Convergence of the Nelder-Mead simplex method to a non-stationary point

K.I.M. McKinnon

MS 96-006

May 1996

(Submitted to SIAM Journal on Optimization)

Department of Mathematics and Statistics
University of Edinburgh, The King's Buildings, Edinburgh EH9 3JZ
Tel. (33) 131 650 5042 E-Mail : ken@maths.ed.ac.uk

CONVERGENCE OF THE NELDER-MEAD SIMPLEX METHOD TO A NON-STATIONARY POINT

K.I.M. MCKINNON*

Abstract. This paper analyses the behaviour of the Nelder-Mead simplex method for a family of examples which cause the method to converge to a non-stationary point. All the examples use continuous functions of two variables. The family of functions contains strictly convex functions with up to three continuous derivatives. In all the examples the method repeatedly applies the inside contraction step with the best vertex remaining fixed. The simplices tend to a straight line which is orthogonal to the steepest descent direction. It is shown that this behaviour cannot occur for functions with more than three continuous derivatives. The stability of the examples is analysed.

Key words. Nelder-Mead method, direct search, simplex, unconstrained minimization

AMS subject classifications. 65K05

1. Introduction. Direct search methods are very widely used in chemical engineering, chemistry and medicine. They are a class of optimization methods which are easy to program, do not require derivatives and are often claimed to be robust for problems with discontinuities or where the function values are noisy. Until recently there was very little theory for these methods. In [6] Torczon produced convergence results for a class of methods called pattern search methods. This class includes several well known direct search methods such as the Spendley Hext and Himsworth simplex method [4], but does not include the most commonly used method, the Nelder-Mead simplex method [3]. In the Nelder-Mead method the simplex can vary in shape from iteration to iteration. Nelder and Mead introduced this feature to allow the simplex to adapt its shape to the local contours for the function, and for many problems this seems to be effective. However it is this change of shape which excludes the Nelder-Mead method from the class of methods covered by the convergence results of Torczon [6].

The Nelder-Mead method uses a small number of function evaluations per iteration and it is the most widely used direct search method. In [5, 1] however Torczon and Dennis report results from tests in which the Nelder-Mead method frequently failed to converge to a local minimum of smooth functions of low dimension. In the cases where failure occured the simplex defining the possible search directions became degenerate, so restriciting the search to a subspace. Until recently there has been little understanding of the convergence behaviour of the Nelder-Mead method. Recently however Tseng [7] introduced a class of simplicial search methods which includes a variant on the Nelder-Mead method and proved convergence to stationary points for continuously differentiable functions. The modification has the effect of preventing the simplex becoming arbitrarily narrow. In a recent report Lagarias et al [2] derive a range of convergence results and prove that the original Nelder-Mead method converges for strictly convex functions of one variable and also prove that for the two variables case the simplex diameters converge to zero. However it is not yet known, even for strictly convex quadratic functions of two variables, whether the method converges to a single point, nor is it know whether it always converges to the minimum of $x^2 + y^2$.

*Deptment of Mathematics and Statistics, The University of Edinburgh, King's Buildings, Edinburgh, EH9 3JZ, UK (ken@maths.ed.ac.uk).

The current paper presents a family of examples of functions of two variables where convergence occurs to a non-stationary point for a range of starting simplices. Some examples have a discontinuous first derivative and others are strictly convex with between one and three continuous derivatives. The simplices converge to a line which is orthogonal to the steepest descent direction. The simplices become arbitrarily narrow, so for these examples the behaviour of the original Nelder-Mead method is different from the modification introduced by Tseng [7].

We assume that the problem to be solved is

$$\min_{v \in \mathbb{R}^2} f(v)$$

For functions defined over $\mathbb{R}^2$ (i.e. functions of two variables) the Nelder-Mead method operates with a simplex in $\mathbb{R}^2$, which is specified by its three vertices. The Nelder-Mead method is described below for the two variable case and without the termination test. The settings for the parameter ρ in $L(\rho)$ are the most commonly used values. A fuller description of the method can be found in the papers by Lagarias et al [2] and Nelder and Mead [3].

