Effect of dimensionality on the Nelder–Mead simplex method

Article *in* Optimization Methods and Software · February 2006

DOI: 10.1080/10556780512331318290

Lixing Han
University of Michigan-Flint

Optimization Methods and Software
Vol. 21, No. 1, February 2006, 1–16

Taylor & Francis
Taylor & Francis Group

Effect of dimensionality on the Nelder–Mead simplex method

LIXING HAN*† and MICHAEL NEUMANN‡

†Department of Mathematics, University of Michigan-Flint, Flint, MI 48502, USA
‡Department of Mathematics, University of Connecticut, Storrs, CT 06269, USA

(Received 5 June 2003; in final form 8 July 2004)

The effect of dimensionality on the widely used Nelder–Mead simplex method for unconstrained optimization is investigated. It is shown that by using the quadratic function $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{x}$, the Nelder–Mead simplex method deteriorates as the dimension increases.

Keywords: Nelder–Mead method; Simplex; Effect of dimensionality; Convergence; Optimization

AMS Subject Classification: 90C56; 90C30; 65K05

1. Introduction

We consider the following unconstrained optimization problem

$$\min f(\mathbf{x}), \tag{1}$$

where the function $f\colon \mathbb{R}^n \to \mathbb{R}$ is called the objective function and $\mathbf{x} = (\xi_1, \xi_2, \ldots, \xi_n)^{\mathrm{T}} \in \mathbb{R}^n$.

In many applications, the objective function f is defined from a complex or convoluted computational structure. Therefore, its derivatives are unavailable or difficult to obtain. In these cases, a direct search method is often used to solve problem (1); for excellent surveys on direct search methods, see refs. [9,22]. One of the most widely used direct search methods is the simplex method of Nelder and Mead [12], see, for example, refs. [1,7,10,13,15,20,22]. Despite its popularity, only a few papers have studied its convergence properties. In recent paper [10], Lagarias *et al.* prove several convergence results for the Nelder–Mead method when applied to strictly convex functions with bounded level sets in low dimensions. Specifically, they show that if the dimension of the problem is $n = 1$, the simplices generated by the Nelder–Mead method converge to a minimizer of the objective function. For dimension $n = 2$, they prove that the simplex diameters converge to zero. However, it is still an open question whether the Nelder–Mead simplices converge to the minimizer for the strictly convex quadratic function $\xi_1^2 + \xi_2^2$. In contrast, several negative results have been found.

*Corresponding author. Fax: 1-810-766-6880; Email: lxhan@umflint.edu

Optimization Methods and Software
ISSN 1055-6788 print/ISSN 1029-4937 online
http://www.tandf.co.uk/journals
DOI: 10.1080/10556780512331318290

In ref. [3], Dennis and Woods give a strictly convex non-differentiable example for which a modified Nelder–Mead method generates a sequence of simplices which converge to a single point at which there is no zero subgradient. Recently, McKinnon [11] presents a family of strictly convex functions in two dimensions with up to the third degree of smoothness and a special initial simplex, for which the Nelder–Mead simplices converge to a non-minimizer. Examples showing that the Nelder–Mead simplices converge to a degenerate simplex instead of a single point can be found in refs. [5, pp. 56–60] and [11, lines 19–20, p. 150].

An important question in the theoretical analysis of the Nelder–Mead method involves the effect of dimensionality [22]. Many authors state that the Nelder–Mead method performs well when the dimension n is small, but may be inefficient when n is large. For example, Byatt [2], Rowan [16], Torczon [18], and Wright [22] find numerically that the Nelder–Mead method can perform very poorly for the quadratic function $\xi_1^2 + \xi_2^2 + \cdots + \xi_n^2$, when n is moderately large, say $n = 32$. However, no theoretical analysis exists for explaining the effect of dimensionality on the Nelder–Mead method. Much analysis needs to be done to complete our understanding why the Nelder–Mead method is so inefficient in higher dimensions; see refs. [10,22].

In this paper, we study the effect of dimensionality on the Nelder–Mead method. Specifically, we consider the quadratic function $\xi_1^2 + \xi_2^2 + \cdots + \xi_n^2$ and show that the Nelder–Mead method inherently becomes less efficient as the dimension increases. The considered example offers insight into understanding the behavior of the Nelder–Mead method in higher dimensions.

This paper is organized as follows. Section 2 gives an outline of the Nelder–Mead method. The effect of dimensionality is studied in Section 3. Section 4 gives some final remarks.

2. The Nelder–Mead simplex method

The Nelder–Mead simplex method for unconstrained optimization was proposed by Nelder and Mead [12], which is a variation of the simplex method of Spendley *et al.* [17]. The Nelder–Mead simplex method maintains a simplex S of approximations to an optimal point. A simplex is a geometric figure in n dimensions of non-zero volume that is the convex hull of $n + 1$ vertices. For instance, a simplex in one dimension is a line segment; a simplex in two dimensions is a triangle; a simplex in three dimensions is a tetrahedron. We denote a simplex with vertices $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_n$, by $S = [\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_n]$.

We now outline the Nelder–Mead method, which is the version given in Lagarias *et al.* [10].

In the Nelder–Mead method, the vertices $\{\mathbf{v}_j\}_{j=0}^n$ of the simplex are sorted according to the objective function values

$$f(\mathbf{v}_0) \leq f(\mathbf{v}_1) \leq \cdots \leq f(\mathbf{v}_n). \tag{2}$$

We refer to $\mathbf{v}_0$ as the *best* vertex, and to $\mathbf{v}_n$ as the *worst* vertex. If several vertices have the same objective values, consistent tie-breaking rules such as those given in Lagarias *et al.* [10] are required for the method to be well defined.

