Convergence of the Nelder-Mead method for convex functions

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 extend several result of Lagarias, Reeds, Wright and Wright [17] for various types of convex functions. The main main convergence theorem of [17] is also generalized to convex functions. We give counterexamples to indicate the tightness of the results.

Keywords: Nelder-Mead simplex method, convergence, convex functions, counterexamples

1 Introduction

The Nelder–Mead (NM) simplex method [22] is a direct search method for the solution of the minimization problem

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

where f is continuous. It is widely used in derivative-free optimization ([30], [14], [5], [1], [15], [26], [18], [12]) and in various application areas (see, e.g., [30], [27]). The Nelder–Mead method can be found in many software libraries as well, such as IMSL, NAG, Matlab, Scilab, Python SciPy and R ([21]). Although there are plenty of numerical testing for the NM method (see, e.g., [28], [20], [26]), only a few theoretical results are known on the convergence (see, e.g., [17], [16], [32], [26], [18]).

Kelley [13], [14] gave a sufficient-decrease condition for the average of the objective function values (evaluated at the simplex vertices) and proved that if this condition is satisfied during the process, then any accumulation point of the simplices is a critical point of f. Han and Neumann [11] investigated the effect of dimensionality on the function $f(x) = x^T x$ $(x \in \mathbb{R}^n)$ (for another approach to dimensionality effects, see Gao and Han [10]). If $f : \mathbb{R}^2 \to \mathbb{R}$ is a strictly convex functions with bounded level sets, then Lagarias, Reeds, Wright and Wright [17] proved that the function values at all simplex vertices converge to the same value. McKinnon [19] gave a strictly convex function $f : \mathbb{R}^2 \to \mathbb{R}$ with continuous derivatives on which the Nelder–Mead algorithm converges to a nonstationary point of f. For the restricted

Nelder–Mead algorithm, where no expansion step is allowed, Lagarias, Poonen and Wright [16] proved that if $f : \mathbb{R}^2 \to \mathbb{R}$ is a twice-continuously differentiable function with bounded level sets and everywhere positive definite Hessian, then it converges to the unique minimizer of f.

The results of [17] and [16] are clearly important steps toward the theoretical foundation/analysis of the Nelder-Mead method. Nesterov et al. [23] considers [17] and [16] as the first non-empirical justification of the NM method for low-dimensional problems. It is a fact that Google Scholar lists 9118 citations of paper [17] and 73 citations of paper [16] as of 9-8-2023.

If the objective function f does not satisfy the strict convexity conditions of Lagarias et al. [17], [16] or $n \neq 2$, then the Nelder–Mead algorithm may have different types of convergence behavior ([7], [8]). The convergence of the Nelder-Mead simplex sequence to a common limit point is studied in [6], [8] and [9] up to 8 dimensions.

Here we extend several results of Lagarias, Reeds, Wright and Wright [17] to convex and quasiconvex functions. The main convergence result of [17] is also generalized to convex functions. The proof exploits many key elements of [17], and it is simpler than that of [17]. We also provide some examples to show the tightness of the new result.

2 The Nelder-Mead simplex method

We first present the Nelder-Mead method according to Lagarias, Reeds, Wright and Wright [17], and derive a matrix form to be used in later analysis.

The initial simplex $S^{(0)}$ is assumed to be nondegenerate and its vertices are denoted by $x_1^{(0)}, x_2^{(0)}, \ldots, x_{n+1}^{(0)} \in \mathbb{R}^n$. We assume that vertices $x_1^{(0)}, x_2^{(0)}, \ldots, x_{n+1}^{(0)}$ are ordered such that

$$f\left(x_1^{(0)}\right) \leq f\left(x_2^{(0)}\right) \leq \cdots \leq f\left(x_{n+1}^{(0)}\right) \tag{1}$$

and this condition is maintained during the iterations of the Nelder-Mead algorithm. The simplex of iteration k is denoted by $S^{(k)} = \left[x_1^{(k)}, x_2^{(k)}, \ldots, x_{n+1}^{(k)}\right] \in \mathbb{R}^{n\times(n+1)}$. Define $x_c^{(k)} = \frac{1}{n}\sum_{i=1}^{n} x_i^{(k)}$ and $x^{(k)}(\lambda) = (1+\lambda)\, x_c^{(k)} - \lambda x_{n+1}^{(k)}$. The reflection, expansion and contraction points of simplex $S^{(k)}$ are defined by

$$x_r^{(k)} = x^{(k)}(1), \quad x_e^{(k)} = x^{(k)}(2), \quad x_{oc}^{(k)} = x^{(k)}\left(\frac{1}{2}\right), \quad x_{ic}^{(k)} = x^{(k)}\left(-\frac{1}{2}\right),$$

respectively. The function values at the vertices $x_j^{(k)}$ and the points $x_r^{(k)}$, $x_e^{(k)}$, $x_{oc}^{(k)}$ and $x_{ic}^{(k)}$ are denoted by $f\left(x_j^{(k)}\right) = f_j^{(k)}$ $(j = 1, \ldots, n+1)$, $f_r^{(k)} = f\left(x_r^{(k)}\right)$, $f_e^{(k)} = f\left(x_e^{(k)}\right)$, $f_{oc}^{(k)} = f\left(x_{oc}^{(k)}\right)$ and $f_{ic}^{(k)} = f\left(x_{ic}^{(k)}\right)$, respectively.

The Nelder-Mead simplex method is defined as follows.

for $k = 0, 1, 2, \dots$

1. **Order:** Order the vertices of $S^{(k)} = \left[x_1^{(k)}, x_2^{(k)}, \dots, x_{n+1}^{(k)}\right]$ so that
$$f_1^{(k)} \leq f_2^{(k)} \leq \cdots \leq f_{n+1}^{(k)}.$$
2. **Reflect:** If $f_1^{(k)} \leq f_r^{(k)} < f_n^{(k)}$, then terminate the iteration with $S^{(k+1)} = \left[x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}, x_r^{(k)}\right]$.
3. **Expand:** If $f_r^{(k)} < f_1^{(k)}$ and $f_e^{(k)} < f_r^{(k)}$, then terminate the iteration with $S^{(k+1)} = \left[x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}, x_e^{(k)}\right]$.
If $f_r^{(k)} < f_1^{(k)}$ and $f_r^{(k)} \leq f_e^{(k)}$, then terminate the iteration with $S^{(k+1)} = \left[x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}, x_r^{(k)}\right]$.
4. **Contract outside:** If $f_n^{(k)} \leq f_r^{(k)} < f_{n+1}^{(k)}$ and $f_{oc}^{(k)} \leq f_r^{(k)}$, then terminate the iteration with $S^{(k+1)} = \left[x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}, x_{oc}^{(k)}\right]$.
5. **Contract inside:** If $f_r^{(k)} \geq f_{n+1}^{(k)}$ and $f_{ic}^{(k)} < f_{n+1}^{(k)}$, then terminate the iteration with $S^{(k+1)} = \left[x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}, x_{ic}^{(k)}\right]$.
6. **Shrink:**
Evaluate f at the points $z_i = \frac{1}{2}\left(x_1^{(k)} + x_i^{(k)}\right)$, $i = 1, 2, \dots, n+1$, and terminate the iteration with $S^{(k+1)} = [z_1, z_2, \dots, z_n, z_{n+1}]$

endfor

Note that the shrink operation may occur only if $f_n^{(k)} \leq f_r^{(k)} < f_{n+1}^{(k)}$ and $f_{oc}^{(k)} > f_r^{(k)}$ or $f_r^{(k)} \geq f_{n+1}^{(k)}$ and $f_{ic}^{(k)} \geq f_{n+1}^{(k)}$. Hence the related logical conditions of the operations 2-6 are mutually exclusive.