The Nelder-Mead method

ORDER: Label the three vertices of the current simplex b, s and w so that their
  corresponding function values f_b, f_s and f_w satisfy $f_b \leq f_s \leq f_w$.
  $m := (b+s)/2$, {the midpoint of the best and second best points}.
  Let $L(\rho)$ denote the function $L(\rho) = m + \rho(m - w)$, {$L$ is the search line}.
  $r := L(1)$; $f_r := f(r)$.
  If $f_r < f_b$ then
    $e := L(2)$; $f_e := f(e)$.
    If $f_e < f_r$ then accept e { Expand} else accept r {Reflect}.
  else $\{f_b \leq f_r\}$ if $f_r < f_s$ then
    Accept r {Reflect}.
  else $\{f_s \leq f_r\}$ if $f_r < f_w$ then
    $c := L(0.5)$; $f_c := f(c)$.
    If $f_c \leq f_r$ then accept c {Outside Contract} else $\rightarrow$ SHRINK.
  else $\{f_w \leq f_r \}$
    $c := L(-0.5)$; $f_c := f(c)$.
    If $f_c < f_w$ then accept c {Inside Contract} else $\rightarrow$ SHRINK.
  Replace w by the accepted point; $\rightarrow$ ORDER.
SHRINK: Replace s by $(s+b)/2$ and w by $(w+b)/2$; $\rightarrow$ ORDER.

The examples in this paper cause the Nelder-Mead method to apply the inside contraction step repeatedly with the best vertex remaining fixed. This behaviour will be referred to as repeated focussed inside contraction (RFIC). No other type of step occurs for these examples and this greatly simplifies their analysis. The examples are very simple and highlight a serious deficiency in the method: the simplices shrink along the steepest descent direction, a direction along which we would like them to expand.

The structure of this paper is as follows. In section 2 the sequence of simplices is derived corresponding to RFIC. In section 3 a family of functions are given which produce this behaviour and result in the method converging to a non-stationary point. In section 4 the range of functions which can give the RFIC behaviour is derived.

Section 5 contains an analysis of how perturbations of the initial simplex affect the RFIC behaviour of the examples in section 3.

2. Analysis of the repeated inside contraction behaviour. Consider a simplex in two dimensions with vertices at 0 (i.e. the origin), $v^{(n+1)}$ and $v^{(n+1)}$. Assume that

$$f(0) < f(v^{(n+1)}) < f(v^{(n)}). \tag{2.1}$$

After the ORDER step of the algorithm, $b = 0$, $s = v^{(n+1)}$ and $w = v^{(n)}$. The Nelder-Mead method calculates $m^{(n)} = v^{(n+1)}/2$, the midpoint of the line joining the best and second best points, and then reflects the worst point, $v^{(n)}$, in $m^{(n)}$ with a reflection factor of $\rho = 1$ to give the point

$$r^{(n)} = m^{(n)} + \rho(m^{(n)} - v^{(n)}) = v^{(n+1)} - v^{(n)}. \tag{2.2}$$

Assume that

$$f(v^{(n)}) < f(r^{(n)}). \tag{2.3}$$

In this case the point $r^{(n)}$ is rejected and the point $v^{(n+2)}$ is calculated using a reflection factor $\rho = -0.5$ in

$$v^{(n+2)} = m^{(n)} + \rho(m^{(n)} - v^{(n)}) = \frac{1}{4}v^{(n+1)} + \frac{1}{2}v^{(n)}.$$

$v^{(n+2)}$ is the midpoint of the line joining $m^{(n)}$ and $v^{(n)}$. Provided (2.1) holds for $v^{(n+1)}$ and $v^{(n+2)}$, the Nelder-Mead method does the inside contraction step rather than a shrink step. The inside contraction step replaces $v^{(n)}$ with the point $v^{(n+2)}$, so that the new simplex consists of $v^{(n+1)}$, $v^{(n+2)}$ and the origin. Provided this pattern repeats, the successive simplex vertices will satisfy the linear recurrence relation

$$4v^{(n+2)} - v^{(n+1)} - 2v^{(n)} = 0.$$

This has the general solution

$$v^{(n)} = A_1\lambda_1^n + A_2\lambda_2^n, \tag{2.4}$$

where $A_i \in \mathbb{R}^2$ and

$$\lambda_1 = \frac{1+\sqrt{33}}{8}, \lambda_2 = \frac{1-\sqrt{33}}{8}. \tag{2.5}$$