The algorithm attempts to formulate a new simplex by replacing the worst vertex $\mathbf{v}_n$ with a new point at which the function has a smaller value than $f(\mathbf{v}_n)$. There are three possible operations: *reflection*, *expansion*, and *contraction* to achieve this. If the earlier mentioned operations cannot give a smaller function value, then a *shrink* step is performed, in which, a set of n new vertices, together with the best vertex $\mathbf{v}_0$ form a new simplex. Each of the earlier mentioned operations is associated with scalar parameters: α (reflection), β (expansion), γ (contraction), and δ (shrinkage). The values of these parameters satisfy $\alpha > 0$, $\beta > 1$,

$0 < \gamma < 1$, and $0 < \delta < 1$. A typical sequence of values for these parameters is

$$\{\alpha, \beta, \gamma, \delta\} = \left\{1, 2, \frac{1}{2}, \frac{1}{2}\right\}. \tag{3}$$

We use this choice of parameters in all our analysis of the Nelder–Mead method and numerical tests in this paper.

Let $\bar{\mathbf{v}}$ be the centroid of the n best vertices. Then

$$\bar{\mathbf{v}} = \frac{1}{n}\sum_{i=0}^{n-1} \mathbf{v}_i. \tag{4}$$

ONE ITERATION OF THE NELDER–MEAD ALGORITHM

1. Sort: *Evaluate f at the $n+1$ vertices of S and sort the vertices so that inequalities* (2) *hold.*
2. Reflection: *Compute the reflection point* $\mathbf{v}_\mathrm{r}$ *from*

$$\mathbf{v}_\mathrm{r} = \bar{\mathbf{v}} + \alpha(\bar{\mathbf{v}} - \mathbf{v}_n).$$

Evaluate $f_\mathrm{r} = f(\mathbf{v}_\mathrm{r})$. *If* $f_0 \le f_\mathrm{r} < f_{n-1}$, *replace* $\mathbf{v}_n$ *with* $\mathbf{v}_\mathrm{r}$.

3. Expansion: *If* $f_\mathrm{r} < f_0$, *then compute the expansion point* $\mathbf{v}_\mathrm{e}$ *from*

$$\mathbf{v}_\mathrm{e} = \bar{\mathbf{v}} + \beta(\mathbf{v}_\mathrm{r} - \bar{\mathbf{v}})$$

and evaluate $f_\mathrm{e} = f(\mathbf{v}_\mathrm{e})$. *If* $f_\mathrm{e} < f_\mathrm{r}$, *replace* $\mathbf{v}_n$ *with* $\mathbf{v}_\mathrm{e}$; *otherwise, replace* $\mathbf{v}_n$ *with* $\mathbf{v}_\mathrm{r}$.

4. Outside Contraction: *If* $f_{n-1} \le f_\mathrm{r} < f_n$, *compute the outside contraction point*

$$\mathbf{v}_\mathrm{oc} = \bar{\mathbf{v}} + \gamma(\mathbf{v}_\mathrm{r} - \bar{\mathbf{v}})$$

and evaluate $f_\mathrm{oc} = f(\mathbf{v}_\mathrm{oc})$. *If* $f_\mathrm{oc} < f_\mathrm{r}$, *replace* $\mathbf{v}_n$ *with* $\mathbf{v}_\mathrm{oc}$.

5. Inside Contraction: *If* $f_\mathrm{r} \ge f_n$, *compute the inside contraction point* x_ic *from*

$$\mathbf{v}_\mathrm{ic} = \bar{\mathbf{v}} - \gamma(\mathbf{v}_\mathrm{r} - \bar{\mathbf{v}})$$

and evaluate $f_\mathrm{ic} = f(\mathbf{v}_\mathrm{ic})$. *If* $f_\mathrm{ic} < f_n$, *replace* $\mathbf{v}_n$ *with* $\mathbf{v}_\mathrm{ic}$; *otherwise, go to step* 6.

6. Shrink: *For* $1 \le i \le n$, *define*

$$\mathbf{v}_i = \mathbf{v}_0 + \delta(\mathbf{v}_i - \mathbf{v}_0).$$

The previously mentioned algorithm is slightly different from the original method in ref. [12]. Specifically, in the original Nelder–Mead method, $\mathbf{v}_\mathrm{e}$ replaces $\mathbf{v}_n$ if it is better than $\mathbf{v}_0$; $\mathbf{v}_\mathrm{r}$ is accepted if $f_\mathrm{r} = f_{n-1}$; and $\mathbf{v}_\mathrm{ic}$ is accepted if $f_\mathrm{r} = f_n$.

To form the initial simplex, it is typical to choose a starting point x_0 that is taken as one of the initial simplex vertices. The remaining n vertices are then generated by setting as $x_0 + \tau_i \mathbf{e}_i$, where $\mathbf{e}_i$ is the unit vector in the ith coordinate, and τ_i is a suitable parameter.

Several termination criteria have been proposed in the literature. For example, given the tolerance, tol, Nelder and Mead [12] suggest termination when

$$\sqrt{\frac{1}{n}\sum_{i=0}^{n}(f_i-\bar{f})^2} \leq \text{ tol}, \tag{5}$$

where

$$\bar{f}=\frac{1}{n+1}\sum_{i=0}^{n} f_i.$$

Another termination criterion suggested by Torczon [18] and Woods [21] is

$$\sigma(S) := \max_{1\leq i\leq n} \|\mathbf{v}_i-\mathbf{v}_0\| \leq \text{ tol } \max(1, \|\mathbf{v}_0\|), \tag{6}$$

where $\|\cdot\|$ denotes the Euclidean norm. Following Kelley [7], we call $\sigma(S)$ the *oriented length* of the simplex S. Some arguments in the favor of this criterion are given in Dennis and Woods [3] and Torczon [18].