There are two rules that apply to reindexing after each iteration. If a nonshrink step occurs, then $x_{n+1}^{(k)}$ is replaced by a new point $v \in \left\{x_r^{(k)}, x_e^{(k)}, x_{oc}^{(k)}, x_{ic}^{(k)}\right\}$. The following cases are possible:

$$f(v) < f_1^{(k)}, \quad f_1^{(k)} \leq f(v) < f_n^{(k)}, \quad f_n^{(k)} \leq f(v) < f_{n+1}^{(k)}.$$

If

$$j^{(k)} = \begin{cases} 1, & \text{if } f(v) < f_1^{(k)} \\ \max_{2 \leq \ell \leq n+1} \left\{f_{\ell-1}^{(k)} \leq f(v) < f_\ell^{(k)}\right\}, & \text{otherwise} \end{cases} . \tag{2}$$

then the new simplex vertices are

$$x_i^{(k+1)} = \begin{cases} x_i^{(k)} & \left(1 \leq i \leq j^{(k)} - 1\right), \\ v & \left(i = j^{(k)}\right), \\ x_{i-1}^{(k)} & \left(i = j^{(k)} + 1, \dots, n+1\right). \end{cases} \tag{3}$$

This rule inserts v into the ordering with the highest possible index. If shrinking occurs, then

$$z_1 = x_1^{(k)}, \quad z_i = \left(x_i^{(k)} + x_1^{(k)}\right)/2 \quad (i = 2, \dots, n+1)$$

plus a reordering takes place. If the insertion rule (3) is adopted at iteration k (no shrink step), then

$$f_i^{(k+1)} = \begin{cases} f_i^{(k)} & \left(1 \le i \le j^{(k)} - 1\right), \\ f(v) & \left(i = j^{(k)}\right), \\ f_{i-1}^{(k)} & \left(i = j^{(k)} + 1, \dots, n+1\right). \end{cases} \tag{4}$$

The insertion rule guarantees that $f_i^{(k+1)} \le f_i^{(k)}$ $(i \ne j^{(k)})$, $f_{j^{(k)}}^{(k+1)} < f_{j^{(k)}}^{(k)}$, and

$$f_1^{(k+1)} \le f_2^{(k+1)} \le \dots \le f_{n+1}^{(k+1)}. \tag{5}$$

holds for $k \ge 0$.

Assume that simplex $S^{(k)} = \left[x_1^{(k)}, x_2^{(k)}, \dots, x_{n+1}^{(k)}\right]$ satisfies condition (5) and define the matrix

$$T(\alpha) = \begin{bmatrix} I_n & \frac{1+\alpha}{n} e \\ 0 & -\alpha \end{bmatrix} \in \mathbb{R}^{n \times (+1)} \quad \left(e = [1, 1, \dots, 1]^T\right).$$

Define the permutation matrix

$$P_j = \left[e_1, \dots, e_{j-1}, e_{n+1}, e_j, \dots, e_n\right] \in \mathbb{R}^{(n+1)\times(n+1)} \quad (j = 1, \dots, n+1),$$

where $j = j^{(k)}$. For nonshrinking operations, the new simplex is given by $S^{(k+1)} = S^{(k)} T(\alpha) P_j$. If shrinking occurs, the new simplex is

$$S^{(k+1)} = S^{(k)} T_{shr} P \quad \left(T_{shr} = \frac{1}{2} I_{n+1} + \frac{1}{2} e_1 e^T\right),$$

where the permutation matrix $P \in \mathscr{P}_{n+1}$ is defined by the ordering condition (5) and $\mathscr{P}_{n+1}$ is the set of all possible permutation matrices of order $n+1$. The following cases are possible

Operation	New simplex
1. Reflection	$S^{(k+1)} = S^{(k)} T(1) P_j$ $(j = 2, \dots, n)$
2a) Expansion($v = x_e^{(k)}$)	$S^{(k+1)} = S^{(k)} T(2) P_1$
2b) Expansion($v = x_r^{(k)}$)	$S^{(k+1)} = S^{(k)} T(1) P_1$
3. Outside contraction	$S^{(k+1)} = S^{(k)} T\left(\frac{1}{2}\right) P_j$ $(j = 1, \dots, n+1)$
4. Inside contraction	$S^{(k+1)} = S^{(k)} T\left(-\frac{1}{2}\right) P_j$ $(j = 1, \dots, n+1)$
5. Shrink	$S^{(k+1)} = S^{(k)} T_{shr} P$ $(P \in \mathscr{P}_{n+1})$

Define the set

$$\mathscr{T}=\left\{T(\alpha)P_j:\alpha\in\left\{-\frac{1}{2},\frac{1}{2}\right\},\ j=1,\ldots,n+1\right\}$$
$$\cup\left\{T_{shr}P:P\in\mathscr{P}_{n+1}\right\}\cup\left\{T(1)P_j:j=1,\ldots,n\right\}\cup\left\{T(2)P_1\right\}. \quad (6)$$

It follows that

$$S^{(k)}=S^{(k-1)}T_kP^{(k)}=S^{(0)}B_k \quad (k\geq 1), \quad (7)$$

where

$$B_k=\prod_{i=1}^{k}T_iP^{(i)} \quad \left(T_iP^{(i)}\in\mathscr{T}\right). \quad (8)$$

Hence the convergence of the simplex sequence $\left\{S^{(k)}\right\}$ depends of the convergence of the infinite matrix product $\prod_{i=1}^{\infty}T_iP^{(i)}$ (see [6], [8]). Set $\mathscr{T}$ consists of $3n+3+(n+1)!$ matrices. For the restricted NM method of Lagarias et al. [16], the expansion step 2a), that is $v=x_e^{(k)}$ is prohibited, and $T(2)P_1$ is not an element of $\mathscr{T}$. The matrices of set $\mathscr{T}$ have a common similarity form (10). Define the matrix

$$F=\begin{bmatrix}1 & -e^T\\ 0 & I_n\end{bmatrix} \quad \left(e=[1,1,\ldots,1]^T\in\mathbb{R}^n\right). \quad (9)$$

Lemma 1. *([6], [8]) For all $T_iP^{(i)}\in\mathscr{T}$, matrix $F^{-1}T_iP^{(i)}F$ has the form*

$$F^{-1}T_iP^{(i)}F=\begin{bmatrix}1 & 0\\ b_i & C_i\end{bmatrix}, \quad (10)$$

where $b_i\in\mathbb{R}^n$ and $C_i\in\mathbb{R}^{n\times n}$ depends on $T_iP^{(i)}$.

The matrices $T_sP^{(s)}$ and C_s are numbered as follows:

$T_sP^{(s)}\in\mathscr{T}$	$\leftrightarrow$	C_s
$T(1)P_{j+1}$	$\leftrightarrow$	$C_j \quad (j=1,\ldots,n-1)$
$T(2)P_1$	$\leftrightarrow$	C_n
$T(1)P_1$	$\leftrightarrow$	C_{n+1}
$T\left(\frac{1}{2}\right)P_j$	$\leftrightarrow$	$C_{n+1+j} \quad (j=1,\ldots,n+1)$
$T\left(-\frac{1}{2}\right)P_j$	$\leftrightarrow$	$C_{2n+2+j} \quad (j=1,\ldots,n+1)$
$T_{shr}P\ (P\in\mathscr{P}_{n+1})$	$\leftrightarrow$	$C_{3n+3+j} \quad (j=1,\ldots,(n+1)!)$

where the numbering of permutations $P\in\mathscr{P}_{n+1}$ follows the `perms` function of Matlab (in actual computations).

3 The results of Lagarias, Reeds, Wright and Wright

Here we summarize the most relevant results of [17] with some additions and observations. The insertion rule (2)-(3)-(4) implies the following simple but important results.

Lemma 2. *(Lemma 3.3 of [17]) If function f is bounded from below on $\mathbb{R}^n$ and only a finite number of shrink iterations occur, then each sequence $\left\{f_i^{(k)}\right\}_{k=0}^{\infty}$ converges to some limit f_i^* for $i = 0,1,\ldots,n+1$ and $f_1^* \leq f_2^* \leq \cdots \leq f_{n+1}^*$.*

This result guarantees a kind of convergence under rather weak conditions. At the same time it also shows that the functions values at the vertices are improving step by step.

