000]$ | (4, 6, 1, 0, 11, 10.7, 10, 5, 3) | −31 | 0.1 | 2.2190 |
| Largescale 300 | $[-\infty, -31]$ | – | – | | – |
| Largescale 300 | $[-31, 2,000]$ | – | – | | – |
| Largescale 300 | $[2,000, \infty]$ | – | – | | – |

## 5 Conclusions

In this paper, new methods based on reference decision vectors have been proposed for solving unconstrained minmax problems or minmax problems with bounds on decision vectors. At the same time, a new method based on necessary conditions similar to KKT conditions of min or max constrained optimization problems is also given for solving constrained minmax optimization problems, in which all minmax optimization solutions of constrained minmax optimization problems can be got by using different $M_1$ and $M_2$ when the cost function $f(x, y)$ is constrained as $M_1 < f(x, y) < M_2$ where $M_1$ and $M_2$ are real numbers. To show effectiveness of these methods, some examples are taken to compare with results in the literature, as a result, it has superiority over the existing methods in the literature. This indicates that the proposed method is very useful.

## Appendix A

In the proof of theorems 1 and 2, the following Lemma 1 is used.

**Lemma 1** *If a matrix A is Hermite matrix and* $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n (\lambda_1, \lambda_2, \ldots, \lambda_n$ *are eigenvalues of matrix A*), *then for arbitrary vector X such that* $\lambda_1 X^* X \le X^* A X \le \lambda_n X^* X$ (*for proof omit here*).

The proof of Theorem 1

*Proof* From $\Delta x = -\eta(\partial f(x, y)/\partial x)$, $\Delta y = \eta(\partial f(x, y)/\partial y)$ and Taylor expansion at point $(x^*, y^*)$, we have,

$$\Delta x(k) = -\eta \frac{\partial^2 f(x^*, y^*)}{\partial^2 x}(x - x^*) - \eta \frac{\partial^2 f(x^*, y^*)}{\partial xy}(y - y^*) + O(x - x^*, y - y^*)$$

$$\Delta y(k) = \eta \frac{\partial^2 f(x^*, y^*)}{\partial yx}(x - x^*) + \eta \frac{\partial^2 f(x^*, y^*)}{\partial^2 y}(y - y^*) + O(x - x^*, y - y^*)$$

It can be written as

$$\Delta 1(k) = \Delta 1(k-1) - \eta \frac{\partial^2 f(x^*, y^*)}{\partial^2 x} \Delta 1(k-1) - \eta \frac{\partial^2 f(x^*, y^*)}{\partial xy} \Delta 2(k-1) + O(|x - x^*|, |y - y^*|)$$

$$\Delta 2(k) = \Delta 2(k-1) + \eta \frac{\partial^2 f(x^*, y^*)}{\partial yx} \Delta 1(k-1) + \eta \frac{\partial^2 f(x^*, y^*)}{\partial^2 y} \Delta 2(k-1) \\ + O(|x - x^*|, |y - y^*|)$$

where $\Delta 1(k) = x(k) - x^*$, $\Delta 2(k) = y - y^*$ and $(\partial^2 f(x, y)/\partial xy) = (\partial f(x, y)/\partial yx)$ for continuously differentiable function $f(x, y)$.

It follows that near $(x^*, y^*)$, we have $\Delta(k) \approx A1\Delta 3(k-1)$ where $\Delta(k) = [\Delta 1(k), \Delta 2(k)]^T$, $\Delta 3(k-1) = [\Delta 1(k-1), -\Delta 2(k-1)]$ and

$$A = \begin{bmatrix} I - \eta \frac{\partial^2 f(x^*, y^*)}{\partial^2 x} & \eta \frac{\partial^2 f(x^*, y^*)}{\partial xy} \\ \eta \frac{\partial^2 f(x^*, y^*)}{\partial xy} & -I + \eta \frac{\partial^2 f(x^*, y^*)}{\partial^2 y} \end{bmatrix}.$$

because $A$ is Hermite matrix, arbitrary eigenvalue of $A * A$ is the square of eigenvalue of $A$. By Lemma 1, thus we have $\min |Z|^2 \, \|\Delta 3(k-1)\| \leq \|\Delta(k)\| \approx \|A1\Delta 3(k-1)\| \leq \max |Z|^2 \, \|\Delta 3(k-1)\|$, because $||\Delta(k-1)|| = ||\Delta 3(k-1)||$, if $|Z| < 1$ holds for arbitrary eigenvalue, then $\Delta(k)$ is convergent as $k \to \infty$ near point $(x^*, y^*)$; if $|Z| > 1$ holds for arbitrary eigenvalue, then $\Delta(k)$ is not convergent as $k \to \infty$ near point $(x^*, y^*)$, thus the proof of this theorem is complete. □