'tol' is set to be zero in the convergence analysis in the following section.

3. The effect of dimensionality

In this section, we analyze the effect of dimensionality when the Nelder–Mead method is applied to minimize the function

$$f(\mathbf{x})=\xi_1^2+\xi_2^2+\cdots+\xi_n^2 \tag{7}$$

with the initial simplex

$$S_0=[\mathbf{0}, \mathbf{v}_1^{(0)}, \ldots, \mathbf{v}_n^{(0)}]. \tag{8}$$

With this choice of the initial simplex, the best vertex remains fixed at $\mathbf{0}\in\mathbb{R}^n$, the origin.

One of our motivations is that in refs. [2,16,18,22], it is found that for the function (7), the Nelder–Mead method suffers from poor performance as dimension n increases. Our second motivation is that the Nelder–Mead method is invariant under affine transformations [Lemma 3.2 of ref. 10]. Thus, we can reduce the study of the Nelder–Mead method for a general strictly convex quadratic function to the study of the function (7).

We comment that quadratic functions have been used to analyze other optimization algorithms in the literature. For instance, Powell [14] employs the quadratic function (7) with $n=2$ to study the behavior of the BFGS and DFP quasi-Newton methods. The quadratic functions have also been used to study the efficiency of the steepest descent method [see, for example, ref. 7] and the more recent Barzilai–Borwein gradient method, see ref. [4].

One concern is the generality of using the function (7). We note that, locally at least, the contours of a sufficiently smooth function appear quadratic. Another concern is the generality of using the initial simplex (8). The initial simplex is chosen because it is, in some sense, the best initial simplex (one of the simplex's vertices is the solution to the problem) and so regardless of choice of a more general simplex it is hoped that eventually it will be as good as the initial simplex chosen. We are grateful to both referees for suggesting discussion on generality of using equations (7) and (8). In particular, this paragraph is based on the suggestions of one referee.

As the function in equation (7) is strictly convex, the Nelder–Mead method never performs the *shrink* step, see ref. [10]. Therefore, at each iteration, a new simplex is formed by replacing the worst vertex by a new, better vertex.

We assume that the Nelder–Mead method generates a sequence of simplices $\{S_k\}$ in $\mathbb{R}^n$, where

$$S_k = [\mathbf{0}, \mathbf{v}_1^{(k)}, \mathbf{v}_2^{(k)}, \ldots, \mathbf{v}_n^{(k)}].$$

We wish that the sequence of simplices $\{S_k\} \to \mathbf{0} \in \mathbb{R}^n$ as $k \to \infty$. To measure the progress of convergence, we introduce the following oriented length of the simplex S_k:

$$\sigma(S_k) := \max_{1 \le j \le n} \|\mathbf{v}_j^{(k)}\|. \tag{9}$$

We say that a sequence of simplices $\{S_k\}$ converges to the minimizer $\mathbf{0} \in \mathbb{R}^n$ of the function in equation (7) if $\lim_{k\to\infty} \sigma(S_k) = 0$.

One of the key features of the performance of an algorithm is its rate of convergence. To measure how fast the Nelder–Mead method converges when it is used to solve the problem (7) with the initial simplex given by equation (8), we introduce the following definition.

DEFINITION 3.1 *Let* $\{S_k\}$ *be a sequence of simplices that converges to the minimizer* 0, *we measure the rate of convergence by*

$$\rho(S_0, n) := \limsup_{k\to\infty} \left(\prod_{i=0}^{k-1} \frac{\sigma(S_{i+1})}{\sigma(S_i)} \right)^{1/k}. \tag{10}$$

This definition implies $\rho(S_0, n) = \limsup_{k\to\infty} (\sigma(S_k)/\sigma(S_0))^{1/k}$. The form of equation (10) is emphasized here, because it can be viewed as the 'geometric mean' of the ratio of the oriented lengths between successive simplices and the minimizer $\mathbf{0}$.

According to the definition, the algorithm is convergent if $0 \le \rho(S_0, n) < 1$. The larger the $\rho(S_0, n)$, the slower the convergence. In particular, the convergence is very slow when $\rho(S_0, n)$ is close to 1. We shall illustrate that as dimension increases, the rate of convergence of the Nelder–Mead method becomes slower.

Our analysis is based on investigating the case that the Nelder–Mead method generates a sequence of simplices in $\mathbb{R}^n$ satisfying

$$S_k = [\mathbf{0}, \mathbf{v}^{(k+n-1)}, \ldots, \mathbf{v}^{(k+1)}, \mathbf{v}^{(k)}], \tag{11}$$

where $\mathbf{0}, \mathbf{v}^{(k+n-1)}, \ldots, \mathbf{v}^{(k+1)}, \mathbf{v}^{(k)}$ are the vertices of the kth simplex, with

$$f(\mathbf{0}) < f(\mathbf{v}^{(k+n-1)}) < \cdots < f(\mathbf{v}^{(k)}), \tag{12}$$

for $k = k_0, k_0 + 1, k_0 + 2, \ldots$, where k_0 is a non-negative integer. Without loss of generality, we assume that $k_0 = 0$. Note that equation (12) means that the vertex replacing the worst vertex becomes the second best vertex in the new simplex.

