A convergence analysis of the Nelder-Mead simplex method

A. Galántai

Óbuda University
John von Neumann Faculty of Informatics
1034 Budapest, Bécsi u. 96/b, Hungary
galantai.aurel@nik.uni-obuda.hu

Abstract: We give a sufficient condition for a certain type of convergence behavior of the Nelder-Mead simplex method and apply this result to several examples. We also give two related examples for the case of repeated shrinking which indicates a kind of local character of the method.

Keywords: Nelder-Mead simplex method, convergence, counterexamples

1 Introduction

The Nelder-Mead simplex method [12] is a direct search algorithm for the unconstrained minimization problem

$$f(x) \to \min \quad (f : \mathbb{R}^n \to \mathbb{R})$$

using a sequence of simplices and function evaluations of their vertices and some related points. Since its publication, the Nelder-Mead simplex method gained high popularity in various application areas and derivative-free optimization [16], [8], [5], [2], [1]. The Nelder-Mead paper [12] has a citation number over 30000 (google scholar, 30-03-2020). In spite of the great number of related papers and variants of the original method quite a few theoretical results are known on the convergence of the Nelder-Mead method (see [4], [7], [6]). A famous two dimensional example by McKinnon [10] shows that Nelder-Mead simplex algorithm may fail to converge to a stationary point of f, even if f is strictly convex and has continuous derivatives. In this paper we also give a few new examples which provide further insights into the convergence properties of the method. In the next section we define the Nelder-Mead simplex method and summarize the main convergence results. In Section 3 we analyze in detail the decision structure of the method in two dimension. In Section 4 we give a sufficient condition for a certain type of convergence behavior of the Nelder-Mead simplex method. We apply this result to several examples in Section 5. In the last section we give two related examples for the case of repeated shrinking which indicate a kind of local character of the method. It has been suspected that the choice of the initial simplex may influence the performance of the simplex method (for experimental results, see e.g. [13], [17]). The results presented in this paper clearly support this assumption as well.

2 The Nelder-Mead simplex method

We use the following generally accepted form of the original method [7]. The vertices of the initial simplex are denoted by $x_1, x_2, \ldots, x_{n+1} \in \mathbb{R}^n$. It is assumed that the vertices $x_1, \ldots, x_{n+1}$ are ordered such that

$$f(x_1) \leq f(x_2) \leq \cdots \leq f(x_{n+1}).$$

Define $x_c = \frac{1}{n}\sum_{i=1}^{n} x_i$. The related evaluation points are

$$x_r = (1+\alpha)x_c - \alpha x_{n+1}, \qquad x_e = (1+\alpha\gamma)x_c - \alpha\gamma x_{n+1},$$

$$x_{oc} = (1+\alpha\beta)x_c - \alpha\beta x_{n+1}, \qquad x_{ic} = (1-\beta)x_c + \beta x_{n+1},$$

where $\alpha = 1$, $\beta = 1/2$, $\gamma = 2$. Then one iteration step of the method is the following.

Operation	Nelder-Mead simplex method
1: Ordering	$f(x_1) \leq \cdots \leq f(x_{n+1})$
Reflect	if $f(x_1) \leq f(x_r) < f(x_n)$, then $x_{n+1} \leftarrow x_r$ and goto 1
Expand	if $f(x_r) < f(x_1)$ and $f(x_e) < f(x_r)$,
	then $x_{n+1} \leftarrow x_e$ and goto 1.
	If $f(x_e) \geq f(x_r)$, then $x_{n+1} \leftarrow x_r$ and goto 1.
Contract outside	If $f(x_n) \leq f(x_r) < f(x_{n+1})$ and $f(x_{oc}) \leq f(x_r)$,
	then $x_{n+1} \leftarrow x_{oc}$ and goto 1.
Contract inside	If $f(x_r) \geq f(x_{n+1})$ and $f(x_{ic}) < f(x_{n+1})$
	then $x_{n+1} \leftarrow x_{ic}$ and goto 1.
Shrink	$x_i \leftarrow (x_i + x_1)/2$, $f(x_i)$ (for all i) and goto 1

There are two rules that apply to reindexing after each iteration. If a nonshrink step occurs, then x_{n+1} is discarded and a new point $v \in \{x_r, x_e, x_{oc}, x_{ic}\}$ is accepted. The following cases are possible:

$$f(v) < f(x_1), \quad f(x_1) \leq f(v) \leq f(x_n), \quad f(v) < f(x_{n+1}).$$

Let

$$j = \begin{cases} 1, & \text{if } f(v) < f(x_1) \\ \max_{2 \leq \ell \leq n+1} \{f(x_{\ell-1}) \leq f(v) \leq f(x_\ell)\}, & \text{otherwise} \end{cases}.$$

Hence

$$x_i^{new} = x_i \ (1 \leq i \leq j-1), \ x_j^{new} = v, \ x_i^{new} = x_{i-1} \ (i = j+1, \ldots, n+1).$$

This type of selection inserts v into the ordering of Step 1 with the highest possible index. If shrinking occurs, then

$$x_1' = x_1, \quad x_i' = (x_i + x_1)/2 \quad (i = 2, \ldots, n+1)$$

plus a reordering takes place. By convention, if $f(x_1') \le f(x_i')$ $(i = 2, \ldots, n)$, then $x_1^{new} = x_1$.