Hence $\lambda_1 \cong 0.84307$ and $\lambda_2 \cong -0.59307$. It follows from (2.2) and (2.4) that

$$r^{(n)} = -A_1\lambda_1^n(1-\lambda_1) - A_2\lambda_2^n(1-\lambda_2). \tag{2.6}$$

It is this repeated inside contraction towards the same fixed vertex which is being referred to as repeated focussed inside contraction (RFIC). Provided the Nelder-Mead method is started from a non-degenerate initial simplex, then no later simplex can be degenerate (see Lagarias et al [2]), so if RFIC occurs then the initial simplex for RFIC is non-degenerate, which implies that A_1 and A_2 in (2.4) are linearly independent.

Consider now the initial simplex with vertices $v^{(0)} = (1,1), v^{(1)} = (\lambda_1, \lambda_2)$ and $(0,0)$. Substituting into (2.4) yields $A_1 = (1,0)$ and $A_2 = (0,1)$. It follows that the particular solution for these initial conditions is $v^{(n)} = (\lambda_1^n, \lambda_2^n)$. This solution is shown in figure 2.1. The general form of the three points needed at one step of the Nelder-Mead method is therefore

$$v^{(n)} = (\lambda_1^n, \lambda_2^n), \tag{2.7}$$

$$v^{(n+1)} = (\lambda_1^{n+1}, \lambda_2^{n+1}), \tag{2.8}$$

$$r^{(n)} = (-\lambda_1^n(1-\lambda_1), -\lambda_2^n(1-\lambda_2)). \tag{2.9}$$

Provided (2.1) and (2.3) hold at these points, the simplex method will take the inside contraction step assumed above.

Note that the x coordinates of $v^{(n)}$ and $v^{(n+1)}$ are positive and the x coordinate of $r^{(n)}$ is negative.

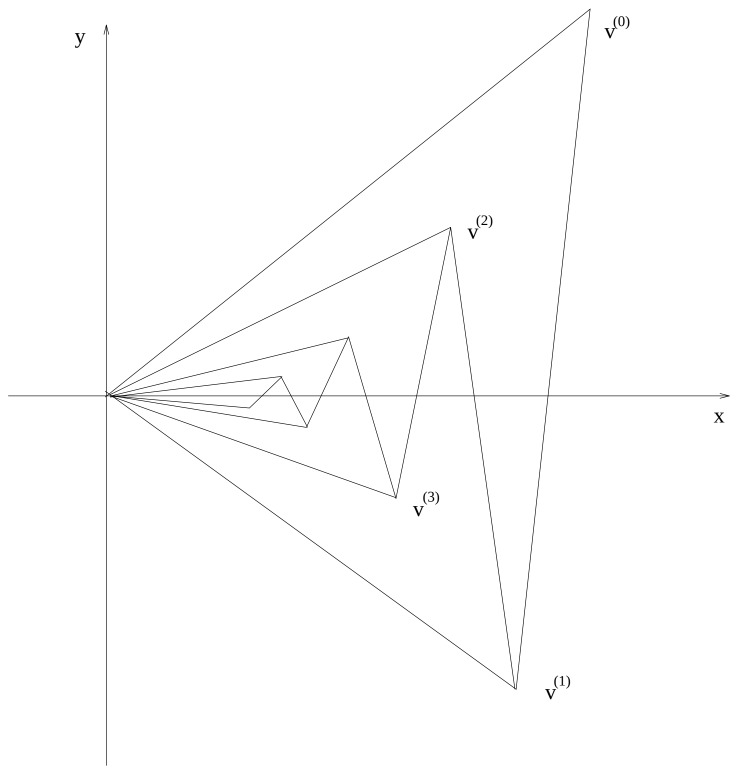

FIG. 2.1. *Successive simplices with repeated focussed inside contractions.*

3. Functions which cause RFIC. Consider the function $f(x,y)$ given by

$$\begin{aligned} f(x,y) &= \theta\phi|x|^\tau + y + y^2, \quad x \le 0, \\ &= \theta x^\tau + y + y^2, \quad x \ge 0, \end{aligned}$$

where θ and ϕ are positive constants. Note that (0,-1) is a descent direction from the origin (0,0) and that f is strictly convex provided $\tau > 1$. f has continuous first derivatives if $\tau > 1$, continuous second derivatives if $\tau > 2$ and continuous third derivatives if $\tau > 3$. Figure 2.2 shows the contour plot of this function and the first two steps of the Nelder-Mead method for the case $\tau = 2$, $\theta = 6$ and $\phi = 60$. Both steps are inside contractions.