Our approach can be considered as a 'favorable case analysis'. We consider the Nelder–Mead method being applied to the function (7), using the initial simplex (8), and having favorable moves (12). We shall demonstrate that even in this favorable situation, the convergence rate $\rho(S_0, n)$ can become very close to 1 as dimension increases. For more difficult functions and more general initial simplices, we expect the Nelder–Mead method suffers from poorer performance in higher dimensions.

3.1 *Recurrence Formulas*

To simplify the analysis, we first consider that one type of step of the Nelder–Mead method is applied repeatedly. This allows us to establish recurrence equations for the successive simplex vertices. As the *shrink* step is never used and moreover the *expansion* step is never used either since the best vertex remains fixed at the origin, we only need to consider the *outside contraction*, *inside contraction*, and *reflection* steps.

The centroid of the n best vertices of S_k is given by

$$\bar{\mathbf{v}}^{(k)} = \frac{\sum_{i=1}^{n-1} \mathbf{v}^{(k+i)}}{n}.$$

Case 1 Outside Contraction: If the *outside contraction* step is repeatedly performed, then by the parameter-choice (3) and the definition of the *outside contraction* step, we have

$$\mathbf{v}^{(k+n)} = \bar{\mathbf{v}}^{(k)} + \frac{1}{2}(\bar{\mathbf{v}}_k - \mathbf{v}^{(k)}).$$

Thus, we have the following recursive formula

$$\mathbf{v}^{(k+n)} = \frac{3}{2n}(\mathbf{v}^{(k+1)} + \cdots + \mathbf{v}^{(k+n-1)}) - \frac{1}{2}\mathbf{v}^{(k)} \tag{13}$$

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

The characteristic equation for the recurrence equation (13) is

$$2n\mu^n - 3\mu^{n-1} - \cdots - 3\mu + n = 0. \tag{14}$$

Case 2 Inside Contraction: If the *inside contraction* step is repeatedly performed, then by the parameter-choice (3) and the definition of the *inside contraction* step, we have

$$\mathbf{v}^{(k+n)} = \bar{\mathbf{v}}^{(k)} - \frac{1}{2}(\bar{\mathbf{v}}_k - \mathbf{v}^{(k)}).$$

Thus, we have the following recursive formula

$$\mathbf{v}^{(k+n)} = \frac{1}{2n}(\mathbf{v}^{(k+1)} + \cdots + \mathbf{v}^{(k+n-1)}) + \frac{1}{2}\mathbf{v}^{(k)} \tag{15}$$

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

The characteristic equation for the recurrence equation (15) is

$$2n\mu^n - \mu^{n-1} - \cdots - \mu - n = 0. \tag{16}$$

Case 3 Reflection: If the *reflection* step is repeatedly performed, then by the parameter-choice (3) and the definition of the *reflection* step, we have

$$\mathbf{v}^{(k+n)} = \bar{\mathbf{v}}^{(k)} + (\bar{\mathbf{v}}_k - \mathbf{v}^{(k)}).$$

Thus, we have the following recursive formula

$$\mathbf{v}^{(k+n)} = \frac{2}{n}(\mathbf{v}^{(k+1)} + \cdots + \mathbf{v}^{(k+n-1)}) - \mathbf{v}^{(k)} \tag{17}$$

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

The characteristic equation for the recurrence equation (17) is

$$n\mu^n - 2\mu^{n-1} - \cdots - 2\mu + n = 0. \tag{18}$$

Recurrence equations (13), (15), and (17) are linear. Their general solutions are of the form

$$\mathbf{v}^{(k)} = \mu_1^k \mathbf{a}_1 + \mu_2^k \mathbf{a}_2 + \cdots + \mu_n^k \mathbf{a}_n, \tag{19}$$

where $\mu_1, \mu_2, \ldots, \mu_n$ are the characteristic values of the characteristic equations (14), (16), and (18). $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are independent vectors in $\mathbb{C}^n$ such that $\mathbf{v}^{(k)} \in \mathbb{R}^n$ for all $k \geq 0$.

For the roots of the characteristic equations (14), (16), and (18), we have the following estimates.

LEMMA 3.2 [see ref. 6] *Let $\mu_{\mathrm{oc}}(n)$, $\mu_{\mathrm{ic}}(n)$, and $\mu_{\mathrm{r}}(n)$ be a root of equations* (14), (16), *and* (18), *respectively. Then we have*

$$|\mu_{\mathrm{oc}}(n)| < 1, \tag{20}$$

$$|\mu_{\mathrm{ic}}(n)| < 1, \tag{21}$$

and

$$|\mu_{\mathrm{r}}(n)| = 1. \tag{22}$$

3.2 *Convergence*

Equations (22) and (19) imply that if the *reflection* step is repeatedly performed for all large k, then the simplex S_k cannot converge to the minimizer $\mathbf{0} \in \mathbb{R}^n$. Therefore, we will consider the convergence when there are infinitely many *inner contraction* or *outer contraction* steps. We have the following theorem.

THEOREM 3.3 *Assume that the Nelder–Mead method generates a sequence of simplices satisfying equations* (11) *and* (12). *If there are infinitely many inner contraction or outer contraction steps*, *then*

$$\lim_{k\to\infty} \mathbf{v}^{(k)} = \mathbf{0} \in \mathbb{R}^n.$$

Proof We consider $n > 1$ only. The $n = 1$ case is analyzed in the next section.