Lemma 3. *(Lemma 3.4 of [17]) If f is bounded below on $\mathbb{R}^n$, no shrink iterations occur and for some integer ℓ ($1 \leq \ell \leq n$)*

$$f_\ell^* < f_{\ell+1}^*, \tag{11}$$

then there is an index K such that for all $k \geq K$, index $j^{(k)}$ satisfies $j^{(k)} > \ell$, i.e. the first ℓ vertices of all simplices remain fixed after iteration K.

Proof. Let $\delta > 0$ be so that $f_\ell^* + \delta = f_{\ell+1}^*$. Then there exist an index $K > 0$ such that for $k \geq K$,

$$f_\ell^* \leq f_\ell^{(k)} < f_\ell^* + \delta = f_{\ell+1}^* \leq f_{\ell+1}^{(i)} \quad (i \geq 0). \tag{12}$$

If $j^{(k)} \leq \ell$ for any $k \geq K$, then the insertion rule (3) imply that $f_{\ell+1}^{(k+1)} = f_\ell^{(k)}$, which is a contradiction. □

Corollary 1. *(Corollary 3.1 of [17]) If f is bounded below on $\mathbb{R}^n$, no shrink iterations occur and $x_1^{(k)} \neq x_1^{(k+1)}$ infinitely many times, then $f_1^* = f_2^* = \cdots = f_{n+1}^*$.*

This case is shown by the Examples 1 and 2 of [8], where only two types of expansion steps occur resulting in unbounded simplex sequences $\left\{S^{(k)}\right\}$.

Lemma 3 and Corollary 1 show an important characteristic of the convergence behavior of the Nelder-Mead method. Corollary 1 can be rephrased so that if for some $1 \leq t \leq n$, $x_t^{(k)} \neq x_t^{(k+1)}$ holds infinitely many times, then $f_t^* = f_{t+1}^* = \cdots = f_{n+1}^*$. Assume that $f_j^* < f_{j+1}^*$ for some j ($t \leq j \leq n$). Then by Lemma 3 there is an index K such that the first j vertices ($j \geq t$) remain fixed for all $k \geq K$, which is contradiction.

In general the limit values $\{f_i^*\}_{i=1}^{n+1}$ can be different as shown by the following two examples. For other cases, see Examples 3, 4 and 5 (see also [7], [8]).

Example 1. Assume that $n = 3$, $f(x_1,x_2,x_3) = g_1(x_1)\, g_2(x_2)\, h(x_3)$ with $h(x) = 1+x^2$,

$$g_1(x) = \max\left(\frac{15}{16} - \frac{21}{16}\left|x - \frac{1}{3}\right|, \left|x - \frac{23}{32}\right| - \frac{7}{32}\right),$$

$$g_2(x) = \max\left(\frac{2}{3} - \frac{1}{2}\left|x - \frac{1}{3}\right|, \left|x - \frac{7}{12}\right| - \frac{1}{12}\right).$$

If the initial simplex is

$$S^{(0)} = \begin{bmatrix} 1 & 0 & 0 & \frac{1}{3} \\ 0 & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

then

$$S^{(k)} = \begin{bmatrix} 1 & 0 & 0 & \frac{1}{3} \\ 0 & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{2^k} \end{bmatrix}, \quad x_r^{(k)} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ -\frac{1}{2^k} \end{bmatrix}, \quad x_{ic}^{(k)} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{2^{k+1}} \end{bmatrix} \quad (k \geq 0).$$

It is easy to check that $f_1^{(k)} = \frac{1}{32}$, $f_2^{(k)} = \frac{1}{6}$, $f_3^{(k)} = \frac{1}{4}$, $f_4^{(k)} = \frac{5}{8} + \frac{5}{2^{2k+3}} = f_r^{(k)}$ and $f_{ic}^{(k)} = \frac{5}{8} + \frac{5}{2^{2k+5}}$. Hence we obtained the inequality

$$f_1^{(k)} < f_2^{(k)} < f_3^{(k)} < f_{ic}^{(k)} < f_4^{(k)} = f_r^{(k)} \quad (k \geq 0)$$

and

$$f_1^* = \frac{1}{32} < f_2^* = \frac{1}{6} < f_3^* = \frac{1}{4} < f_4^* = \frac{5}{8}.$$

Note that f has a global minimum point at $(1,1,0)$ with $f_{\min} = \frac{1}{48}$. However $x_4^{(k)} \to \left(\frac{1}{3}, \frac{1}{3}, 0\right)$ and $f_4^* = \frac{5}{8}$. f also has an isolated local minimum point at $(0,0,0)$ with $f(0,0,0) = \frac{1}{4}$. Also note that only vertex x_4 is changing.

Example 2. Assume that

$$f(x,y) = \min\left(\max\left(\left|y + \frac{1}{2}\right|, 1\right), \left|y - \frac{3}{2}\right|\right) + x^2$$

and

$$S^{(0)} = \begin{bmatrix} 0 & 0 & -\frac{1}{2} \\ 1 & 0 & \frac{1}{2} \end{bmatrix}.$$

Then

$$S^{(k)} = \begin{bmatrix} 0 & 0 & -\frac{1}{2^{k+1}} \\ 1 & 0 & \frac{1}{2} \end{bmatrix}, \quad x_r^{(k)} = \begin{bmatrix} \frac{1}{2^{k+1}} \\ \frac{1}{2} \end{bmatrix}, \quad x_{ic}^{(k)} = \begin{bmatrix} -\frac{1}{2^{k+2}} \\ \frac{1}{2} \end{bmatrix} \quad (k \geq 0)$$

and

$$f_1^{(k)} = \frac{1}{2} < f_2^{(k)} = 1 < f_{ic}^{(k)} = \frac{1}{2^{2k+4}} + 1 < f_3^{(k)} = \frac{1}{2^{2k+2}} + 1 = f_r^{(k)}.$$

Hence $x_1^{(k)} = x_1^{(0)}$, $x_2^{(k)} = x_2^{(0)}$, $x_3^{(k)} \to \left[0, \frac{1}{2}\right]^T = \frac{1}{2}\left(x_1^{(0)} + x_2^{(0)}\right)$ and $f_3^{(k)} \to 1$. f has a global minimum point at $\left(0, \frac{3}{2}\right)$ with $f_{\min} = 0$ and $f_1^* = \frac{1}{2} < f_2^* = f_3^*$.

In general we can only say that if $f_t^* < f_{t+1}^*$ and t is the maximal index, then the first t vertices will be unchanged for all $k \geq K$ with some $K > 0$.

Definition 1. *Let f be a function defined on a convex set $S \subset \mathbb{R}^n$. (i) The function f is said to be convex on S if for every $x, y \in S$*

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y), \quad \forall \lambda \in [0,1]. \tag{13}$$

(ii) The function f is said to be strictly convex on S if for every $x, y \in S$, $x \neq y$,

$$f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y), \quad \forall \lambda \in (0,1).$$

Lagarias et al. [17] showed that no shrinking occurs if f is strictly convex.

Lemma 4. *(Lemma 3.5 of [17]) Assume that f is strictly convex on $\mathbb{R}^n$. Then no shrink steps may occur.*

Lagarias et al. [17] proved the following two results also for strictly convex functions.

Lemma 5. *(Lemma 3.6 of [17]) If f is strictly convex on $\mathbb{R}^n$ and bounded below, then $f_n^* = f_{n+1}^*$.*

Theorem 1. *(Theorem 5.1 of [17]) Assume that f is a strictly convex function on $\mathbb{R}^2$ with bounded level sets. Assume that $S^{(0)}$ is nondegenerate. Then the Nelder-Mead simplex method converges in the sense that $f_1^* = f_2^* = f_3^* = f^*$.*

Theorem 1 is generally considered as the main result of [17]. We now prove the following supplement to this.