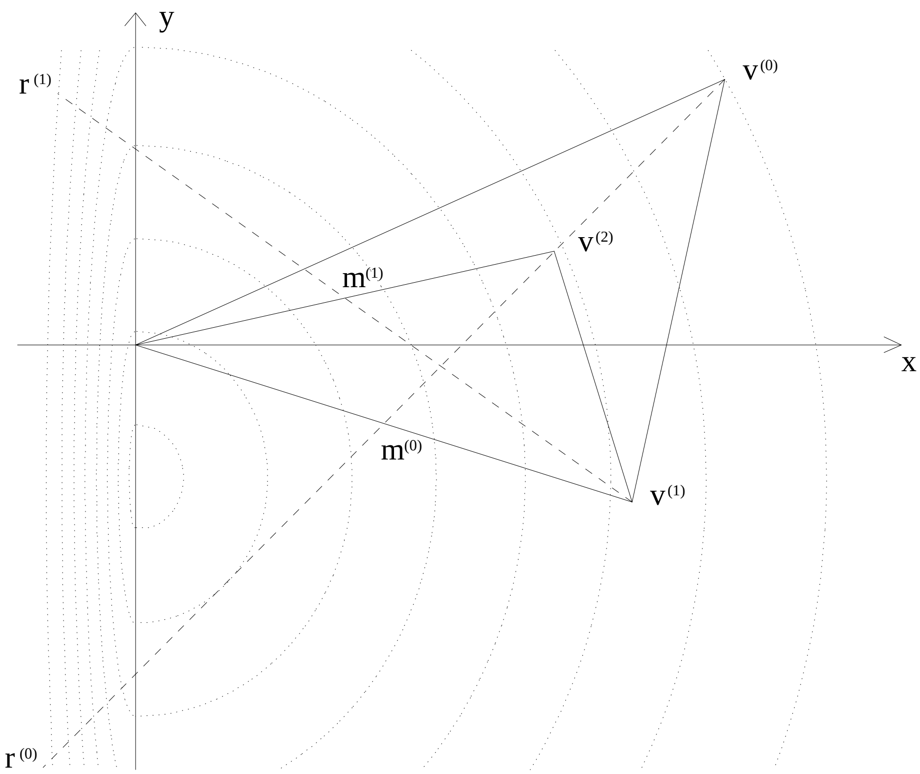

FIG. 2.2. *$f(x, y) = 360x^2 + y + y^2$ if $x \leq 0$ and $f(x, y) = 6x^2 + y + y^2$ if $x \geq 0$, i.e. function in section 3 for case $\tau = 2, \theta = 6, \phi = 60$.*

Define $\hat{\tau}$ to be such that

$$\lambda_1^{\hat{\tau}} = |\lambda_2|, \tag{3.1}$$

so $\hat{\tau}$ is given by

$$\hat{\tau} = \frac{\ln|\lambda_2|}{\ln \lambda_1} \cong 3.0605. \tag{3.2}$$

In what follows assume that τ satisfies

$$0 < \tau < \hat{\tau}. \tag{3.3}$$

Since $0 < \lambda_1 < 1$, it therefore follows that

$$\lambda_1^{\tau} > \lambda_1^{\hat{\tau}} = |\lambda_2|. \tag{3.4}$$

Using (2.7) and (2.9) it follows that

$$\begin{aligned} f(v^{(n)}) &= \theta\lambda_1^{\tau n} + \lambda_2^n + \lambda_2^{2n} \\ \text{and} \quad f(r^{(n)}) &= \phi\theta(\lambda_1^{\tau n}(1-\lambda_1)^{\tau}) - \lambda_2^n(1-\lambda_2) + \lambda_2^{2n}(1-\lambda_2)^2. \end{aligned}$$

Hence $f(v^{(n)}) > f(v^{(n+1)})$ iff

$$\theta\lambda_1^{\tau n}(1-\lambda_1^{\tau}) > \lambda_2^n(\lambda_2 - 1) + \lambda_2^{2n}(\lambda_2^2 - 1).$$