First, note that $\{f(\mathbf{v}^{(k)})\}$ is a decreasing sequence and bounded below. Thus,

$$\lim_{k\to\infty} f(\mathbf{v}^{(k)}) = f^*. \tag{23}$$

We will show that $f^* = 0$ as this is equivalent to $\lim_{k\to\infty} \mathbf{v}^{(k)} = 0$. Assume that $f^* > 0$. We seek a contradiction.

Similar to ref. [10], we introduce matrix notation

$$M_k = [\mathbf{v}^{(k+n-1)} \cdots \mathbf{v}^{(k+1)}\ \mathbf{v}^{(k)}],$$

where $\mathbf{v}^{(i)} \in \mathbb{R}^n$ is a column vector for $i = k, \ldots, k+n-1$. Then equations (11) and (12) imply

$$M_{k+1} = M_k A_k,$$

where A_k is a matrix of the form

$$A_k = \begin{bmatrix} \dfrac{1+\tau_k}{n}\mathbf{e}_{n-1} & I_{n-1} \\ -\tau_k & \mathbf{0}_{n-1}^{\mathrm{T}} \end{bmatrix},$$

where $\mathbf{e}_{n-1} = (1, 1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^{n-1}$, I_{n-1} is the identity matrix of order $n-1$, $\mathbf{0}_{n-1} = (0, 0, \ldots, 0)^{\mathrm{T}} \in \mathbb{R}^{n-1}$, and the parameter τ_k has one of the following three possible values:

$$\tau_k = 1 \text{ (reflection)}, \quad \tau_k = \frac{1}{2} \text{ (outer contraction)}, \quad \tau_k = -\frac{1}{2} \text{ (inner contraction)}.$$

Define

$$\mathbf{w}^{(i,k)} = (u_i^{(k+n-1)}, \ldots, u_i^{(k)})^{\mathrm{T}} \in \mathbb{R}^n$$

for $i = 1, 2, \ldots, n$, where $u_i^{(j)}$ is the ith component of vector $\mathbf{v}^{(j)}$, $j = k, \ldots, k+n-1$. Then

$$M_k^{\mathrm{T}} = [\mathbf{w}^{(1,k)}, \ldots, \mathbf{w}^{(n,k)}].$$

Introduce a positive definite quadratic function in $\mathbb{R}^n$

$$\Psi(\mathbf{x}) = \mathbf{x}^{\mathrm{T}} H \mathbf{x},$$

where

$$H = \begin{bmatrix} n & -1 & \cdots & -1 \\ -1 & n & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots \\ -1 & -1 & \cdots & n \end{bmatrix}_{n \times n}.$$

By direct computation, we have

$$H - A_k H A_k^{\mathrm{T}} = (1-\tau_k^2) \begin{bmatrix} \dfrac{1}{n} & \cdots & \dfrac{1}{n} & -1 \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{1}{n} & \cdots & \dfrac{1}{n} & -1 \\ -1 & \cdots & -1 & n \end{bmatrix}_{n \times n}.$$

Therefore, for any fixed i: $1 \le i \le n$, we have

$$\begin{aligned} \Psi(\mathbf{w}^{(i,k)}) - \Psi(\mathbf{w}^{(i,k+1)}) &= (\mathbf{w}^{(i,k)})^{\mathrm{T}} H \mathbf{w}^{(i,k)} - (\mathbf{w}^{(i,k)})^{\mathrm{T}} A_k H A_k^{\mathrm{T}} \mathbf{w}^{(i,k)} \\ &= n(1-\tau_k^2) \left(\frac{u_i^{(k+n-1)} + \cdots + u_i^{(k+1)}}{n} - u_i^{(k)} \right)^2. \end{aligned} \tag{24}$$

As $\tau_k = 1, 1/2,\ \text{or} - 1/2$, $\{\Psi(\mathbf{w}^{(i,k)})\}$ is a non-increasing sequence in k for any fixed i. Note that this sequence is bounded below. Thus, it has a finite limit and the series

$$\sum_{k=0}^{\infty}(\Psi(\mathbf{w}^{(i,k)}) - \Psi(\mathbf{w}^{(i,k+1)})) \tag{25}$$

is convergent.

Because there are infinitely many *inner contraction* or *outer contraction* steps, we consider the subsequence when $\tau_{k_l} = 1/2$ or $-1/2$ is used. Using equations (24) and (25), we have

$$\lim_{k_l\to\infty}\left(\frac{u_i^{(k_l+n-1)} + \cdots + u_i^{(k_l+1)}}{n} - u_i^{(k_l)}\right) = 0.$$

for $i = 1, 2, \ldots, n$. We choose sufficiently large k_l so that

$$u_i^{(k_l)} = \frac{u_i^{(k_l+n-1)} + \cdots + u_i^{(k_l+1)}}{n} + \epsilon_i,$$

where $|\epsilon_i| < \sqrt{f^*}(n-1)/(2n^2\sqrt{n})$, for $i = 1, 2, \ldots, n$. Thus,

$$\mathbf{v}^{(k_l)} = \frac{\mathbf{v}^{(k_l+n-1)} + \cdots + \mathbf{v}^{(k_l+1)}}{n} + \epsilon,$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^{\mathrm{T}} \in \mathbb{R}^n$. Using the triangular inequality, we have for sufficiently large k_l,

$$\|\mathbf{v}^{(k_l)}\| \le \frac{1}{n}(\|\mathbf{v}^{(k_l+n-1)}\| + \cdots + \|\mathbf{v}^{(k_l+1)}\|) + \frac{n-1}{2n^2}\sqrt{f^*}.$$