Theorem 2. *Assume that f is strictly convex function on $\mathbb{R}^2$ with bounded level sets and $S^{(0)}$ is nondegenerate. Then for any accumulation point S' of the simplex sequence $\left\{S^{(k)}\right\}_{k=0}^{\infty}$, $S' = [x', x', x']$ and $f_i(x') = f^*$ ($i = 1,2,3$).*

Proof. f has a unique global minimum point $x_{\min}$ with $f_{\min} = f(x_{\min})$. The only local minimizer of f is the global minimizer. The compactness of level sets implies that $\left\{S^{(k)}\right\}_{k=0}^{\infty} \subset \times_{i=1}^{n+1} L\left(f, f\left(x_{n+1}^{(0)}\right)\right)$ is also bounded. The Bolzano-Weierstrass theorem implies that we can select a subsequence $\left\{k_j\right\}_{j=0}^{\infty}$, $k_j \to \infty$, such that $S^{(k_j)} \to S^* = [x_1^*, x_2^*, x_3^*]$ and $f_i^{(k_j)} = f\left(x_i^{(k_j)}\right) \to f(x_i^*) = f^*$ ($i = 1,2,3$). It is also clear that $f^* \geq f_{\min}$. It is not possible that x_1^*, x_2^* and x_3^* are pairwise different, that is $x_i^* \neq x_j^*$ ($i \neq j$, $i, j = 1,2,3$). If they are pairwise different, there are two possible cases. (i) If x_1^*, x_2^* and x_3^* are not collinear, they form a nondegenerate triangle. The strict convexity of f implies that for $y \in (x_1^*, x_3^*)$, $y \in (x_3^*, x_2^*)$ or $y \in (x_2^*, x_1^*)$, $f(y) < f^*$. Hence we have three different local minimum points, which is contradiction. (ii) If x_1^*, x_2^* and x_3^* are collinear, then one point is between the other two. For simplicity, assume that $x_3^* \in (x_2^*, x_1^*)$. Then we have two local minimum points in the intervals (x_1^*, x_3^*) and (x_3^*, x_2^*), which is also contradiction. Hence we cannot have three pairwise different limit points. It is also not possible that we have two different limit points $x_i^*$and x_j^* ($i, j \in \{1,2,3\}$). Assume that we have two different limit points $x^*_{i_1}$, $x^*_{i_2}$ ($x^*_{i_1} \neq x^*_{i_2}$),while the third limit point $x^*_{i_3}$ is identical with one of the two different limit points (i_1, i_2, i_3 is a permutation of the numbers $1,2,3$ so that $i_1 < i_2$). Let $S_j = S\left(x^*_{i_2}, \delta\right)$ $(j = 1,2)$ be an open ball with $\delta < \left\|x^*_{i_1} - x^*_{i_2}\right\|/2$. Assume first that $x^*_{i_3} = x^*_{i_1}$. Let $K > 0$ be large enough such that $x^{(k_j)}_{i_1}, x^{(k_j)}_{i_3} \in S_1$ and $x^{(k_j)}_{i_2} \in S_2$ for all $k_j \geq K$. The strict convexity of f implies that

$$f(y^*) = f\left(\frac{1}{2}\left(x^*_{i_1} + x^*_{i_2}\right)\right) < \frac{1}{2} f\left(x^*_{i_1}\right) + \frac{1}{2} f\left(x^*_{i_2}\right) = f^* \leq f_1^{(k)} \quad (k \geq 0).$$

The continuity of f implies that

$$f\left(y^{(k_j)}\right) = f\left(\frac{1}{2}\left(x^{(k_j)}_{i_1} + x^{(k_j)}_{i_2}\right)\right) \to f\left(\frac{1}{2}\left(x^*_{i_1} + x^*_{i_2}\right)\right) < f^* \quad (k_j \to \infty)$$

and $f\left(z^{(k_j)}\right) = f\left(\frac{1}{2}\left(x^{(k_j)}_{i_2} + x^{(k_j)}_{i_3}\right)\right) \to f(y^*) < f^*$ $(k_j \to \infty)$. Hence for a sufficiently large $K' \geq K$, we have two local minimum points in the open intervals $\left(x^{(k_j)}_{i_1}, x^{(k_j)}_{i_2}\right)$ and $\left(x^{(k_j)}_{i_2}, x^{(k_j)}_{i_3}\right)$, which is contradiction. Assume now that $x^*_{i_3} = x^*_{i_2}$ and $x^{(k_j)}_{i_1} \in S_1$ and $x^{(k_j)}_{i_2}, x^{(k_j)}_{i_3} \in S_2$ for $k_j \geq K$. For $k_j \to \infty$,

$$f\left(w^{(k_j)}\right) = f\left(\frac{1}{2}\left(x^{(k_j)}_{i_1} + x^{(k_j)}_{i_3}\right)\right) \to f\left(\frac{1}{2}\left(x^*_1 + x^*_2\right)\right) < f^*.$$

Thus we have two local minimum points in the open intervals $\left(x^{(k_j)}_{i_1}, x^{(k_j)}_{i_2}\right)$ and $\left(x^{(k_j)}_{i_1}, x^{(k_j)}_{i_3}\right)$, which is contradiction. Thus the accumulation point S^* is of the form $S^* = [x^*_1, x^*_1, x^*_1]$. $\square$

Note that f^* is the smallest value of f achievable by the NM method (see Lemma 2). The strictly convex example of McKinnon [19] indicates that f^* can be different from $f_{\min}$. Hence, in general, we cannot expect that the limit or accumulation point is the minimum point.

If $f^* = f_{\min}$, we have only one accumulation point $[x_{\min}, x_{\min}, x_{\min}]$. If $\left\{x^{(k_j)}_i\right\}$ is such that $x^{(k_j)}_i \to y \neq x_{\min}$ $(i = 1,2,3)$ and $f(y) = f_{\min}$. Then for any $z \in (x_{\min}, y^*)$, $f(z) < f_{\min}$, which is contradiction. So there is only one accumulation point.

The result of Theorem 2 is somewhat reminiscent to that of Kelley [13].

4 Extensions for various types of convex functions

Here we extend several results of the previous section for convex and quasiconvex functions.

Definition 2. *Let f be defined on a convex set $S \subset \mathbb{R}^n$. The function f is said to be quasiconvex on S if*

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \max\{f(x_1), f(x_2)\}$$

for every $x_1, x_2 \in S$ and for every $\lambda \in [0,1]$ or, equivalently

$$f(x_1) \geq f(x_2) \Rightarrow f(x_1) \geq f(x_1 + \lambda(x_2 - x_1))$$

for every $x_1, x_2 \in S$ and for every $\lambda \in [0,1]$.

Definition 3. *Let f be defined on a convex set $S \subset \mathbb{R}^n$. The function f is said to be strictly quasiconvex on S if*

$$f(\lambda x_1 + (1-\lambda)x_2) < \max\{f(x_1), f(x_2)\}$$

for every $x_1, x_2 \in S$ and for every $\lambda \in [0,1]$ or, equivalently

$$f(x_1) \geq f(x_2) \Rightarrow f(x_1) > f(x_1 + \lambda(x_2 - x_1))$$

for every $x_1, x_2 \in S$, $x_1 \neq x_2$, and for every $\lambda \in (0,1)$.

If f is convex, then it is also quasiconvex. If f is strictly convex, then it is strictly quasiconvex. If f is strictly quasiconvex, then it is quasiconvex (see, e.g. Avriel et al. [2] or Cambini et al. [4], Thm. 2.2.1).

The nonshrinking assumption is an essential element in the results of Lagarias et al. [17]. The strict convexity assumption of f excludes shrinking and for such functions Lemma 2 clearly holds.

For quasiconvex (convex) functions, where shrinking may occur, we have the following convergence result.