Since $\lambda_1^\tau > |\lambda_2|$ and $\lambda_2^2 - 1 < 0$, this is true for all $n \geq 0$ if θ is such that

$$\theta(1-\lambda_1^\tau) > |\lambda_2 - 1|. \tag{3.5}$$

Also $f(v^{(n+1)}) > f(0)$ iff

$$\theta\lambda_1^{\tau(n+1)} + \lambda_2^{n+1} + \lambda_2^{2(n+1)} > 0.$$

Since, $\lambda_1^\tau > |\lambda_2|$ this is true for all $n \geq 0$ if

$$\theta > 1. \tag{3.6}$$

Also $f(r^{(n)}) > f(v^{(n)})$ iff

$$\begin{aligned} &\phi\theta(\lambda_1^{\tau n}(1-\lambda_1)^\tau) - \lambda_2^n(1-\lambda_2) + \lambda_2^{2n}(1-\lambda_2)^2 > \theta\lambda_1^{\tau n} + \lambda_2^n + \lambda_2^{2n}, \\ \iff \quad &\theta\lambda_1^{\tau n}(\phi(1-\lambda_1)^\tau - 1) > \lambda_2^n(2-\lambda_2) - \lambda_2^{2n}((1-\lambda_2)^2 - 1). \end{aligned}$$

Since $\lambda_2 < 0$ and $\lambda_1^\tau > |\lambda_2|$, this is true for all $n \geq 0$ if θ and ϕ are such that

$$\theta(\phi(1-\lambda_1)^\tau - 1) > (2-\lambda_2). \tag{3.7}$$

For any τ in the range given by (3.3), θ can be chosen so that (3.5) and (3.6) hold and then ϕ can be chosen so that (3.7) holds. It then follows that (2.1) and (2.3) will hold, so the inside contraction step will be taken at every iteration and the simplices will be as derived in section 2. The method will therefore converge to the origin, which is not a stationary point. Examples of values of θ and ϕ which make (3.5), (3.6) and (3.7) hold are: for $\tau = 1$, $\theta = 15$ and ϕ=10; for $\tau = 2$, $\theta = 6$ and $\phi = 60$; for $\tau = 3$, $\theta = 6$ and $\phi = 400$.

4. Necessary conditions for RFIC to occur. In this section we will derive necessary conditions for RFIC to occur. For notational convenience the results are given for RFIC with the origin as focus, but by change of origin they can be applied to any point.

It follows from the description of the algorithm that a necessary condition for RFIC to occur is

$$f_0 = f(0) \leq f(v^{(n+1)}) \leq f(v^{(n)}) \leq f(r^{(n)}). \tag{4.1}$$

(The examples in section 3 satisfy the strict form of the (4.1) relations as given in (2.1) and (2.3).)

If f is s times differentiable at the origin then f can be written in the form $f(v) = p_s(v) + o(\|v\|^s)$, where p_s is a polynomial of degree at most s, and $D^i f(0) = D^i p_s(0)$ for $i = 0, ..., s$, i.e. the derivatives of f and p_s agree. Making a change of variable to z-space using $v = A_1 z_1 + A_2 z_2$, f and p_s can be viewed as functions of $(z_1, z_2) \in \mathbb{R}^2$. When the necessary derivatives exist, define

$$f_0 = f(0), \quad g_i = \frac{\partial f}{\partial z_i}(0), \quad h = \frac{1}{2}\frac{\partial^2 f}{\partial z_1^2}(0) \text{ and } k = \frac{1}{6}\frac{\partial^3 f}{\partial z_1^3}(0).$$

Then (g_1, g_2) is the gradient of f in z-space, and g_i, h and k are the z_i, z_1^2 and z_1^3 coefficients in the Taylor expansion of f in z-space. Since $|\lambda_2| < \lambda_1$ and (2.4) holds, $\|v^{(n)}\| = o(\lambda_1^n)$, so

$$f(v^{(n)}) = p_s(v^{(n)}) + o(|\lambda_1|^{sn}). \tag{4.2}$$

THEOREM 1. *If the origin is the focus of repeated inside contraction starting from a simplex with limiting direction* A_1*, then*
(a) if f *is differentiable at the origin, then* $g_1 = 0$*,*
(b) if f *is 2 times differentiable at the origin, then* $h = 0$*,*
(c) if f *is 3 times differentiable at the origin, then* $k = 0$*.*