The proof of Theorem 2

*Proof* By Taylor expansion at point $(x^*, y^*)$, we have

$$\frac{\partial f(x, y)}{\partial x} = \frac{\partial f(x^*, y^*)}{\partial x} + \frac{\partial^2 f(x^*, y^*)}{\partial^2 x}(x(k-1) - x^*) + \frac{\partial^2 f(x^*, y^*)}{\partial xy}(y(k-1) - y^*) \\ + O(x - x^*, y - y^*)$$

$$\frac{\partial f(x, y)}{\partial y} = \frac{\partial f(x^*, y^*)}{\partial y} + \frac{\partial^2 f(x^*, y^*)}{\partial xy}(x(k-1) - x^*) + \frac{\partial^2 f(x^*, y^*)}{\partial^2 y}(y(k-1) - y^*) \\ + O(x - x^*, y - y^*)$$

Thus, from Eq. 6, we have near $(x^*, y^*)$,

$$x(k) \approx x(k-1) - \eta_1 \frac{\partial^2 f(x^*, y^*)}{\partial^2 x}(x(k-1) - x^*) - \eta_1 \frac{\partial^2 f(x^*, y^*)}{\partial xy}(y(k-1) - y^*) \\ + \eta_2(x(k-1) - x(k-2))$$

$$y(k) \approx y(k-1) + \eta_1 \frac{\partial^2 f(x^*, y^*)}{\partial xy}(x(k-1) - x^*) + \eta_1 \frac{\partial^2 f(x^*, y^*)}{\partial^2 y}(y(k-1) - y^*) \\ + \eta_2(y(k-1) - y(k-2))$$

It follows that $\begin{matrix} \Delta 1(k) \approx \Delta 1(k-1) - a\Delta 1(k-1) - b\Delta 2(k-2) + \eta_2 \Delta 1(k-1) - \eta_2 \Delta 1(k-2)) \\ \Delta 2(k) \approx \Delta 2(k-1) + b\Delta 1(k-1) + c\Delta 2(k-1) - \eta_2 \Delta 2(k-1) + \eta_2 \Delta 2(k-1)) \end{matrix}$,

$$\begin{bmatrix} \Delta 1(k) \\ \eta_2^{0.5} \Delta 1(k-1) \\ \Delta 2(k) \\ -\eta_2^{0.5} \Delta 2(k-1) \end{bmatrix} \approx A1 \begin{bmatrix} \Delta 1(k-1) \\ -\eta_2^{0.5} \Delta 1(k-2) \\ -\Delta 2(k-1) \\ -\eta_2^{0.5} \Delta 2(k-2) \end{bmatrix},$$

$$A1 = \begin{bmatrix} I + \eta_2 I - a & \eta_2^{0.5} I & b & 0 \\ \eta_2^{0.5} I & 0 & 0 & 0 \\ b & 0 & -I + \eta_2 I - c & \eta_2^{0.5} I \\ 0 & 0 & \eta_2^{0.5} I & 0 \end{bmatrix}$$

where $\|(\Delta 1(k),\ \eta_2^{0.5}\Delta 1(k-1),\ \Delta 2(k),\ -\eta_2^{0.5}\Delta 2(k-1))\| = \|(\Delta 1(k), \eta_2^{0.5}\Delta 1(k-1), \Delta 2(k), \eta_2^{0.5}\ \Delta 2(k-1))\|\ \|(\Delta 1(k),\ -\eta_2^{0.5}\ \Delta 1(k-1),\ -\Delta 2(k),\ -\eta_2^{0.5}\ \Delta 2(k-1))\| = \|(\Delta 1(k),\ \eta_2^{0.5}\ \Delta 1(k-1),\ \Delta 2(k),\ \eta_2^{0.5}\ \Delta 2(k-1))\|$, $a = \eta_1(\partial^2 f(x^*, y^*)/\partial^2 x), b = \eta_1(\partial^2 f(x^*, y^*)/\partial xy), c = \eta_1(\partial^2 f(x^*, y^*)/\partial^2 y), \Delta 1(.) = x(.)-x^*$ and $\Delta 2(.) = y(.)-y^*$. Because there exists a $||.||$ for arbitrary given $\varepsilon > 0$ such that $\max\{|\lambda(A1)|\} \le \|A1\| \le \max\{|\lambda(A1)|\} + \varepsilon$(where $\lambda$ is an eigenvalue of matrix $A$) holds, if