Lemma 6. *If f is quasiconvex and bounded below on $\mathbb{R}^n$, then for $i = 0, 1, \ldots, n+1$, the sequence $\left\{f_i^{(k)}\right\}_{k=0}^{\infty}$ is monotone decreasing and converges to some limit f_i^*. Furthermore $f_1^* \leq f_2^* \leq \cdots \leq f_{n+1}^*$.*

Proof. If shrinking occurs then $f(z_j) \leq f\left(\frac{1}{2}\left(x_1^{(k)} + x_j^{(k)}\right)\right) \leq f_j^{(k)}$ $(j = 1, \ldots, n+1)$. Since $f(z_i) \leq f_i^{(k)} \leq f_j^{(k)}$, there are at least j new function values $f(z_i)$ that are less than or equal to $f_j^{(k)}$. After ordering the vertices of the shrunken simplex $S^{(k+1)}$, the new function values at the vertices satisfy $f_j^{(k+1)} \leq f_j^{(k)}$ $(j = 1, 2, \ldots, n+1)$. Hence each sequence $\left\{f_i^{(k)}\right\}_{k=0}^{\infty}$ is monotone decreasing and bounded from below. □

For quasiconvex functions, the shrinking can be characterized as follows.

Lemma 7. *(i) If f is strictly quasiconvex on $\mathbb{R}^n$, then no shrink steps may occur. (ii) If f is quasiconvex on $\mathbb{R}^n$, then shrinking may occur if and only if $f_r^{(k)} \geq f_{n+1}^{(k)}$ and $f_{ic}^{(k)} = f_{n+1}^{(k)}$.*

Proof. Shrinking may occur if either $f_n^{(k)} \le f_r^{(k)} < f_{n+1}^{(k)}$ and $f_{oc}^{(k)} > f_r^{(k)}$ or $f_{n+1}^{(k)} \le f_r^{(k)}$ and $f_{ic}^{(k)} \ge f_{n+1}^{(k)}$. By definition $f_c^{(k)} \le f_n^{(k)}$. If f is strictly quasiconvex, then $f_{oc}^{(k)} = f\left(\frac{1}{2}\left(x_r^{(k)} + x_c^{(k)}\right)\right) < \max\left\{f_r^{(k)}, f_c^{(k)}\right\} \le \max\left\{f_r^{(k)}, f_n^{(k)}\right\}$. If $f_n^{(k)} \le f_r^{(k)} < f_{n+1}^{(k)}$, then $f_{oc}^{(k)} < f_r^{(k)}$ and $x_{oc}^{(k)}$ is the incoming vertex. Similarly,

$$f_{ic}^{(k)} = f\left(\frac{1}{2}\left(x_c^{(k)} + x_{n+1}^{(k)}\right)\right) < \max\left\{f_c^{(k)}, f_{n+1}^{(k)}\right\} = f_{n+1}^{(k)}$$

and $x_{ic}^{(k)}$ is the incoming vertex. If f is quasiconvex and $f_n^{(k)} \le f_r^{(k)} < f_{n+1}^{(k)}$, then

$$f_{oc}^{(k)} = f\left(\frac{1}{2}\left(x_r^{(k)} + x_c^{(k)}\right)\right) \le \max\left\{f_r^{(k)}, f_c^{(k)}\right\} \le \max\left\{f_r^{(k)}, f_n^{(k)}\right\} = f_r^{(k)}.$$

If $f_r^{(k)} \ge f_{n+1}^{(k)}$, then $f_{ic}^{(k)} = f\left(\frac{1}{2}\left(x_c^{(k)} + x_{n+1}^{(k)}\right)\right) \le f_{n+1}^{(k)}$. It follows that shrinking occurs if and only if $f_r^{(k)} \ge f_{n+1}^{(k)}$ and $f_{ic}^{(k)} = f_{n+1}^{(k)}$. □

Corollary 2. *If f is convex on $\mathbb{R}^n$ and an index $1 \le j \le n$ exists such that $f_j^{(k)} < f_{n+1}^{(k)}$, then $f_{ic}^{(k)} < f_{n+1}^{(k)}$ and no shrinking occurs in iteration k.*

Proof. Assume that $f_r^{(k)} \ge f_{n+1}^{(k)}$ and $f_j^{(k)} < f_{n+1}^{(k)}$ $(j \le n)$. Then $f_{ic}^{(k)} \le \frac{1}{2n}\sum_{i=1}^{n} f_i^{(k)} + \frac{1}{2} f_{n+1}^{(k)} < f_{n+1}^{(k)}$. □

We can characterize the situation $f_r^{(k)} \ge f_{n+1}^{(k)} = f_{ic}^{(k)}$ as follows.

Lemma 8. *Assume that f is quasiconvex on $\mathbb{R}^n$ and $f_{ic}^{(k)} = f_{n+1}^{(k)}$. If $f_c^{(k)} < f_{n+1}^{(k)}$, then f is constant on the interval $\left[x_{ic}^{(k)}, x_{n+1}^{(k)}\right]$.*

Proof. It is clear that $f_c^{(k)} \le f_n^{(k)}$. By definition $f(z) \le f_{n+1}^{(k)}$ if $z \in \left[x_{ic}^{(k)}, x_{n+1}^{(k)}\right]$. Assume that there exists an element $z \in \left(x_{ic}^{(k)}, x_{n+1}^{(k)}\right)$ such that $f(z) < f_{n+1}^{(k)}$. Then $x_{ic}^{(k)} \in \left[x_c^{(k)}, z\right]$ and $f_{ic}^{(k)} \le \max\left\{f(z), f_c^{(k)}\right\} < f_{n+1}^{(k)}$, which is contradiction. □

Lemma 9. *Assume that f is convex on $\mathbb{R}^n$ and $f_r^{(k)} \ge f_{n+1}^{(k)} = f_{ic}^{(k)}$. Then f is constant on the interval $\left[x_c^{(k)}, x_{n+1}^{(k)}\right]$.*

Proof. $f_c = f_{n+1}^{(k)}$, for if $f_c < f_{n+1}^{(k)}$, then $f_{ic} \le \frac{1}{2}\left(f_c^{(k)} + f_{n+1}^{(k)}\right) < f_{n+1}^{(k)}$, which is a contradiction. Hence $f(z) \le f_{n+1}^{(k)}$ for $z \in \left[x_c^{(k)}, x_{n+1}^{(k)}\right]$. Assume that there is a point $z_1 \in \left(x_c^{(k)}, x_{ic}^{(k)}\right)$ such that $f(z) < f_{n+1}^{(k)}$. Then $x_{ic}^{(k)} \in \left(z_1, x_{n+1}^{(k)}\right)$ and $f_{ic}^{(k)} \le \lambda f(z_1) + (1-\lambda) f_{n+1}^{(k)} < f_{n+1}$ for some $\lambda \in (0,1)$, which is contradiction. If there is a point $z_2 \in \left(x_{ic}^{(k)}, x_{n+1}^{(k)}\right)$ such that $f(z_2) < f_{n+1}^{(k)}$, then $x_{ic}^{(k)} \in \left(x_c^{(k)}, z_2\right)$, and $f_{ic}^{(k)} \leq \mu f_{n+1}^{(k)} + (1-\mu) f(z_2) < f_{n+1}$ for some $\mu \in (0,1)$. □

We prove the following variant of Lemma 3.

Lemma 10. *Assume that f is convex and bounded below on $\mathbb{R}^n$. If for some integer ℓ $(1 \leq \ell \leq n)$*

$$f_\ell^* < f_{\ell+1}^*, \tag{14}$$

then there is an index $K > 0$ such that for all $k \geq K$, $j^{(k)} > \ell$, that is, the first ℓ vertices do not change.