As $\|\mathbf{v}^{(k)}\| \ge \sqrt{f^*}$ for all k, we have

$$\sqrt{f^*} \le \frac{1}{n-1}(\|\mathbf{v}^{(k_l+n-1)}\| + \cdots + \|\mathbf{v}^{(k_l+1)}\|).$$

Thus,

$$\|\mathbf{v}^{(k_l)}\| \le \left(\frac{1}{n} + \frac{1}{2n^2}\right)(\|\mathbf{v}^{(k_l+n-1)}\| + \cdots + \|\mathbf{v}^{(k_l+1)}\|).$$

Let $\|\mathbf{v}^{(k_l+j_0)}\| = \max\{\|\mathbf{v}^{(k_l+j)}\|\colon 1 \le j \le n-1\}$. Then

$$\|\mathbf{v}^{(k_l)}\| \le \left(\frac{1}{n} + \frac{1}{2n^2}\right)(n-1)\|\mathbf{v}^{(k_l+j_0)}\| < \|\mathbf{v}^{(k_l+j_0)}\|.$$

This contradicts inequalities (12). The proof is complete. ■

3.3 *Rate of convergence for dimensions* $n = 1$ *and* 2

When $n = 1$, the quadratic function in equation (7) becomes

$$f(\xi_1) = \xi_1^2. \tag{26}$$

The initial simplex is chosen as

$$S_0 = [0, \mathbf{v}^{(0)}],$$

where $\mathbf{v}^{(0)} = 1$. It is easy to show that the kth Nelder–Mead simplex is given by

$$S_k = [0, \mathbf{v}^{(k)}], \quad \mathbf{v}^{(k)} = 1/2^k.$$

As $k \to \infty$, the sequence of simplices $\{S_k\}$ converges to the minimizer $\mathbf{0}$, i.e.,

$$\sigma(S_k) \to 0.$$

Notice that

$$\frac{\sigma(S_{k+1})}{\sigma(S_k)} = \frac{1}{2}$$

for any k. Therefore, we have the following claim.

CLAIM 3.4 *Let f be the function* (26). *Suppose that the Nelder–Mead method is applied to f with the initial simplex* [0,1]. *Then the Nelder–Mead method applies the inside contraction repeatedly and generates a sequence of simplices $S_k = [0, 1/2^k]$ which converges to the minimizer* $\mathbf{0}$. *The rate of convergence is* $1/2$.

We now consider the quadratic function in $\mathbb{R}^2$:

$$f(\xi_1, \xi_2) = \xi_1^2 + \xi_2^2. \tag{27}$$

We shall show that, with some choice of the initial simplex, the Nelder–Mead method applies the *outside contraction* step repeatedly and generates a sequence of simplices

$$\{S_k\} = [\mathbf{0}, \mathbf{v}^{(k+1)}, \mathbf{v}^{(k)}]$$

satisfying

$$f(\mathbf{v}^{(k+1)}) < f(\mathbf{v}^{(k)}).$$

We comment that this type of behavior has been presented by McKinnon [11] in his analysis of nonconvergence examples for the Nelder–Mead method.

When $n = 2$, the characteristic equation (14) becomes

$$4\mu^2 - 3\mu + 2 = 0.$$

It has roots

$$\mu_{1,2} = r(\cos\theta \pm i \sin\theta) = \frac{3}{8} \pm \frac{\sqrt{23}}{8} i$$

with $r = \sqrt{2}/2$ and $\theta = \tan^{-1}(\sqrt{23}/3)$.

If we choose the initial simplex as $S_0 = [\mathbf{0}, \mathbf{v}^{(0)}, \mathbf{v}^{(1)}]$ with

$$\mathbf{v}^{(0)} = (1, 0)^{\mathrm{T}}, \quad \mathbf{v}^{(1)} = (3/8, -\sqrt{23}/8)^{\mathrm{T}},$$

then the solution for the recurrence equation (13) is given by

$$\mathbf{v}^{(k)} = \mathbf{a}_1 \mu_1^k + \mathbf{a}_2 \mu_2^k, \tag{28}$$

where

$$\mathbf{a}_1 = (1/2, i/2)^{\mathrm{T}}, \quad \mathbf{a}_2 = (1/2, -i/2)^{\mathrm{T}}.$$

For this initial simplex, we can show that the Nelder–Mead method applies the *outside contraction* step repeatedly. In fact, we have the following claim.

CLAIM 3.5 *Let f be the function given in equation (27). Suppose that the Nelder–Mead method is applied to f with the initial simplex* $[\mathbf{0}, (3/8, -\sqrt{23}/8)^{\mathrm{T}}, (1, 0)^{\mathrm{T}}]$. *Then the Nelder–Mead simplex method applies the outside contraction step repeatedly, and generates a sequence of simplices $S_k = [\mathbf{0}, \mathbf{v}^{(k+1)}, \mathbf{v}^{(k)}]$, with $\mathbf{v}^{(k)}$ defined by equation* (28). *Furthermore, we have*

$$\lim_{k \to \infty} \sigma(S_k) = 0$$

with the rate of convergence

$$\rho(S_0, 2) = \frac{\sqrt{2}}{2}.$$

Proof We can rewrite $\mathbf{v}^{(k)}$ in equation (28) as

$$\mathbf{v}^{(k)} = (r^k \cos k\theta, -r^k \sin k\theta)^{\mathrm{T}}.$$

Thus, the function value at the vertex $\mathbf{v}^{(k)}$ is

$$f^{(k)} = r^{2k}.$$

Consider the *reflection* point for the kth simplex S_k:

$$\mathbf{v}_{\mathrm{r}}^{(k)} = \bar{\mathbf{v}}_k + (\bar{\mathbf{v}}_k - \mathbf{v}^{(k)}) = \mathbf{v}^{(k+1)} - \mathbf{v}^{(k)}.$$