$$\begin{vmatrix} I+\eta_2 I - a - ZI & \eta_2^{0.5} I & b & 0 \\ \eta_2^{0.5} I & -ZI & 0 & 0 \\ b & 0 & -I+\eta_2 I - c - ZI & \eta_2^{0.5} I \\ 0 & 0 & \eta_2^{0.5} I & -ZI \end{vmatrix} = 0, |Z| < 1, \text{ that is,}$$

$$\begin{vmatrix} -Z^2 I + (1+\eta_2 - a)Z + \eta_2 I & bZ \\ bZ & -Z^2 I + (1+\eta_2 - c)Z + \eta_2 I \end{vmatrix} = 0, |Z| < 1,$$

then $\|A1\| < 1$, this implies that $\|\Delta 1(k)\| \to 0$, $\|\Delta 2(k)\| \to 0$ as $k \to \infty$. In addition, because $A1$ is Hermite matrix, any eigenvalue of $A1 * A1$ is the square of eigenvalue of $A1$, thus by Lemma 1, we have

$$\min |Z|^2\ \|\Delta 3(k-1)\| \le \|\Delta(k)\| \approx \|A1\Delta 3(k-1)\| \le \max |Z|^2\ \|\Delta 3(k-1)\|$$

where $\Delta(k) = [\Delta 1(k), \eta_2^{0.5}\Delta 1(k-1), \Delta 2(k), -\eta_2^{0.5}\Delta 2(k-1)]$, $\Delta 3(k-1) = [\Delta 1(k-1), -\eta_2^{0.5}\Delta 1(k-2), -\Delta 2(k-1), -\eta_2^{0.5}\Delta 2(k-2)]$

Because $||\Delta(k-1)|| = ||\Delta 3(k-1)||$, If $|Z| > 1$ for arbitrary eigenvalue, then $\Delta(k)$ is not convergent as $k \to \infty$ near point $(x^*, y^*)$ thus, the proof of this theorem is complete. □

The proof of Theorem 3

*Proof* Assuming that point $(x^*, y^*)$ is a solution of $(\partial f(x^*, y^*)/\partial x) = (\partial f(x^*, y^*)/\partial y) = 0$, by mean value theory of function, we have,

$$\frac{\partial f(x, y)}{\partial x} \approx \frac{\partial f^2(x^*, y^*)}{\partial x^2}(x - x^*) + \frac{\partial^2 f(x^*, y^*)}{\partial xy}(y - y^*), \frac{\partial f(x, y)}{\partial y} = \frac{\partial f^2(x^*, y^*)}{\partial xy}(y - y^*) + \frac{\partial^2 f(x^*, y^*)}{\partial y^2}(y - y^*)$$

According to Eq. 8, we have

$$\begin{aligned} x(k)-x^* &\approx x(k-1)-x^*+\eta_1 \tfrac{\partial^2 f(x^*,y^*)}{\partial x^2}(x(k-1)-x^*)+\eta_1 \tfrac{\partial^2 f(x^*,y^*)}{\partial xy}(y(k-1)-y^*) \\ &\quad -\eta_2(x(k-1)-x^*-x_r(k-1)+x^*) \\ y(k)-y^* &\approx y(k-1)-y^*-\eta_1 \tfrac{\partial^2 f(x^*,y^*)}{\partial y^2}(y(k-1)-y^*)]-\eta_1 \tfrac{\partial^2 f(x^*,y^*)}{\partial yx}(x(k-1)-x^*) \\ &\quad -\eta_2(y(k-1)-y^*-y_r(k-1)+y^*) \\ x_r(k)-x^* &= x_r(k-1)-x^*+\eta_3(x(k-1)-x^*-x_r(k-1)+x^*),\ y_r(k)-y^* = y_r-y^* \\ &\quad +\eta_3(y(k-1)-y^*-y_r(k-1)+y^*) \end{aligned}$$