Proof. (a) From (4.1) it follows that a necessary condition for RFIC to occur is that $f_0 \le f(v^{(n)})$ and $f_0 \le f(r^{(n)})$. This is true iff

$$\begin{aligned} f_0 &\le f_0 + g_1\lambda_1^n + g_2\lambda_2^n + o(\lambda_1^n), \\ \text{and} \quad f_0 &\le f_0 - g_1\lambda_1^n(1-\lambda_1) - g_2\lambda_2^n(1-\lambda_2) + o(\lambda_1^n). \end{aligned}$$

Since $|\lambda_2| < \lambda_1 < 1$, this cannot occur for all n unless $g_1 = 0$.

(b) Since f is 2 times differentiable at the origin, part (a) holds, so $g_1 = 0$. Hence $p_2(v^{(n)}) - (f_0 + g_2\lambda_2^n + h\lambda_1^{2n}) = O(|\lambda_1\lambda_2|^n) = o(\lambda_1^{4n})$, since $|\lambda_2| < \lambda_1^3$. From this and (4.2) it follows that

$$f(v^{(n)}) = f_0 + g_2\lambda_2^n + h\lambda_1^{2n} + o(\lambda_1^{2n})$$

From (4.1) it follows that a necessary condition for RFIC to occur is that $f_0 \le f(v^{(n)})$ and $f(v^{(n)}) \le f(r^{(n)})$. This is true iff

$$\begin{aligned} f_0 &\le f_0 + g_2\lambda_2^n + h\lambda_1^{2n} + o(\lambda_1^{2n}) \\ \text{and} \quad 0 &\le -g_2\lambda_2^n(2-\lambda_2) - h\lambda_1^{2n+1}(2-\lambda_1) + o(\lambda_1^{2n}). \end{aligned}$$

Since $|\lambda_2| < \lambda_1^2 < 1$, this cannot occur for all n unless $h = 0$.

(c) Since f is 3 times differentiable at the origin, parts (a) and (b) hold, so $g_1 = 0$ and $h = 0$. Hence $p_3(v^{(n)}) - (f_0 + g_2\lambda_2^n + k\lambda_1^{3n}) = O(|\lambda_1\lambda_2|^n) = o(\lambda_1^{4n})$. From this and (4.2) it follows that

$$f(v^{(n)}) = f_0 + g_2\lambda_2^n + k\lambda_1^{3n} + o(\lambda_1^{3n})$$

From (4.1) it follows that a necessary condition for RFIC to occur is that $f_0 \le f(v^{(n)})$ and $f_0 \le f(r^{(n)})$. This is true iff

$$\begin{aligned} f_0 &\le f_0 + g_2\lambda_2^n + k\lambda_1^{3n} + o(\lambda_1^{3n}) \\ \text{and} \quad f_0 &\le f_0 - g_2\lambda_2^n(1-\lambda_2) - k\lambda_1^{3n}(1-\lambda_1)^3 + o(\lambda_1^{3n}). \end{aligned}$$

Since $\lambda_1^{3n} > |\lambda_2|$, this cannot occur for all n unless $k = 0$. □

THEOREM 2. *If* f *has a non-zero gradient at the origin and in a neighbourhood of the origin can be expressed in the form*

$$f(v) = p_4(v) + o(\|v\|^{\hat{\tau}}), \tag{4.3}$$

where p_4 *is at least 4 times differentiable at the origin, and if the initial simplex is not degenerate, then the origin cannot be the focus of repeated inside contractions.*

Proof. Assume the origin is the focus of repeated contractions.