Then

$$\mathbf{v}_{\mathrm{r}}^{(k)} = (r^{k+1} \cos(k+1)\theta - r^k \cos k\theta, -r^{k+1} \sin(k+1)\theta + r^k \sin k\theta)^{\mathrm{T}},$$

and the value of function (27) at $\mathbf{v}_{\mathrm{r}}^{(k)}$ is

$$\begin{aligned} f_{\mathrm{r}}^{(k)} &= \|\mathbf{v}^{(k+1)} - \mathbf{v}^{(k)}\|^2 \\ &= r^{2k} + r^{2k+2} - 2r^{2k+1}[\cos k\theta \cos(k+1)\theta + \sin k\theta \sin(k+1)\theta] \\ &= r^{2k}[1 - 2r\cos\theta + r^2]. \end{aligned}$$

As $r = \sqrt{2}/2$ and $r \cos\theta = 3/8$, we have that

$$f^{(k+1)} < f_{\mathrm{r}}^{(k)} < f^{(k)}.$$

Moreover, using the *outside contraction* gives

$$\mathbf{v}^{(k+2)} = (r^{k+2} \cos(k+2)\theta, -r^{k+2} \sin(k+2)\theta)^{\mathrm{T}}$$

with

$$f^{(k+2)} = r^{2(k+2)} < f_{\mathrm{r}}^{(k)}.$$

Thus, the Nelder–Mead method applies the *outside contraction* step and the next simplex is $S_{k+1} = [\mathbf{0}, \mathbf{v}^{(k+2)}, \mathbf{v}^{(k+1)}]$. This pattern repeats for $k = 0, 1, 2, \ldots$

Notice that

$$\sigma(S_k) = r^k$$

for $k = 0, 1, 2, \ldots$ Therefore, we have

$$\sigma(S_k) \to 0, \quad \text{as} \quad k \to \infty.$$

As

$$\frac{\sigma(S_{k+1})}{\sigma(S_k)} = r = \frac{\sqrt{2}}{2},$$

for all $k \geq 0$, the rate of convergence is

$$\rho(S_0, 2) = \frac{\sqrt{2}}{2}.$$

■

3.4 *When dimension n ≥ 3*

We consider the quadratic function (7) and study of the effect of dimensionality on the Nelder–Mead when the dimension $n \geq 3$. The initial simplex is chosen as

$$S_0 = [\mathbf{0}, \mathbf{v}^{(n-1)}, \ldots, \mathbf{v}^{(1)}, \mathbf{v}^{(0)}]. \tag{29}$$

In this situation, the analysis becomes more complicated, because each of the three steps *outside contraction, inside contraction* and *reflection* can be used by the Nelder–Mead method.

As we have seen, if one type of step is applied repeatedly, then the Nelder–Mead method generates a sequence of simplices S_k of the form (11) with $\mathbf{v}^{(k)}$ satisfying equation (19). The characteristic values, especially those have largest modulus play an important role in the behavior of the Nelder–Mead method as $k \to \infty$.

As we are concerned with the effect of dimensionality when n becomes large, we now consider how the moduli of characteristic values behave as $n \to \infty$. We have the following theorem.

THEOREM 3.6 [see ref. 6] *Let $\mu_{\rm oc}(n)$ and $\mu_{\rm ic}(n)$ be a root of equations* (14) *and* (16), *respectively. Then we have*

$$\lim_{n\to\infty} |\mu_{\rm oc}(n)| = 1, \tag{30}$$

and

$$\lim_{n\to\infty} |\mu_{\rm ic}(n)| = 1. \tag{31}$$

For a closer examination of the characteristic values, we introduce following Definition.

DEFINITION 3.7 *Let $\mu_1, \mu_2, \ldots, \mu_n$ be the roots of equation* (14) *and $\nu_1, \nu_2, \ldots, \nu_n$ the roots of equation* (16). *We define vectors*

$$w = \mathrm{Sort}((|\mu_1|, |\mu_2|, \ldots, |\mu_n|)), \quad z = \mathrm{Sort}((|\nu_1|, |\nu_2|, \ldots, |\nu_n|)),$$

where Sort is in ascending order. We denote

$$S_{\rm oc}(n) = w_1, \quad \mathrm{SS}_{\rm oc}(n) = w_2, \quad \mathrm{SL}_{\rm oc}(n) = w_{n-1}, \quad L_{\rm oc}(n) = w_n,$$

and

$$S_{\rm ic}(n) = z_1, \quad \mathrm{SS}_{\rm ic}(n) = z_2, \quad \mathrm{SL}_{\rm ic}(n) = z_{n-1}, \quad L_{\rm ic}(n) = z_n.$$

Clearly, $L_{oc}(n)$ and $S_{oc}(n)$ are the largest and smallest moduli of the characteristic values of equation (14), respectively. $SL_{oc}(n)$ and $SS_{oc}(n)$ are the second largest (or the largest) and second smallest (or the smallest) moduli of the characteristic values of equation (14), respectively. Similar argument goes to the inner contraction case.