Here we can also write $x(\lambda) = (1+\lambda)x_c - \lambda x_{n+1}$ and so

$$x_r = x(1) = 2x_c - x_{n+1}, \qquad x_e = x(2) = 3x_c - 2x_{n+1},$$

$$x_{oc} = x\left(\tfrac{1}{2}\right) = \tfrac{3}{2}x_c - \tfrac{1}{2}x_{n+1}, \qquad x_{ic} = x\left(-\tfrac{1}{2}\right) = \tfrac{1}{2}x_c + \tfrac{1}{2}x_{n+1}.$$

Lagarias, Poonen and Wright [6] defined a restricted version of the above method, where expansion steps are not allowed.

Kelly [4], [3] developed a sufficient decrease condition for the average of the object function values (evaluated at the vertices) and proves that if this condition is satisfied during the process, then any accumulation point of the simplices is a critical point of f. For other variants of the Nelder-Mead algorithm, see Tseng [15], Nazareth-Tseng [11], Pryce-Coope-Byatt [14].

Lagarias, Reeds, Wright and Wright [7] proved that if the function f is strictly convex on $\mathbb{R}^2$ with bounded level sets and the initial simplex is non-degenerate, the function values at all simplex vertices converge to the same value. They also proved that the simplex diameters are converging and the simplices in the standard Nelder-Mead algorithm have diameters converging to zero ([7] Theorems 5.1, 5.2).

For the restricted version of the Nelder-Mead method, Lagarias, Poonen, Wright [6] showed also in $\mathbb{R}^2$ that for any non-degenerate starting simplex and any twice-continuously differentiable function with everywhere positive definite Hessian and bounded level sets, the algorithm always converges to the minimizer.

In the light of the above convergence results McKinnon's nonconvergence example is particularly interesting (see, e.g. [6], [18]). McKinnon [10] constructed the function

$$f(x,y) = \begin{cases} \theta\phi\,|x|^{\tau} + y + y^2, & \text{if } x \le 0 \\ \theta x^{\tau} + y + y^2, & \text{if } x \ge 0 \end{cases} \tag{1}$$

where ϕ, θ, τ are positive constants. This f is strictly convex and has continuous first derivatives if $\tau > 1$. For this function, the Nelder-Mead simplex algorithm may fail to converge to a stationary point. In particular, with $\phi = 6$ and $\theta = 60$, the counterexample works for $0 \le \tau \le \widehat{\tau}$, and it does not work for $\tau > \widehat{\tau}$, where $\widehat{\tau} \approx 3.0605$.

Wright [18] raises several open questions concerning the Nelder-Mead method such as

- Why is it sometimes so effective (compared to other direct search methods) in obtaining a rapid improvement in f?

- One failure mode is known (McKinnon [10]) – but are there other failure modes?
- Why, despite its apparent simplicity, should the Nelder-Mead method be difficult to analyze mathematically?

Our purpose here is to show other failure modes indicating the complicated convergence structure or behavior of the method.

3 The Nelder-Mead method in two dimensions

Assume that $f : \mathbb{R}^2 \to \mathbb{R}$ is continuous and the i^{th} simplex $S^{(i)} = \left[x_1^{(i)}, x_2^{(i)}, x_3^{(i)}\right]$ is such that $f\left(x_1^{(i)}\right) \leq f\left(x_2^{(i)}\right) \leq f\left(x_3^{(i)}\right)$ for all $i \geq 0$. The related refection, expansion and contraction points are denoted by $x_r^{(i)}$, $x_e^{(i)}$, $x_{oc}^{(i)}$ and $x_{ic}^{(i)}$, respectively. For the given parameters, $x_c^{(0)} = \frac{1}{2}\left(x_1^{(i)} + x_2^{(i)}\right)$,

$$x_r^{(i)} = 2x_c^{(0)} - x_3^{(i)}, \qquad x_e^{(i)} = 3x_c^{(0)} - 2x_3^{(i)},$$

$$x_{oc}^{(i)} = \tfrac{3}{2}x_c^{(0)} - \tfrac{1}{2}x_3^{(i)}, \qquad x_{ic}^{(i)} = \tfrac{1}{2}x_c^{(0)} + \tfrac{1}{2}x_3^{(i)}.$$

Taking all possible cases into account, the $(i+1)^{\text{th}}$ iteration of the Nelder-Mead method can be written as follows:

1. If $f\left(x_1^{(i)}\right) \leq f\left(x_r^{(i)}\right) < f\left(x_2^{(i)}\right)$, then $S^{(i+1)} = \left[x_1^{(i)}, x_r^{(i)}, x_2^{(i)}\right]$.

2a) If $f\left(x_e^{(i)}\right) < f\left(x_r^{(i)}\right) < f\left(x_1^{(i)}\right)$, then $S^{(i+1)} = \left[x_e^{(i)}, x_1^{(i)}, x_2^{(i)}\right]$.