Proof. There is an index such that for all $k \geq K$,

$$f_\ell^* \leq f_\ell^{(k)} < f_\ell^* + \delta = f_{\ell+1}^* \leq f_{\ell+1}^{(i)} \quad (i \geq 0). \tag{15}$$

Since $f_\ell^{(k)} < f_{n+1}^{(k)}$, there is no shrinking (Corollary 2) and the result follows from the insertion rule (3). □

We extend Lemma 5 for convex and strictly quasiconvex functions.

Lemma 11. *If f is convex (strictly quasiconvex) and bounded below on $\mathbb{R}^n$, then $f_n^* = f_{n+1}^*$.*

Proof. It follows from Lemma 6 that $f_n^* \leq f_{n+1}^*$. Assume that $f_n^* < f_{n+1}^*$. If f is convex (strictly quasiconvex), then by Lemma 10 (Lemma 3) there exists an index K such that for all $k \geq K$, $j^{(k)} > n$. Hence the first n vertices do not change and only $x_{ic}^{(k)}$ or $x_{oc}^{(k)}$ is accepted in place of $x_{n+1}^{(k)}$ for $k \geq K$. Hence $x_i^{(k+1)} = x_i^{(K)}$ $(i = 1, 2, \ldots, n)$, $S^{(k+1)} = S^{(k)} T(\alpha)$ $\left(\alpha \in \left\{-\frac{1}{2}, \frac{1}{2}\right\}\right)$, and for $k \geq K$,

$$S^{(k)} = S^{(K)} \prod_{i=K}^{k} T(\alpha_i) = S^{(K)} T\left((-1)^{k-K} \prod_{i=K}^{k} \alpha_i\right) \quad \left(\alpha_i \in \left\{-\frac{1}{2}, \frac{1}{2}\right\}\right). \tag{16}$$

Thus for $k \to \infty$,

$$S^{(k)} \to S^{(K)} T(0) = S^{(K)} \begin{bmatrix} I_n & \frac{1}{n} e \\ 0 & 0 \end{bmatrix} = \left[x_1^{(K)}, \ldots x_n^{(K)}, \frac{1}{n} \sum_{i=1}^{n} x_i^{(K)}\right].$$

Hence $f_n^* = f\left(x_n^{(K)}\right)$, $f_{n+1}^* = f\left(x_C^{(K)}\right) = f\left(\frac{1}{n} \sum_{i=1}^n x_i^{(K)}\right) \leq f\left(x_n^{(K)}\right) = f_n^*$, which is contradiction. In the case of strict quasiconvexity we have $f_{n+1}^* < f\left(x_n^{(K)}\right) = f_n^*$, which is contradiction. □

Remark 1. If f is quasiconvex on $\mathbb{R}^2$ and bounded below, then $f_2^* = f_3^*$ also holds. The proof, however, is more complicated due to the possible presence of shrinking operations (case $f_r^{(k)} \geq f_{n+1}^{(k)} = f_{ic}^{(k)}$). Hence the proof is omitted here.

The convexity assumption on f is indeed essential here. If f is not convex or strictly quasiconvex, then it may happen that $f_n^* < f_{n+1}^*$ (see Example 4 and Lemma 5 of [8]). The following example is related to saddle points (see, e.g. [7] or [8]).

Example 3. Assume that $f(x,y)$ is separable of the form

$$f(x,y) = g(x) - h(y), \tag{17}$$

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 $x < 0$, $g(-x) \geq g(x)$ $(x \geq 0)$, $h(y) > 0$ for $y > 0$, $h(0) = 0$ and $h(-y) \geq h(y)$ for $y \geq 0$. Set 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 $x_i^{(k)} = x_i^{(0)}$ $(i = 1,2)$, $x_3^{(k)} = \left(\frac{b}{2^k}, 0\right)$. Hence $x_3^{(k)}$ converges to the saddle point $x_c^{(0)} = (0,0) = \frac{1}{2}\left(x_1^{(0)} + x_2^{(0)}\right)$, and here $f_2^* = -h(a) < f_3^* = 0$.

5 A generalization of Theorem 1 to convex functions

We now prove the following generalization of Theorem 1, which was the main result of Lagarias, Reeds, Wright and Wright [17].

Theorem 3. *Assume that f is a convex function on $\mathbb{R}^2$ and bounded below. Then the Nelder-Mead simplex method converges in the sense that $f_1^* = f_2^* = f_3^*$.*

Proof. It follows (Lemmas 6 and 11) that $f_1^* \leq f_2^* = f_3^*$. Assume now that $f_1^* < f_2^*$. It follows from Lemma 10 that there exists an index $K > 0$ such that for $k \geq K$, $x_1^{(k)} = x_1^{(K)}$. Hence only $x_2^{(k)}$ and $x_3^{(k)}$ may change. The insertion rule (and the impossibility of shrinking) implies that only the following cases are possible
(i) $f_1^{(k)} \leq f_r^{(k)} < f_2^{(k)}$;
(ii) $f_2^{(k)} \leq f_r^{(k)} < f_3^{(k)}$ and $f_1^{(k)} \leq f_{oc}^{(k)} \leq f_r^{(k)}$;
(iii) $f_3^{(k)} \leq f_r^{(k)}$ and $f_1^{(k)} \leq f_{ic}^{(k)} < f_3^{(k)}$.
In case (i) $x_r^{(k)}$ replaces $x_2^{(k)}$ and $S^{(k+1)} = S^{(k)}T(1)P_2$. For the other two cases we have to assume that $K > 0$ is big enough, so that for $k \geq K$, $f_1^* \leq f_1^{(k)} < f_1^* + \varepsilon$, $f_2^* \leq f_i^{(k)} < f_2^* + \varepsilon$ $(i = 2,3)$, where $\varepsilon > 0$ is such that $f_1^* + 4\varepsilon \leq f_2^*$. In case (ii)

$$\begin{aligned} f_{oc}^{(k)} = f\left(\frac{1}{2}x_c^{(k)} + \frac{1}{2}x_r^{(k)}\right) &\leq \frac{1}{4}f_1^{(k)} + \frac{1}{4}f_2^{(k)} + \frac{1}{2}f_r^{(k)} \\ &\leq \frac{1}{4}\left(f_1^* + \varepsilon\right) + \frac{3}{4}\left(f_2^* + \varepsilon\right) \leq f_2^*. \end{aligned}$$

Hence $x_{oc}^{(k)}$ replaces $x_2^{(k)}$ and $S^{(k+1)} = S^{(k)}T\left(\frac{1}{2}\right)P_2$. In case (iii)

$$\begin{aligned} f_{ic}^{(k)} = f\left(\frac{1}{2}x_c^{(k)} + \frac{1}{2}x_3^{(k)}\right) &\leq \frac{1}{4}f_1^{(k)} + \frac{1}{4}f_2^{(k)} + \frac{1}{2}f_3^{(k)} \\ &\leq \frac{1}{4}\left(f_1^* + \varepsilon\right) + \frac{3}{4}\left(f_2^* + \varepsilon\right) \leq f_2^*. \end{aligned}$$

Hence $x_{ic}^{(k)}$ replaces $x_2^{(k)}$ and $S^{(k+1)} = S^{(k)} T\left(-\frac{1}{2}\right) P_2$. It follows that

$$S^{(k+K)} = S^{(K)} \prod_{j=1}^{k} T(\alpha_j) P_2, \quad k \geq 0,\, \alpha_j \in \left\{1, \frac{1}{2}, -\frac{1}{2}\right\}.$$

Lemma 1 implies that

$$F^{-1} T(1) P_2 F = \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & 1 & 1 \\ 0 & -1 & 0 \end{array}\right] = \begin{bmatrix} 1 & 0 \\ 0 & C_1 \end{bmatrix} \quad \left(C_1 \in \mathbb{R}^{2\times 2}\right),$$

$$F^{-1} T\left(\frac{1}{2}\right) P_2 F = \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & \frac{3}{4} & 1 \\ 0 & -\frac{1}{2} & 0 \end{array}\right] = \begin{bmatrix} 1 & 0 \\ 0 & C_5 \end{bmatrix} \quad \left(C_5 \in \mathbb{R}^{2\times 2}\right),$$

$$F^{-1} T\left(-\frac{1}{2}\right) P_2 F = \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & \frac{1}{4} & 1 \\ 0 & \frac{1}{2} & 0 \end{array}\right] = \begin{bmatrix} 1 & 0 \\ 0 & C_8 \end{bmatrix} \quad \left(C_8 \in \mathbb{R}^{2\times 2}\right).$$