The first three derivatives of f and p_4 at the origin are the same. Theorem 1 shows that $g_1 = h = k = 0$. Hence $p_4(v^{(n)}) - (f_0 + g_2\lambda_2^n) = O(|\lambda_1\lambda_2|^n) = o(\lambda^{4n})$. Since $\hat{\tau} < 4$ and $o(\|v^{(n)}\|^{\hat{\tau}}) = o(\lambda_1^{\hat{\tau}n})$ and $\lambda_1^{\hat{\tau}} = |\lambda_2|$ (by the definition of $\hat{\tau}$), it follows that

$$f(v^{(n)}) = f_0 + g_2\lambda_2^n + o(|\lambda_2|^n).$$

From (4.1) it follows that a necessary condition for RFIC to occur is that $f_0 \leq f(v^{(n)})$ and $f_0 \leq f(v^{(n+1)})$. Since $\lambda_2 < 0$, this cannot occur for all n unless $g_2 = 0$. However since a condition of the theorem is that the gradient is non-zero at the origin and since $g_1 = 0$, it is not possible that $g_2 = 0$. This contradicts the original assumption and so proves that the origin cannot be the focus of repeated contractions. □

Theorem 2 shows that RFIC cannot occur for sufficiently smooth functions, the limit being slightly more that three times differentiable. The examples in section 3 show that if the conditions of theorem 2 do not hold then RFIC is possible.

5. Perturbations of the initial simplex. In this section the behaviour of the examples is analysed for perturbations of the starting simplex. The perturbed position for the vertex at the origin must be on the y axis, otherwise the contracting simplex will eventually lie within a region where all derivatives of the function exist, and theorems 1 and 2 show that a non stationary point cannot be the focus of RFIC in such a region. Also if $\tau > 1$, the gradient exists where $x = 0$ and its direction is parallel to the y axis. If follows from theorem 1 that the only initial simplices which can yield RFIC are those with the dominant eigenvector A_1 perpendicular to the y axis. We therefore consider only perturbations where the vertex at the origin is perturbed to $(0, y_0)$ giving the general form

$$v^{(n)} = \begin{bmatrix} 0 \\ y_0 \end{bmatrix} + \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \lambda_1^n + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \lambda_2^n, \tag{5.1}$$

and when $\tau > 1$ we take $y_1 = 0$. The reflected point is then given by

$$r^{(n)} = \begin{bmatrix} 0 \\ y_0 \end{bmatrix} - \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \lambda_1^n (1 - \lambda_1) - \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \lambda_2^n (1 - \lambda_2). \tag{5.2}$$

We are considering y_0, $x_1 - 1$, y_1, x_2 and $y_2 - 1$ to be close to zero.

Repeating the analysis of section 3 gives $f(v^{(n)}) > f(v^{(n+1)})$ iff

$$\begin{aligned}
&\theta \lambda_1^{\tau n} x_1^\tau \left(\left(1 + \frac{x_2}{x_1} \left(\frac{\lambda_2}{\lambda_1} \right)^n \right)^\tau - \left(1 + \frac{x_2}{x_1} \left(\frac{\lambda_2}{\lambda_1} \right)^{n+1} \right)^\tau \lambda_1^\tau \right) \\
&+\lambda_1^n (1 - \lambda_1) y_1 (1 + 2y_0 + \lambda_1^n (1 + \lambda_1) y_1 + \lambda_2^n (1 + \lambda_2) y_2) \\
> \; &\lambda_2^n (1 - \lambda_2) y_2 (1 + 2y_0 + \lambda_1^n (1 + \lambda_1) y_1) + \lambda_2^{2n} (\lambda_2^2 - 1) y_2^2.
\end{aligned}$$

Also $f(v^{(n)}) > f(0, y_0)$ iff

$$\begin{aligned}
&\theta \lambda_1^{\tau(n+1)} x_1^\tau \left(1 + \frac{x_2}{x_1} \left(\frac{\lambda_2}{\lambda_1} \right)^{n+1} \right)^\tau + y_1 \lambda_1^{n+1} (1 + 2y_0 + y_1 \lambda_1^{n+1} + y_2 \lambda_2^{n+1}) \\
&+ y_2 \lambda_2^{n+1} (1 + 2y_0 + y_1 \lambda_1^{n+1}) + y_2^2 \lambda_2^{n+1} > 0.
\end{aligned} \tag{5.3}$$