It follows that $\Delta(k) \approx A\Delta(k-1)$, $\Delta(k) = \left[\ \Delta x(k)\ \ \Delta x_r(k)\ \ \Delta y(k)\ \ \Delta y_r(k)\ \right]^T$

$$\text{where } A = \begin{bmatrix} (1-\eta_2)I+\eta_1 \frac{\partial^2 f(x^*,y^*)}{\partial x^2} & \eta_2 I & \eta_1 \frac{\partial^2 f(x^*,y^*)}{\partial xy} & 0 \\ \eta_3 I & (1-\eta_3)I & 0 & 0 \\ \eta_1 \frac{\partial^2 f(x^*,y^*)}{\partial yx} & 0 & (1-\eta_2)I-\eta_1 \frac{\partial^2 f(x^*,y^*)}{\partial y^2} & \eta_2 I \\ 0 & 0 & \eta_3 I & (1-\eta_3)I \end{bmatrix}$$

$$A_1 = \begin{bmatrix} (1-\eta_2)I & \eta_2 I & \eta_1 \frac{\partial^2 f}{\partial xy} & 0 \\ 0 & (1-\eta_3)I & 0 & 0 \\ 0 & 0 & (1-\eta_2)I & \eta_2 I \\ 0 & 0 & 0 & (1-\eta_3)I \end{bmatrix}, \quad A_2 = \begin{bmatrix} \eta_1 \frac{\partial^2 f}{\partial x^2} & 0 & 0 & 0 \\ \eta_3 I & 0 & 0 & 0 \\ \eta_1 \frac{\partial^2 f}{\partial yx} & 0 & -\eta_1 \frac{\partial^2 f}{\partial f^2} & 0 \\ 0 & 0 & \eta_3 I & 0 \end{bmatrix}$$

thus $A = A_1 + A_2$, $\|A\| \leq \|A_1\| + \|A_2\|$. In addition, because there exists a $||.||$ such that $\|A\| \leq \rho(A) + \varepsilon$, $\rho(A) = \max\{|\lambda|, \lambda \in \lambda(A)\}$, $A \in C^{n*n}$ where $\lambda$ is an eigenvalue vector of matrix $A$, and $\varepsilon$ is arbitrary small positive number, $\|A_1\| \leq \rho(A_1) + \varepsilon_1$, $\|A_2\| \leq \rho(A_2) + \varepsilon_2$. If $\rho(A_1) + \rho(A_2) < 1$, then $||A|| < 1$, this implies that Eq. 8 is convergent. From $|A_1 - \lambda(A_1)| = 0$, $|A_2 - \lambda(A_2)| = 0$, we have $(1-\eta_2-\lambda(A_1)) = 0$, $(1-\eta_3-\lambda(A_1)) = 0$ and $|\eta_1(\partial^2 f/\partial x^2) - \lambda(A_2)||\eta_1(\partial^2 f/\partial y^2) + \lambda(A_2)|\lambda(^2 A_2) = 0$, This implies that $\lambda(A_1) = 1-\eta_2$, $1-\eta_3$, $\lambda(A_2) = 0$, $|\eta_1(\partial^2 f/\partial x^2) - \lambda(A_2)| = 0$, $|\eta_1(\partial^2 f/\partial y^2) + \lambda(A_2)| = 0$, thus we have $|\lambda(A_2)| \leq \eta_1 \left\|(\partial^2 f/\partial x^2)\right\|$ and $|\lambda(A_2)| \leq \eta_1 \left\|(\partial^2 f/\partial y^2)\right\|$. Because $\left\|(\partial^2 f/\partial x^2)\right\|$ and $\left\|(\partial^2 f/\partial y^2)\right\|$ are bounded, if $\eta_1$is very small, then $|\lambda(A_2)| << 1$, it follows that $\|A\| \leq \max\{|1-\eta_2|, |1-\eta_3|) + \varepsilon$, $\varepsilon$ is arbitrary small positive number. If $\eta_1$, $\eta_2$ and $\eta_3$ are chosen as numbers such that $\max\{|1-\eta_2|, |1-\eta_3|\} + \varepsilon < 1$, then $||A|| < 1$, this implies that Eq. 8 is convergent, for example $\eta_1$ is chosen as a very small number and $\eta_2, \eta_3$ are chosen as number close to 1. Thus the proof of this theorem is complete. □

The proof of Theorem 5

*Proof* To prove Theorem 5, the following lemmas are used.

Gordan lemma: Given vectors $A_1, A_2, \ldots, A_l \in R^N$, there exists no zero vector $\mu \in R^L$ with components $\mu_i$, $i = 1, 2, \ldots, L$such that$\sum_{j=1}^{l} \mu_i A_i = 0$ holds if and only if there dose not exist vector $P$ such that $A_i^T P < 0$, $j = 1, 2, \ldots, L$ hold (For proof omit).