2b) If $f\left(x_r^{(i)}\right) < f\left(x_1^{(i)}\right)$ and $f\left(x_e^{(i)}\right) \geq f\left(x_r^{(i)}\right)$, then $S^{(i+1)} = \left[x_r^{(i)}, x_1^{(i)}, x_2^{(i)}\right]$.

3) If $f\left(x_2^{(i)}\right) \leq f\left(x_r^{(i)}\right) < f\left(x_3^{(i)}\right)$ and $f\left(x_{oc}^{(i)}\right) \leq f\left(x_r^{(i)}\right)$, then three cases are possible:

3a) If $f\left(x_{oc}\right) < f\left(x_1^{(i)}\right)$, then $S^{(i+1)} = \left[x_{oc}^{(i)}, x_1^{(i)}, x_2^{(i)}\right]$.

3b) If $f\left(x_1^{(i)}\right) \leq f\left(x_{oc}^{(i)}\right) < f\left(x_2^{(i)}\right)$, then $S^{(i+1)} = \left[x_1^{(i)}, x_{oc}^{(i)}, x_2^{(i)}\right]$.

3c) If $f\left(x_2^{(i)}\right) \leq f\left(x_{oc}^{(i)}\right) \leq f\left(x_r^{(i)}\right)$, then $S^{(i+1)} = \left[x_1^{(i)}, x_2^{(i)}, x_{oc}^{(i)}\right]$.

4) If $f\left(x_r^{(i)}\right) \geq f\left(x_3^{(i)}\right) > f\left(x_{ic}^{(i)}\right)$, then three cases are possible:

4a) If $f\left(x_{ic}^{(i)}\right) < f\left(x_1^{(i)}\right)$, then $S^{(i+1)} = \left[x_{ic}^{(i)}, x_1^{(i)}, x_2^{(i)}\right]$.

4b) If $f\left(x_1^{(i)}\right) \leq f\left(x_{ic}^{(i)}\right) < f\left(x_2^{(i)}\right)$, then $S^{(i+1)} = \left[x_1^{(i)}, x_{ic}^{(i)}, x_2^{(i)}\right]$.

4c) If $f\left(x_2^{(i)}\right) \leq f\left(x_{ic}^{(i)}\right) < f\left(x_3^{(i)}\right)$, then $S^{(i+1)} = \left[x_1^{(i)}, x_2^{(i)}, x_{ic}^{(i)}\right]$.

5) Shrinking occurs if and only if

$$f\left(x_2^{(i)}\right) \leq f\left(x_r^{(i)}\right) < f\left(x_3^{(i)}\right) \text{ and } f\left(x_{oc}^{(i)}\right) > f\left(x_r^{(i)}\right)$$

or

$$f\left(x_r^{(i)}\right) \geq f\left(x_3^{(i)}\right) \text{ and } f\left(x_{ic}^{(i)}\right) \geq f\left(x_3^{(i)}\right)$$

holds. Then $S^{(i+1)} = \left[x_1^{(i+1)}, x_2^{(i+1)}, x^{(i+1)}\right]$, where

$$\left\{x_1^{(i+1)}, x_2^{(i+1)}, x^{(i+1)}\right\} = \left\{x_1^{(i)}, \frac{1}{2}\left(x_2^{(i)} + x_1^{(i)}\right), \frac{1}{2}\left(x_3^{(i)} + x_1^{(i)}\right)\right\}$$

whose order is determined by the requirement

$$f\left(x_1^{(i+1)}\right) \leq f\left(x_2^{(i+1)}\right) \leq f\left(x^{(i+1)}\right).$$

McKinnon investigated the Nelder-Mead method concerning a repeated case 4b), that is the behavior

$$f\left(x_1^{(i)}\right) \leq f\left(x_{ic}^{(i)}\right) < f\left(x_2^{(i)}\right) < f\left(x_3^{(i)}\right) \leq f\left(x_r^{(i)}\right) \quad (i \geq 0). \tag{2}$$

His construction keeps one vertex ($x_1^{(0)} = (0,0)$) fixed, while vertices $x_2^{(i)}$ and $x_3^{(i)}$ converge to $x_1^{(0)}$.

In the next two sections we give a sufficient condition under which the Nelder-Mead method repeats case 4c) and also show its application to several functions in various situations. Our construction keeps two vertices ($x_1^{(0)}$ and $x_2^{(0)}$) fixed, while the third one converges to the midpoint of the line segment $\overline{x_1^{(0)} x_2^{(0)}}$.

4 A condition for repeated inside contraction

Here we give a sufficient condition under which the inside contraction (case 4c)

$$f\left(x_1^{(i)}\right) \leq f\left(x_2^{(i)}\right) \leq f\left(x_{ic}^{(i)}\right) < f\left(x_3^{(i)}\right) \leq f\left(x_r^{(i)}\right) \tag{3}$$

is repeated.