Hence

$$S^{(k+K)} = S^{(K)} F \begin{bmatrix} 1 & 0 \\ 0 & \prod_{j=1}^{k} C_{i_j} \end{bmatrix} F^{-1}.$$

Consider $\mathscr{C}_5 = \left\{\prod_{j=1}^{5} C_{i_j} : i_j \in \{1,5,8\}\right\}$ which is the set of all possible products of C_{i_j}'s of length 5. Here we have $\max\{\rho(G) : G \in \mathscr{C}_5\} = 1$, where $\rho(A)$ denotes the spectral radius of matrix A. The maximum spectral radius appears exactly at one element ($C_1C_1C_1C_1C_1$), while

$$\max\{\rho(G) : G \in \mathscr{C}_5 \backslash \{C_1C_1C_1C_1C_1\}\} = 0.8431.$$

Define the matrix norm

$$\|A\|_s = \left\|S^{-1}AS\right\|_2 \quad \left(S = \begin{bmatrix} 1.5934 & -0.9069 \\ 0 & 1.6413 \end{bmatrix}\right). \tag{18}$$

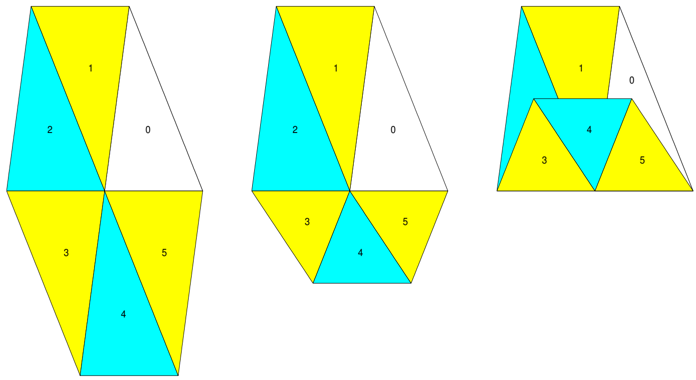
Then $Q = \max\{\|G\|_S : G \in \mathscr{C}_5\} = 1.1217$, and only the 3 elements of

$$\mathscr{F} = \{C_1C_1C_1C_1C_1, C_1C_1C_5C_1C_1, C_1C_1C_8C_1C_1\} \subset \mathscr{C}_5$$

have norm greater than 1. For the rest of the elements,

$$q = \max\{\|G\|_S : G \in \mathscr{C}_5 \backslash \mathscr{F}\} = 0.9707. \tag{19}$$

The three cases of $\mathscr{F}$ coincide with those that are mentioned in the proof of Lemma 5.1 of [17] (p. 141). They represent the following simplex chains of $\mathscr{F}$ in order



In these cases the incoming vertex in step 5, coincides with the worst vertex of the ”starting simplex”, which is impossible by the insertion rule (3). Also note that matrix $T(1)P_2$ is a 6-involutory matrix (see Trench [29]). So we can exclude these three cases from $\mathscr{C}_5$ and define the set

$$\widehat{\mathscr{C}_5} = \mathscr{C}_5 \setminus \mathscr{F}. \tag{20}$$

Assume that $k = 5m + r$ ($m, r \in \mathbb{N},\ 0 \leq r < 5$). Then

$$\prod_{j=1}^{k} C_{i_j} = \left[\prod_{j=1}^{m}\left(C_{i_{5(j-1)+1}} \cdots C_{i_{5j}}\right)\right] C_{i_{5m+1}} \cdots C_{i_{5m+r}}.$$

If all $C_{i_{5(j-1)+1}} \cdots C_{i_{5j}}$'s are from $\widehat{\mathscr{C}_5}$, then

$$\left\|\prod_{j=1}^{k} C_{i_j}\right\|_S \leq q^m Q^r \leq q^m Q^4 = q^{\left\lfloor \frac{k}{5} \right\rfloor} Q^4,$$

where $\lfloor x \rfloor$ denotes the lower integer part of x. Hence for the allowed simplex chains, we have $\prod_{j=1}^{k} C_{i_j} \to 0$ ($k \to \infty$) and

$$S^{(k+K)} = S^{(K)} F \begin{bmatrix} 1 & 0^T \\ 0 & \prod_{j=1}^{k} C_{i_j} \end{bmatrix} F^{-1} \to S^{(K)} F \begin{bmatrix} 1 & 0^T \\ 0 & 0 \end{bmatrix} F^{-1}.$$

Since

$$F \begin{bmatrix} 1 & 0^T \\ 0 & 0 \end{bmatrix} F^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

we have

$$S^{(k+K)} \to S^{(K)} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \left[x_1^{(K)}, x_1^{(K)}, x_1^{(K)}\right].$$

It follows that $f_1^* = f\left(x_1^{(K)}\right)$, $f_2^* = f\left(x_1^{(K)}\right)$ and $f_3^* = f\left(x_1^{(K)}\right)$, which is contradiction. Hence $f_1^* = f_2^* = f_3^*$. $\square$

Remark 2. The bondedness of the level sets of f is not assumed unlike Lagarias, Reeds, Wright and Wright [17].

Remark 3. The proof exploits the Jensen inequality for convex functions. Nikodem [24] proved that a function f defined on a convex open subset D of $\mathbb{R}^n$ is convex if and only if it is midconvex and quasi-convex (for extensions, see [3], [33]). Hence the proof does not apply to quasiconvex functions.

We consider now the limitations of Theorem 3. The following example (see also Example 2 of [8]) indicates that the boundedness of f from below is essential.

Example 4. The expansion point $x_r^{(k)}$ is the incoming vertex infinitely many times if

$$f_r^{(k)} < f_1^{(k)} < f_2^{(k)} < f_3^{(k)} \text{ and } f_r^{(k)} \le f_e^{(k)} \quad (k \ge 0). \tag{21}$$

Consider the simplex sequence

$$S^{(k)} = \begin{bmatrix} c+(k+1)\delta & c+k\delta & c+(k-1)\delta \\ (-1)^k\gamma & (-1)^{k+1}\gamma & (-1)^k\gamma \end{bmatrix} \quad (k \ge 0) \tag{22}$$

with

$$x_r^{(k)} = \begin{bmatrix} c+(k+2)\delta \\ (-1)^{k+1}\gamma \end{bmatrix}, \quad x_e^{(k)} = \begin{bmatrix} c+\left(k+\frac{7}{2}\right)\delta \\ 2(-1)^{k+1}\gamma \end{bmatrix}.$$

Observe that the simplex sequence $\left\{S^{(k)}\right\}$ is unbounded, while $\operatorname{diam}\left(S^{(k)}\right)$ is constant. If $\gamma \ge \frac{\sqrt{\delta}}{2}$, the convex function $f_1(x,y) = -\frac{1}{2}x + y^2$ satisfies (21) and $f_i^{(k)} \to -\infty$ $(i = 1,2,3,\ k \to \infty)$. Note that f_1 has no bounded level sets. The function $f_2(x,y) = e^{-x} + y^2$ is strictly convex, its Hesse matrix is everywhere positive definite and f_2 is bounded below. For a large enough $c > 0$ and for any small $\delta, \gamma > 0$, it satisfies (21), and $f_2\left(x_i^{(k)}\right) \to \gamma^2$ $(i = 1,2,3)$ for $k \to \infty$, while $\inf f(x,y) = 0$. Note that the level sets of f_2 are not bounded.