**Lemma 2** *Assuming* $J = \{j | g_j(\bar{X}) = 0,\ 1 \leq j \leq m\}$, *there exists* $\lambda_0 > 0$ *such that* $g_i(\bar{X} + \lambda_P) > 0$, $j \in J$ *for a vector P and arbitrary* $\lambda_\in[0, \lambda_0]$ *if and only if* $\nabla g_i(\bar{X})^T P > 0$, $j \in J$ *hold.*

*Proof* For a vector $P$, $\lambda_0 > 0$ and arbitrary $\lambda_\in[0, \lambda_0]$, by the Taylor expansion, we have $g_i(\bar{X}+\lambda_P) = g_i(\bar{X}) + \lambda\nabla g_i(\bar{X})^T P + o(\lambda)$, $j \in J$. If $\lambda$ is very small number and $\nabla g_i(\bar{X})^T P > 0$, $j \in J$, then $g_i(\bar{X}+\lambda_P) > 0$, $j \in J$ holds; on the other hand, if $g_i(\bar{X}+\lambda_P) > 0$, $j \in J$ holds, then $g_i(\bar{X}+\lambda_P) - g_i(\bar{X}) > 0$, $j \in J$ and $(dg_i(\bar{X}+\lambda P)/d\lambda) = \nabla g_i(\bar{X})^T P > 0$, as $\lambda \to 0$. (The proof is complete).

**Lemma 3** *If there exists* $\lambda_0 > 0$ *such that* $f(\bar{X}+\lambda P, Y) > f(\bar{X}, Y)$ *for a vector P and arbitrary* $\lambda_\in[0,\ \lambda_0]$ *and Y if and only if* $\nabla f(\bar{X})^T P > 0$ *holds.*

*Proof* For a vector $P$, $\lambda_0 > 0$ and arbitrary $\lambda_\in[0, \lambda_0]$, by the Taylor expansion, we have $f(\bar{X}+\lambda P, Y) = f(\bar{X}, Y) + \lambda\nabla f(\bar{X}, Y)^T P + o(\lambda)$, if $\lambda$ is very small number and $\nabla f(\bar{X})^T P > 0$, then $f(\bar{X}+\lambda P, Y) > f(\bar{X}, Y)$. In the other hand, if $f(\bar{X}+\lambda P, Y) > f(\bar{X}, Y)$, then $f(\bar{X}+\lambda P, Y) - f(\bar{X}, Y) > 0$ and $df(\bar{X}+\lambda P, Y)/d\lambda = \nabla f(\bar{X}, Y)^T P > 0$, as $\lambda \to 0$. (the proof is complete). □

**Lemma 4** *If there exists* $\lambda_0 > 0$ *such that* $f(X, \bar{Y}+\lambda P) < f(X, \bar{Y})$ *for a vector P and arbitrary* $\lambda_\in[0,\ \lambda_0]$*and X if and only if* $\nabla f(X, \bar{Y})^T P < 0$ *holds (the proof, omit).*

Because point $(x^*, y^*)$ is an optimization solution of problem (3), $f(x^*, y) \leq f(x^*, y^*) \leq f(x, y^*)$, and at point $(x^*, y^*)$, there do not exit $P1$ and $P2$ such that $\begin{cases} \nabla_x f(x^*, y^*)P1 < 0 \\ \nabla_x g_i(x^*, y^*)P1 > 0 \end{cases}$ and $\begin{cases} \nabla_y f(x^*, y^*)P2 > 0 \\ \nabla_y g_i(x^*, y^*)P2 > 0 \end{cases}$ hold, thus the proof of this theorem is complete.