Assume that $f\left(x_1^{(0)}\right) \leq f\left(x_2^{(0)}\right) < f\left(x_3^{(0)}\right)$. The points $x_r^{(0)}$, $x_{ic}^{(0)}$ are located on the straight line defined by the points $x_c^{(0)}$ and $x_3^{(0)}$, where $x_c^{(0)} = \frac{1}{2}\left(x_1^{(0)} + x_2^{(0)}\right)$. The equation for this line is given by

$$\varphi(t) = (1+t)x_c^{(0)} - tx_3^{(0)}. \tag{4}$$

For $t \in [-1,1]$, we have $x_3^{(0)} = \varphi(-1)$, $x_{ic}^{(0)} = \varphi\left(-\frac{1}{2}\right)$, $x_c^{(0)} = \varphi(0)$, $x_r^{(0)} = \varphi(1)$.

If
$$f\left(x_1^{(0)}\right) \le f\left(x_2^{(0)}\right) \le f\left(x_{ic}^{(0)}\right) < f\left(x_3^{(0)}\right) \le f\left(x_r^{(0)}\right), \tag{5}$$
that is
$$f\left(x_1^{(0)}\right) \le f\left(x_2^{(0)}\right) \le f\left(\varphi\left(-\frac{1}{2}\right)\right) < f\left(\varphi\left(-1\right)\right) \le f\left(\varphi\left(1\right)\right), \tag{6}$$
then $x_{ic}^{(0)} = \varphi\left(-\frac{1}{2}\right)$ is selected in the first iteration of the Nelder-Mead simplex method and
$$x_1^{(1)} = x_1^{(0)}, \quad x_2^{(1)} = x_2^{(0)}, \quad x_3^{(1)} = x_{ic} = \varphi\left(-\frac{1}{2}\right). \tag{7}$$

Since $x_1^{(0)}$ and $x_2^{(0)}$ do not change, $x_c^{(0)}$ also remains and the next $x_r^{(1)}$ and $x_{ic}^{(1)}$ will be on the line segment $x(t)$ ($t \in \left[-\frac{1}{2}, \frac{1}{2}\right]$).

Let $t_k = \frac{1}{2^k}$ and assume that we have performed k consecutive steps so that
$$x_1^{(k)} = x_1^{(0)}, \quad x_2^{(k)} = x_2^{(0)}, \quad x_3^{(k)} = \varphi\left(t_k\right) \tag{8}$$
and
$$f\left(x_1^{(0)}\right) \le f\left(x_2^{(0)}\right) < f\left(x_3^{(k)}\right). \tag{9}$$
Since $x_c^{(k)} = x_c^{(0)}$ and
$$\begin{aligned} x(\lambda) &= (1+\lambda)x_c^{(0)} - \lambda x_3^{(k)} = (1+\lambda)x_c^{(0)} - \lambda\left((1+t_k)x_c^{(0)} - t_k x_3^{(0)}\right) \\ &= (1-\lambda t_k)x_c^{(0)} + \lambda t_k x_3^{(0)}, \end{aligned}$$
we have
$$x_r^{(k)} = x^{(k)}(1) = (1-t_k)x_c^{(0)} + t_k x_3^{(0)} = \varphi\left(-t_k\right)$$
and
$$x_{ic}^{(k)} = x^{(k)}\left(-\frac{1}{2}\right) = \left(1+\frac{t_k}{2}\right)x_c^{(0)} + \frac{t_k}{2}x_3^{(0)} = \varphi\left(\frac{t_k}{2}\right).$$

If
$$f\left(x_2^{(0)}\right) \le f\left(\varphi\left(\frac{t_k}{2}\right)\right) < f\left(\varphi\left(t_k\right)\right) \le f\left(\varphi\left(-t_k\right)\right), \tag{10}$$
then
$$x_1^{(k+1)} = x_1^{(0)}, \quad x_2^{(k+1)} = x_2^{(0)}, \quad x_3^{(k+1)} = \varphi\left(t_{k+1}\right) \quad \left(t_{k+1} = -\frac{1}{2^{k+1}}\right). \tag{11}$$

If the above conditions hold for all k values (and f is a continuous function), that is

(i) $f\left(x_1^{(0)}\right) \le f\left(x_2^{(0)}\right) < f\left(x_3^{(0)}\right)$;

(ii) $f\left(\varphi\left(-t_k\right)\right) \ge f\left(\varphi\left(t_k\right)\right) > f\left(\varphi\left(\frac{t_k}{2}\right)\right) \ge f\left(x_2^{(0)}\right)$ ($t_k = -\frac{1}{2^k}$, $k = 0, 1, \ldots$), then $x_3^{(k)} \to x_c^{(0)}$, $f\left(x_3^{(k)}\right) \to f\left(x_c^{(0)}\right)$, while $x_1^{(k)}$ and $x_2^{(k)}$ ($f\left(x_j^{(k)}\right) = f\left(x_j^{(0)}\right)$, $j = 1,2$) remain fixed. Hence the simplices converge the line segment $\overline{x_1^{(0)} x_2^{(0)}}$, while $\lim_{k\to\infty} x_3^{(k)} = x_c^{(0)}$ is the midpoint of this line segment. Also, the diameters of the simplices do not converge to 0.