The example also shows that the Nelder-Mead method is sensitive to the initial simplex (see also [25], [31]).

The next example shows that even if we have the convergence of simplex vertices to finite limit points, they are not necessarily identical.

Example 5. Consider the (pig-through) function

$$f(x,y) = \frac{1}{2}(x+y)^2,$$

which is convex and bounded below. Take the initial simplex

$$S^{(0)} = \begin{bmatrix} -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & \frac{1}{4}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} & \frac{1}{4}\sqrt{2} \end{bmatrix}.$$

Then

$$S^{(k)} = \begin{bmatrix} -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & \frac{\sqrt{2}}{2^{k+2}} \\ \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} & \frac{\sqrt{2}}{2^{k+2}} \end{bmatrix}$$

with $x_1^{(k)} = x_1^{(0)}$, $x_2^{(k)} = x_2^{(0)}$, $f_1^{(k)} = f_2^{(k)} = 0$, $f_3^{(k)} = \frac{1}{2^{2k+2}}$. Since $x_c^{(k)} = (0,0)$, $x_r^{(k)} = -\left(\frac{\sqrt{2}}{2^{k+2}}, \frac{\sqrt{2}}{2^{k+2}}\right)$, $f_r^{(k)} = \frac{1}{2^{2k+2}}$, $x_{ic}^{(k)} = \left(\frac{\sqrt{2}}{2^{k+3}}, \frac{\sqrt{2}}{2^{k+3}}\right)$ and $f_{ic}^{(k)} = \frac{1}{2^{2k+4}}$, $x_3^{(k+1)} = x_{ic}^{(k)}$. Hence $x_3^{(k)} \to (0,0)$ and $f_3^{(k)} \to 0$ $(k \to \infty)$. Note that $\lim_k x_3^{(k)} = \frac{1}{2}\left(x_1^{(0)} + x_2^{(0)}\right)$.

6 Closing remarks

Considering the counterexamples of this paper and those of [19], [7], [6], [8] we may conclude that even for (strictly) convex functions, there is no guarantee that the simplex sequence $\left\{S^{(k)}\right\}$ converge to a minimum point. Hence the most general convergence result for the NM method is Lemma 2 of Lagarias et al. [17], which proves the monotone convergence of the function values at the vertices and the limit values are the best available (the NM method starting from $S^{(0)}$ generates a unique simplex sequence with the aforementioned property).

The ideal situation would be as follows: the simplex sequence converges to a limit of the form $\widehat{x}e^T$, where $\widehat{x}$ is a (local/global) minimum or at least stationary point of f. The matrix form of the Nelder-Mead method is

$$S^{(k+1)} = S^{(k)} T_k P^{(k)} \quad (k = 0, 1, \ldots)$$

which is a nonstationary iteration with $\left\|T_k P^{(k)}\right\| \geq 1$. Hence the convergence depends on the convergence of the right infinite matrix product $\prod_{i=1}^{\infty} T_i P^{(i)}$. The convergence of simplex vertices to a common limit point is studied in [6], [8] and [9]. If $S^{(k)} \to \widehat{x}e^T$ for some vector $\widehat{x}$, then $f_1^* = \cdots = f_{n+1}^* = f(\widehat{x})$.

Under the assumptions of Lemma 2 or Lemma 6 plus those of [6], [8] and [9] we may conclude that $f(\widehat{x})$ is the best available value (starting from $S^{(0)}$). As noted earlier $f(\widehat{x})$ is not necessarily a local minimum of f. However it is certainly an improvement of the initial function value $f\left(x_1^{(0)}\right)$.

References

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

[2] M. Avriel, W.E. Diewert, S. Schaible, and I. Zang. *Generalized Concavity*. SIAM, 2010.

[3] F.A. Behringer. Convexity is equivalent to midpoint convexity combined with strict quasiconvexity. *Optimization*, 24:219–228, 1992.

[4] A. Cambini and L. Martein. *Generalized Convexity and Optimization*. Springer, 2017.

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

[6] A. Galántai. Convergence theorems for the Nelder-Mead method. *J. Comput. Appl. Mech.*, 15(2):115–133, 2020.

[7] A. Galántai. A convergence analysis of the Nelder-Mead simplex method. *Acta Polytech. Hung.*, 18(5):93–105, 2021.

[8] A. Galántai. Convergence of the Nelder-Mead method. *Numer. Algorithms*, 90:1043–1072, 2022.

[9] A. Galántai. A stochastic convergence result for the Nelder-Mead simplex method. *Mathematics*, 11:1998., 2023.

[10] F. Gao and L. Han. Implementing the nelder-mead simplex algorithm with adaptive parameters. *Comput Optim Appl*, 51:259–277, 2012.

[11] L. Han and M. Neumann. Effect of dimensionality on the Nelder-Mead simplex method. *Optim. Method. Softw.*, 21(1):1–16, 2006.

[12] P. Heeman. On derivative-free optimisation methods. Master's thesis, University of Bergen, 2023.

[13] C.T. Kelley. Detection and remediation of stagnation in the Nelder-Mead algorithm using an sufficient decrease condition. *SIAM J. Optimiz.*, 10(1):43–55, 1999.

[14] C.T. Kelley. *Iterative Methods for Optimization*. SIAM, Philadelphia, USA, 1999.

[15] M.J. Kochenderfer and T.A. Wheeler. *Algorithms for Optimization*. The MIT Press, 2019.

[16] J.C. Lagarias, B. Poonen, and M.H. Wright. Convergence of the restricted Nelder-Mead algorithm in two dimensions. *SIAM J. Optimiz.*, 22(2):501–532, 2012.

[17] J.C. Lagarias, J.A. Reeds, M.H. Wright, and P.E. Wright. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM J. Optimiz.*, 9(1):112–147, 1998.

[18] J. Larson, M. Menickelly, and S.M. Wild. Derivative-free optimization methods. *Acta Numerica*, 28:287–404, 2019.

[19] K.I.M. McKinnon. Convergence of the Nelder-Mead simplex method to a nonstationary point. *SIAM J. Optimiz.*, 9(1):148–158, 1998.

[20] J.J. Moré and S.M. Wild. Benchmarking derivative-free optimization algorithms. *SIAM J. Optimiz.*, 20(1):172–191, 2009.

[21] C. J. Nash. On best practice optimization methods in R. *Journal of Statistical Software*, 60(2):1–14, 2014.

[22] J. A. Nelder and R. Mead. A simplex method for function minimization. *Comput. J.*, 7:308–313, 1965.

[23] Y. Nesterov and V. Spokoiny. Random gradient-free minimization of convex functions. *Found Comput Math*, 17:527–566, 2017.

[24] K. Nikodem. On some class of midkonvex functions. *Annales Polonici Mathematici*, L:145–151, 1989.

[25] J.M. Parkinson and D. Hutchinson. An investigation into the efficiency of variants on the simplex method. In Lootsma. F.A., editor, *Numerical Methods for Non-linear Optimization*, pages 115–135, London, New York, 1972. Academic Press.

[26] L.M. Rios and N.V. Sahinidis. Derivative-free optimization: a review of algorithms and comparison of software implementations. *J Glob Optim*, 56:1247–1293, 2013.

[27] A.S. Rykov. *System Analysis. Models and Methods of Decision Making and Search Engine Optimization (in Russian)*. MISiS, 2009.

[28] H.-P. Schwefel. *Evolution and Optimum Seeking*. Wiley-Interscience, 1995.

[29] W.F. Trench. Characterization and properties of matrices with k-involutory symmetries. *Linear Algebra and its Applications*, 429:2278–2290, 2008.

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

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

[32] M.H. Wright. Nelder, Mead, and the other simplex method. *Doc. Math., Extra Volume: Optimization Stories*, pages 271–276, 2012.

[33] X.M. Yang, K.L. Teo, and X.Q. Yang. A characterization of convex function. *Applied Mathematics Letters*, 13:27–30, 2000.