The proof of Theorem 6

Assuming $(\bar{X}, \bar{Y})$ is a optimization solution of problem (3), by Gordan lemma and Theorem 5, except that $\lambda_i = 0, \gamma_i = 0, g_i(\bar{X}, \bar{Y}) > 0$, we have

(a) $g_i(\bar{X}, \bar{Y}) = 0, j \in J$,
(b) $\lambda_0 \nabla f_x(\bar{X}, \bar{Y}) - \sum_i^J \lambda_i \nabla_x g_i(\bar{X}, \bar{Y}) = 0$,
(c) $-\gamma_0 \nabla f_y(\bar{X}, \bar{Y}) - \sum_i^J \gamma_i \nabla_y g_i(\bar{X}, \bar{Y}) = 0$,

Because the set $\{\nabla g_j | \lambda_j \neq 0, \gamma_j \neq 0\}$ is linearly independent, it follows that

(a) $g_i(\bar{X}, \bar{Y}) = 0, j \in J$,
(b) $\nabla f_x(\bar{X}, \bar{Y}) - \sum_i^J \lambda_i \nabla_x g_i(\bar{X}, \bar{Y}) = 0$,
(c) $\nabla f_y(\bar{X}, \bar{Y}) + \sum_i^J \gamma_i \nabla_y g_i(\bar{X}, \bar{Y}) = 0$,

thus the proof of this theorem is complete.

**Acknowledgements** This work is supported by Shanghai leading academic discipline project (T0103).

## References

1. Shi, D.S., Ling, C.: Minmax theorems and cone saddle point of uniformly same -order vector valued functions. J. Optim. Theory Appl. **84**(3), 575–587 (1995)
2. Greco, G.H., Horvath, C.D.: A topological minimax theorem. J. Optim. Theory Appl. **113**(3), 513–536 (2002)
3. Frenk, J.F.B., Kassay, G.: Minimax results and finite-dimensional separation. J. Optim. Theory Appl. **113**(2), 409–421 (2002)
4. Ghosh, M.K., McDonald, D., Sinha, S.: Zero-sum stochastic games with partial information. J. Optim. Theory Appl. **121**(1), 99–118 (2004)
5. Frenk, J.B.G., Kassay, G., Kolumban, J.: On equivalent results in minmax theory. Eur. J. Oper. Res. **157**, 46–58 (2004)
6. Achtziger, W.: Multiple local truss topology and sizing optimization some properties of minmax compliance. J. Optim. Theory Appl. **98**(2), 255–277 (1998)
7. Tan, K.K., Yu, J.: New minmax inequality with applications to existence theorems of equilibrium point. J. Optim. Theory Appl. **82**(1), 105–120 (1994)
8. Polak, E., Royset, J.O.: Algorithms for finite and semi-infinite min–max–min problems using adaptive smoothing techniques. J. Optim. Theory Appl. **119**(3), 421–457 (2003)
9. Polak, E., Royset, J.O.: Algorithms with adaptive smoothing for finite minimax problems. J. Optim. Theory Appl. **119**(3), 459–484 (2003)
10. Yu, Y.H., Gao, L.: Nonmonotone line search algorithm for constrained minimax problems. J. Optim. Theory Appl. **115**(2), 419–446 (2002)
11. Yu, G.: Min–max optimization of several classical discrete optimization problems. J. Optim. Theory Appl. **98**(1), 221–242 (1998)
12. Frenk, J.B.G., Gromicho, J., Zhang, S.: General models in minmax continuous localtion theory and solution thechniques. J. Optim. Theory Appl. **89**(1), 39–63 (1996)
13. Vincent, T.L., Goh, B.S.: Trajectory-following algorithms for min–max optimization problem. J. Optim. Theory Appl. **75**(3), 501–519 (1992)
14. Fischer, A.: New constrained optimization reformulation of complementarity problems. J. Optim. Theory Appl. **97**(1), 105–117 (1998)

15. Pu, D., Tian, W.: Globally convergent inexact generalized Newton's methods for nonsmooth equations. J. Comput. Appl. Math. **138**, 37–49 (2002)
16. Neumaier, A.: Introduction to numerical analysis. New York Cambridge university press, Cambridge UK (2001)
17. Hernandez, M.A., Rubio, M.J.: A modification of Newton's method for non-differentiable equations. J. Comput. Appl. Math. **164–165**, 409–417 (2004)
18. Amat, S., Busquier, S., Gutierrez, J.M.: Geometric constructions of iterative functions to solve nonlinear equations. J. Comput. Appl. Math. **157**, 197–205 (2003)
19. Xiaojun, C.: Superlinear convergence of smoothing quasi-Newton methods for non-smooth equations. J. Comput. Appl. Math. **80**, 105–126 (1997)
20. Xu, J.-j.: Convergence of partially asynchronous block quasi-Newton methods for nonlinear systems of equations. J. Comput. Appl. Math. **103**, 307–321 (1999)
21. Andreas, F.: Asynchronous parallel for enclosing solutions of nonlinear equations. J. Comput. Appl. Math. **60**, 47–62 (1995)