A sufficient condition for the requested behavior can be formulated as follows.

Theorem 1. Assume that $S^{(0)} = \left[x_1^{(0)}, x_2^{(0)}, x_3^{(0)}\right]$ is such that

$$f\left(x_1^{(0)}\right) \leq f\left(x_2^{(0)}\right) < f\left(x_3^{(0)}\right).$$

If, in addition, f is such that
(a) $f(\varphi(t))$ is continuous on $[-1,1]$;
(b) $f(\varphi(t)) \geq f(\varphi(-t))$ for $t \in [0,1]$;
(c) $f(x(t))$ is strictly monotone decreasing on $[-1,0]$;
(d) $f(\varphi(t)) > f(\varphi(0)) = f\left(x_c^{(0)}\right) \geq f\left(x_2^{(0)}\right)$ ($t \in [-1,1], t \neq 0$),
then

$$f\left(x_1^{(i)}\right) \leq f\left(x_2^{(i)}\right) \leq f\left(x_{ic}^{(i)}\right) < f\left(x_3^{(i)}\right) \leq f\left(x_r^{(i)}\right) \tag{12}$$

holds for all $i = 0,1,2,\ldots$, $x_3^{(i)} \to x_c^{(0)}$, and $f\left(x_3^{(i)}\right) \to f\left(x_c^{(0)}\right)$.

Proof. Assume that for some $-1 \leq t < 0$, $x_3 = \varphi(t)$ and

$$f\left(x_1^{(0)}\right) \leq f\left(x_2^{(0)}\right) < f(\varphi(t)).$$

Then $x_r = \varphi(-t)$, $x_{ic} = \varphi\left(\frac{t}{2}\right)$, and (b) and (c) imply that

$$f(x_r) = f(\varphi(-t)) \geq f(\varphi(t)) = f(x_3) > f\left(\varphi\left(\frac{t}{2}\right)\right) = f(x_{ic}).$$

Condition (d) implies that $f\left(x_1^{(0)}\right) \leq f\left(x_2^{(0)}\right) < f\left(\varphi\left(\frac{t}{2}\right)\right)$. □

In the next section we apply this sufficient condition to a few functions and show some different types of convergence behavior.

5 Examples of repeated inside contractions

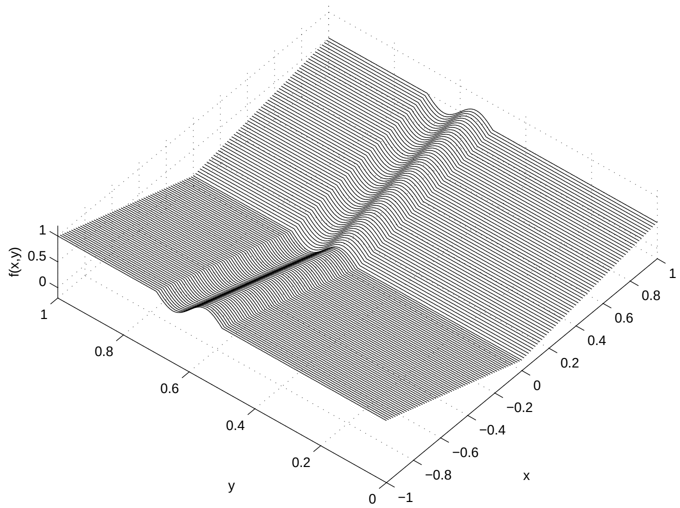
Define the first function as

$$f(x,y) = \frac{1}{4}(x+|x|) + \frac{1}{2}|x-|x|| + g(y), \tag{13}$$

where

$$g(y) = \begin{cases} 0.2\sin(10\pi y - 5\pi), & \text{if } 0.5 \leq y \leq 0.7 \\ 0 & \text{otherwise} \end{cases}. \tag{14}$$

This function is shown on the next figure.



Define the initial simplex vertices as $x_1^{(0)} = (0,0.5)$, $x_2^{(0)} = (0,0.7)$, $x_3^{(0)} = (0.5,0.6)$. Then $f\left(x_1^{(0)}\right) = 0$, $f\left(x_2^{(0)}\right) = 0$, $f\left(x_3^{(0)}\right) = 0.25$, $x_c^{(0)} = [0,0.6]$, $\varphi(t) = [-0.5t, 0.6]$ and

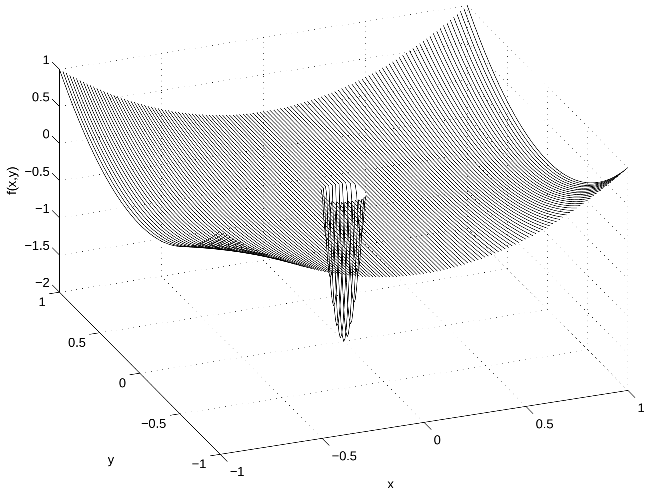
$$f(\varphi(t)) = \begin{cases} \frac{|t|}{4}, & \text{if } t < 0 \\ \frac{t}{2} & \text{if } t > 0 \end{cases}.$$