Note that for $x_1 - 1$ and x_2 sufficiently close to zero, the x coordinate of $r^{(n)}$ is negative, so the negative x case for the form of f holds. Hence $f(r^{(n)}) > f(v^{(n)})$ iff

$$\begin{aligned}
&\theta \lambda_1^{\tau n} x_1^\tau \left(\phi \left(1 - \lambda_1 - \frac{x_2}{x_1} \left(\frac{\lambda_2}{\lambda_1} \right)^n (1 - \lambda_2) \right)^\tau - \left(1 + \frac{x_2}{x_1} \left(\frac{\lambda_2}{\lambda_1} \right)^n \right)^\tau \right) \\
&- y_1 \lambda_1^n (2 - \lambda_1)(1 + 2y_0 + y_1 \lambda_1^{n+1} + y_2 \lambda_2^{n+1}) \\
> \; &y_2 \lambda_2^n (2 - \lambda_2)(1 + 2y_0 + y_1 \lambda_1^{n+1}) + y_2^2 \lambda_2^n (2 - \lambda_2).
\end{aligned}$$

Since the corresponding inequalities are strict in section 3 and all the functions are continuous, it follows that there exists a symmetric neighbourhood of $y_0 = 0$, $x_1 = 1$, $y_1 = 0$, $x_2 = 0$ and $y_2 = 1$ in which the above three relations hold for $n = 0$. Since $|\lambda_1| < 1$ and $|\lambda_2| < 1$, it follows that if $\tau \leq 1$, the inequalities still hold for all $n \geq 0$. If $\tau > 1$ then the RFIC behaviour will not change in the neighbourhood provided $y_1 = 0$. The set of possible perturbations which maintain the RFIC behaviour is therefore of dimension 4 for $\tau > 1$ and of dimension 5 for $\tau \leq 1$.

Because of this, the behaviour of the examples are stable against small numerical perturbations caused by rounding error when $\tau \leq 1$ and are not stable when $\tau > 1$. This behaviour is confirmed by numerical tests. Rounding error introduces a component of the larger eigenvector in the y direction and this is enough to prevent the algorithm converging to the origin when $\tau > 1$. The example with $\tau = 1$ is not strictly convex, however a strictly convex example which is numerically stable can be constructed by taking the average of examples with $\tau = 1$ and with $\tau = 2$.

6. Conclusions. A family of functions of two variables has been presented which cause the Nelder-Mead method to converge to a non-stationary point. Members of the family are strictly convex with up to 3 continuous derivatives. The examples cause the Nelder-Mead method to perform the inside contraction step repeatedly with the best vertex remaining fixed. It has been shown that this behaviour cannot occur for smoother functions. These examples are the best behaved functions currently known which cause the Nelder-Mead to converge to a non-stationary point. They provide a limit to what can be proved about the convergence of the Nelder-Mead method.

There are 6 values necessary to specify the initial simplex for functions of two variables. It has been shown that for examples in the family which have a discontinuous first derivative, there is a neighbourhood of the initial simplex of dimension 5 in which all the simplices exhibit the same behaviour. These examples appear to be numerically stable. For those examples in the family where the gradient exists the dimension of the neighbourhood is only 4. These examples are numerically unstable and so are unlikely to occur in practice due to rounding errors, even for starting simplices within the neighbourhood. However even in this case the method can spend a very large number of steps close to a non-stationary point before numerical errors perturb the simplices enough to escape.

7. Acknowledgements. The author would like to thank Margaret Wright and Tony Gilbert for help in clarifying the problem and for valuable suggestions.

REFERENCES

[1] J. E. Dennis and V. J. Torczon, *Direct search methods on parallel machines*, SIAM Journal on Optimization, 1 (1991), pp. 448–474.

[2] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, *Convergence behavior of the Nelder-Mead simplex algorithm in low dimensions*, Submitted to SIOPT, (1996).

[3] J. A. Nelder and R. Mead, *A simplex method for function minimization*, Computer Journal, 7 (1965), pp. 308–313.

[4] W. Spendley, G. R. Hext, and F. R. Himsworth, *Sequential applications of simplex designs in optimization and evolutionary operation*, Technometrics, 4 (1962), pp. 441–461.

[5] V. J. Torczon, *Multi-directional search: a direct search algorithm for parallel machines*, Tech. Report TR90-7, Department of Computational and Appled Mathematics, Rice University, 1990. PhD thesis.

[6] ———, *On the convergence of the multi-directional search algorithm*, SIAM Journal on Optimization, 1 (1991), pp. 123–145.

[7] P. Tseng, *Fortified-descent simplicial search method: a general approach*, Submitted to SIOPT, (1995).