This function clearly satisfies conditions (a)-(d). Hence we have the repeated x_{ic} behavior and the convergence $x_3^{(i)} \to x_{ic}^{(0)}$. Note that the limit point $x_c^{(0)}$ is not a local minimum point of $f(x,y)$!

Our next example is the function

$$f(x,y) = x^2y^2 + \frac{1}{\varepsilon^2}\left(x^2+y^2\right) - 1 - \left|\frac{1}{\varepsilon^2}\left(x^2+y^2\right) - 1\right| \quad (\varepsilon > 0), \qquad (15)$$

where $\varepsilon < 1$ is small enough. For $\varepsilon = 0.1$, this function is shown on the next figure.



This function has a unique global minimum point at the origin and a continuum number of non-isolated local minimum points along the x and y axes. Select the initial vertices as $x_1^{(0)} = (0,0)$, $x_2^{(0)} = (0,2\varepsilon)$, $x_3^{(0)} = (\varepsilon,\varepsilon)$. Then $f\left(x_1^{(0)}\right) = -2$, $f\left(x_2^{(0)}\right) = 0$, $f\left(x_3^{(0)}\right) = \varepsilon^4$, $x_c^{(0)} = (0,\varepsilon)$, $\varphi(t) = (-t\varepsilon,\varepsilon)$ and $f(x(t)) = \varepsilon^4 t^2$. Hence Theorem 1 applies. Note that $x_1^{(0)}$ is the global minimum point and $x_2^{(0)}$ is a non-isolated local minimum point, while $x_3^{(i)}$ converges to $x_c^{(0)}$ which is not a local minimum point. As ε can be chosen arbitrarily small, this problem can occur arbitrarily close to the global minimum point.

The third example is related to saddle points. Assume that $f(x,y)$ is separable in the form

$$f(x,y) = g(x) - h(y), \tag{16}$$

where g and h are continuous real functions, $g(x) > 0$ for $x \neq 0$, $g(0) = 0$, $g(x)$ is strictly monotone increasing for $x \geq 0$, $g(x)$ is strictly monotone decreasing for $x < 0$, $g(-x) \geq g(x)$ $(x \geq 0)$, $h(y) \geq 0$ for $y \neq 0$, $h(0) = 0$ and $h(-y) \geq h(y)$ for $y \geq 0$. Select the initial vertices as $x_1^{(0)} = (0,-a)$, $x_2^{(0)} = (0,a)$ and $x_3^{(0)} = (b,0)$ with $a,b > 0$. Then $x_c^{(0)} = (0,0)$,

$$f\left(x_1^{(0)}\right) = -h(-a) \leq -h(a) = f\left(x_2^{(0)}\right) \leq 0 = f\left(x_c^{(0)}\right) < g(b) = f\left(x_3^{(0)}\right).$$

Since $\varphi(t) = (-bt,0)$ and $f(\varphi(t)) = g(-bt) \geq g(bt)$ for $t \in [0,1]$, Theorem 1 implies that $x_3^{(i)}$ converges to the saddle point $x_c^{(0)} = (0,0)$. The same result holds, if the above conditions are restricted to an open neighborhood of the origin.

This saddle point phenomenon also appears if $f(x,y)$ is not separable. Using Theorem 1 it is easy to check this for the function

$$f(x,y) = x^2 - y^2 + \frac{1}{\varepsilon^2}\left(x^2 + (y-2\varepsilon)^2\right) - 1 - \left|\frac{1}{\varepsilon^2}\left(x^2 + (y-2\varepsilon)^2\right) - 1\right| \\ + \frac{1}{\varepsilon^2}\left(x^2 + (y+2\varepsilon)^2\right) - 1 - \left|\frac{1}{\varepsilon^2}\left(x^2 + (y+2\varepsilon)^2\right) - 1\right|$$

with $\varepsilon > 0$ small enough and initial vertices $x_1^{(0)} = (0, 2\varepsilon)$, $x_2^{(0)} = (0, -2\varepsilon)$ and $x_3^{(0)} = (1, 0)$. Note that $x_1^{(0)}$ and $x_2^{(0)}$ are are the two local minimum points of f.

6 A note on repeated shrinking

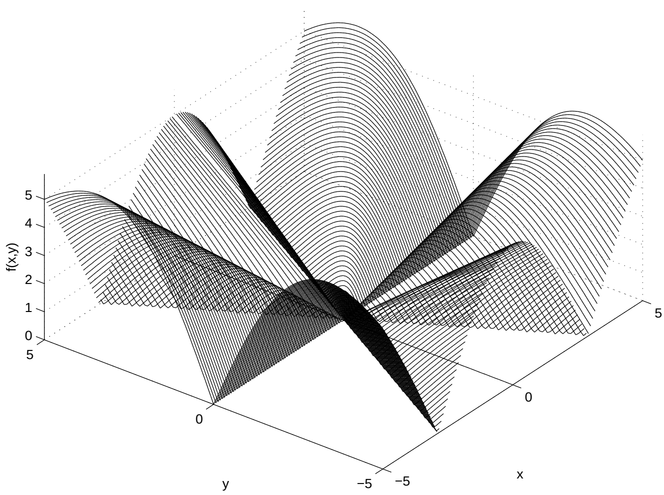
In the case of repeated shrinking the algorithm behaves as a contraction procedure and the vertices of the simplex sequence converge to a common limit point. The following simple examples indicate that this fail safe convergence may have unwanted consequences if we seek for the global minimum point.

Assume that $f\left(x_1^{(i)}\right) \le f\left(x_2^{(i)}\right) \le f\left(x_3^{(i)}\right)$. We use the second condition for shrinking, that is

$$f\left(x_r^{(i)}\right) \ge f\left(x_3^{(i)}\right) \ \wedge f\left(x_{ic}^{(i)}\right) \ge f\left(x_3^{(i)}\right).$$

Consider the function

$$f(x,y) = \begin{cases} \left|\frac{y^3 - 3x^2 y}{x^2 + y^2}\right|, & \text{if } (x,y) \neq (0,0) \\ 0, & \text{if } (x,y) = (0,0). \end{cases} \tag{17}$$



Observe that $0 \le f(x,y) \le 3|y|$, $f(x,0) = 0$, $f\left(x, \sqrt{3}x\right) = f\left(x, -\sqrt{3}x\right) = 0$, and f has a continuum of non-isolated global minimum points.

Let

$$x_1^{(0)} = (0,0), \quad x_2^{(0)} = \left(-1, \sqrt{3}\right), \quad x_3^{(0)} = \left(1, \sqrt{3}\right).$$

Then $f\left(x_1^{(0)}\right) = f\left(x_2^{(0)}\right) = f\left(x_3^{(0)}\right) = 0$, $x_c^{(0)} = \left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$, $x_r^{(0)} = (-2,0)$, $x_{ic}^{(0)} = \left(\frac{1}{4}, \frac{3\sqrt{3}}{4}\right)$. Since $f\left(x_r^{(0)}\right) = f\left(x_3^{(0)}\right) = 0$, and $f\left(x_{ic}^{(0)}\right) = \frac{9\sqrt{3}}{14} > f\left(x_3^{(0)}\right) = 0$, we perform a shrink operation:

$$x_1^{(1)} = (0,0), \quad x_2^{(1)} = \left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right), \quad x_3^{(1)} = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$$

where $f\left(x_j^{(1)}\right) = 0$ $(j = 1,2,3)$ and the previous argument can be applied again. After the i-th iteration we have

$$x_1^{(i)} = (0,0), \quad x_2^{(i)} = \frac{1}{2^i}\left(-1, \sqrt{3}\right), \quad x_3^{(i)} = \frac{1}{2^i}\left(1, \sqrt{3}\right),$$

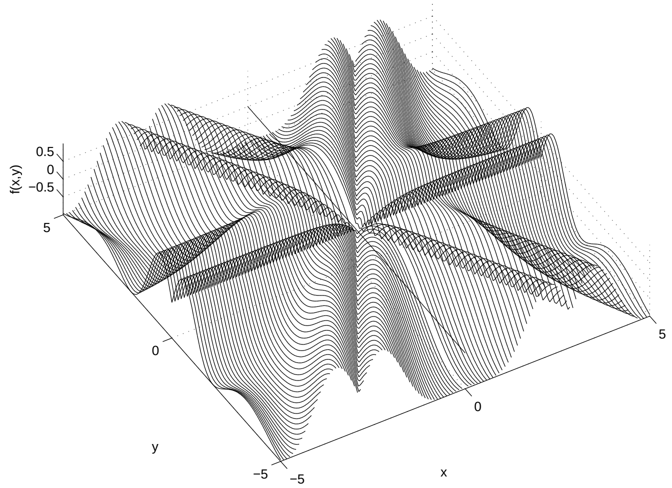
where $f\left(x_j^{(i)}\right) = 0$ $(j = 1,2,3)$. Then

$$x_c^{(i)} = \frac{1}{2^{i+1}}\left(-1, \sqrt{3}\right), \quad x_r^{(i)} = \left(-\frac{1}{2^{i-1}}, 0\right), \quad x_{ic}^{(i)} = \left(\frac{1}{2^{i+2}}, \frac{3\sqrt{3}}{2^{i+2}}\right),$$

$f\left(x_r^{(i)}\right) = 0$ and $f\left(x_{ic}^{(i)}\right) = \frac{9\sqrt{3}}{14}2^{-i} > f\left(x_3^{(i)}\right) = 0$. Hence we have a repeated shrinking and the convergence $x_2^{(i)}, x_3^{(0)} \to x_1^{(0)}$.

Consider the following modification of function (17)

$$f(x,y) = \begin{cases} \sin\left(\left|\frac{y^3-3x^2y}{x^2+y^2}\right|\right), & \text{if } (x,y) \neq (0,0) \\ 0, & \text{if } (x,y) = (0,0). \end{cases}$$



The modified function has non-isolated local minimum points and several global minimum points. For this function, $f(x,0)=0$, $f\left(x,\sqrt{3}x\right)=f\left(x,-\sqrt{3}x\right)=0$ also hold. Take again the initial simplex

$$x_1^{(0)}=(0,0), \quad x_2^{(0)}=\left(-1,\sqrt{3}\right), \quad x_3^{(0)}=\left(1,\sqrt{3}\right).$$

Then $f\left(x_1^{(0)}\right)=f\left(x_2^{(0)}\right)=f\left(x_3^{(0)}\right)=0$, $x_r^{(0)}=(-2,0)$, $x_{ic}^{(0)}=\left(\frac{1}{4},\frac{3\sqrt{3}}{4}\right)$, $f\left(x_r^{(0)}\right)=0$, and $f\left(x_{ic}^{(0)}\right)=\sin\frac{9\sqrt{3}}{14}>f\left(x_3^{(0)}\right)=0$. So we must perform a shrinking and we obtain a repeated shrinking with the same sequences $x_2^{(i)}$, $x_3^{(i)}$, $x_c^{(i)}$, $x_r^{(i)}$, $x_{ic}^{(i)}$ with one difference: $f\left(x_{ic}^{(i)}\right)=\sin\left(\frac{9\sqrt{3}}{14}2^{-i}\right)$.

The obtained result is essentially the result of the previous example. However $x_2^{(i)},x_3^{(i)}$ converge to the local non-isolated minimizer x_1, while the global minimum of f is not achieved.

We note that the Nelder-Mead simplex algorithm is also used in the context of global optimization (see, e.g. [9]). The last two examples indicate a kind of local character of the method.

References

[1] Audet, C., Hare, W: Derivative-free and Blackbox Optimization, Springer, 2017

[2] Conn, A.R., Scheinberg, K., Vicente, L.N.: Introduction to Derivative-Free Optimizations, SIAM, 2009

[3] Kelley, C.T.: Iterative Methods for Optimization, SIAM, 1999

[4] Kelley, CT.: Detection and remediation of stagnation in the Nelder-Mead algorithm using an sufficient decrease condition, SIAM Journal of Optimization, 10 (1999) 33–45

[5] Kolda, T.G., Lewis, R.M., Torczon, V.: Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods, SIAM Review 45, 3 (2003) 385–482

[6] Lagarias, J.C., Poonen, B., Wright, M.H.: Convergence of the restricted Nelder-Mead algorithm in two dimensions, SIAM J. Optim., Vol. 22, No. 2, pp. 501-532, 2012

[7] Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E.: Convergence properties of the Nelder-Mead simplex method in low dimensions, SIAM J. Optim., Vol. 9, No. 1, pp. 112-147, 1998

[8] Lewis, R.M., Torczon, V., Trosset, M.W.: Direct search methods: then and now, Journal of Computational and Applied Mathematics 124 (2000) 191–207

[9] Matlab Global Optimization Toolbox User's Guide, Version R2016a, MathWorks, 2016

[10] McKinnon, K.I.M.: Convergence of the Nelder-Mead simplex method to a nonstationary point, SIAM j. Optim. 9, 1 (1998) 148-158

[11] Nazareth, L., Tseng, P.: Gilding the Lily: A variant for the Nelder-Mead algorithm based on golden-section search, Computational Optimization and Applications, 22 (2002) 133–144

[12] Nelder, J. A., Mead, R.: A simplex method for function minimization, Computer Journal, 7 (1965) 308–313, doi:10.1093/comjnl/7.4.308

[13] Parkinson, J.M., Hutchinson, D.: An investiagtion into the efficiency of variants on the simplex method, in: F.A. Lootsma (ed): Numerical Method for Non-linear Optimization, Academic Press, 1972, pp. 115–135

[14] Price, C.J., Coope, I.D., Byatt, D.: A convergent variant of the Nelder-Mead algorithm, Journal of Optimization Theory and Applications, 113, 1 (2002), 5–19

[15] Tseng, P.: Fortified-descent simplicial search method: A general approach, SIAM J. Optim. 10, 1 (1999) 269–288

[16] Walters, F.H., Morgan, S.L., Parker, L.R., Deming, S.N.: Sequential Simplex Optimization, CRC Press LLC, 1991

[17] Wessing, S.: Proper initialization is crucial for the Nelder-Mead simplex search, Optimization Letters volume 13 (2019) 847–856

[18] Wright, M.H.: Nelder, Mead, and the other simplex method, Documenta Mathematica, Exra Volume: Optimization Stories (2012) 271-276