All these values depend on the dimension n. To see how they change as n increases, we computed them for various n. Figure 1 gives $L_{oc}(n)$, $SL_{oc}(n)$, $S_{oc}(n)$, and $SS_{oc}(n)$ for $3 \leq n \leq 100$. Figure 2 gives $L_{ic}(n)$, $SL_{ic}(n)$, $S_{ic}(n)$, and $SS_{ic}(n)$ for $3 \leq n \leq 100$. In figures 1 and 2, the dashed line corresponds to L, the solid line to SL, the dots to SS, and the circles to S.

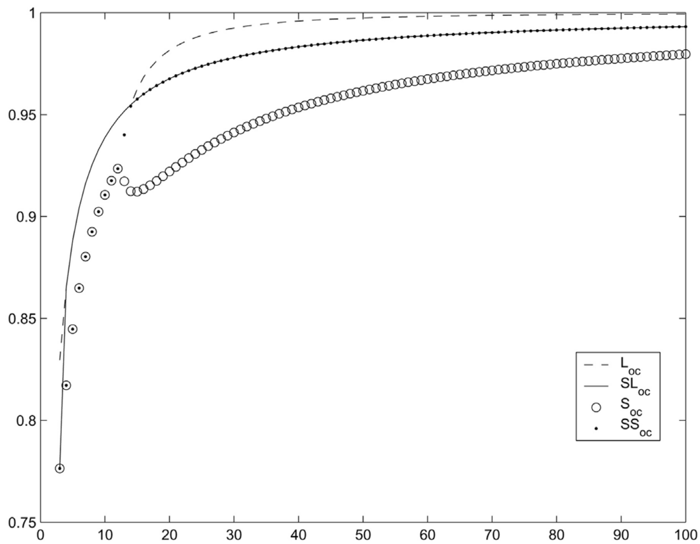

Figure 1. L_{oc}, SL_{oc}, S_{oc}, and SS_{oc} when $3 \leq n \leq 100$.

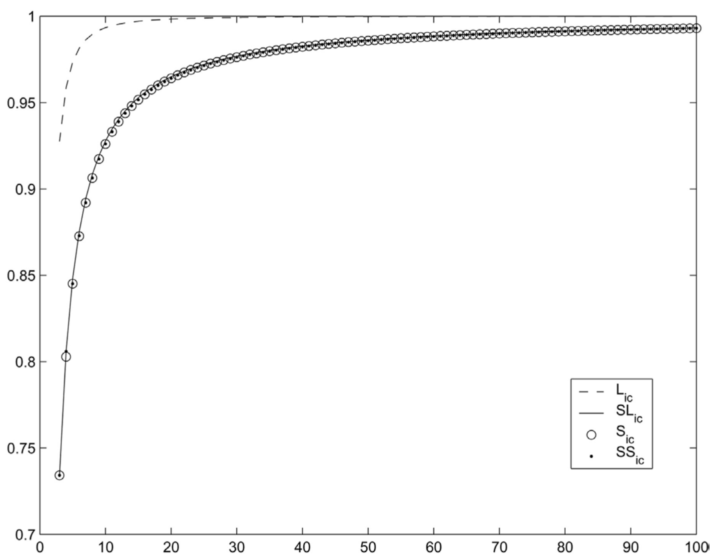

Figure 2. L_{ic}, SL_{ic}, S_{ic}, and SS_{ic} when $3 \leq n \leq 100$.

Figures 1 and 2 lead to the following observations. They are drawn from our numerical results for $3 \leq n \leq 100$. However, further numerical experiments seem to indicate that they are also true for $n > 100$.

OBSERVATION 3.8 *For* $3 \leq n \leq 100$, *we have*

1. $L_{\text{oc}}(n)$, $L_{\text{ic}}(n)$, $\text{SL}_{\text{oc}}(n)$, $\text{SL}_{\text{ic}}(n)$, *and* $S_{\text{ic}}(n)$ *are increasing functions of n.* $S_{\text{oc}}(n)$ *is increasing when* $n \geq 15$.
2. $L_{\text{oc}}(n) < L_{\text{ic}}(n) < 1$ *for* $n \geq 3$. $S_{\text{oc}}(n) < S_{\text{ic}}(n) < 1$ *for* $n \geq 5$.
3. *The characteristic values of equation* (14) *are clustered in modulus except that the one with the largest modulus* L_{oc} *and the one with the smallest modulus* S_{oc} *for* $n \geq 15$.
4. *The characteristic values of equation* (16) *are clustered in modulus except that the one with the largest modulus* L_{ic}.

From Observation 3.8, Lemma 3.2, and Theorem 3.6, if the Nelder–Mead method applies one type of step (*outside contraction* or *inside contraction*) repeatedly, then it makes less and less progress per iteration toward the minimizer as the dimension n increases. In particular, Theorem 3.6 shows that $|\mu_{\text{oc}}(n)|$ and $|\mu_{\text{ic}}(n)|$ are close to 1 as n is large (for instance, when $n = 32$, $L_{\text{oc}} \approx 0.9935$ and $L_{\text{ic}} \approx 0.9993$ are already very close to 1). If mixed steps are used, it is expected that the Nelder–Mead method deteriorates as the dimension increases. This is illustrated by the numerical experiments reported in the next section.

3.5 *Numerical results*

To study the effect of dimensionality on the Nelder–Mead method, we did some numerical experiments. We were particularly interested in how the rate of convergence of the Nelder–Mead method changes when the dimension n increases. In these tests, the rate of convergence of the Nelder–Mead method $\rho(S_0, n)$ was computed numerically for various $n \in [3, 100]$. For each such n, the Nelder–Mead method was applied to the function (7) 10 times. The method was stopped when $\sigma(S_k) \leq 10^{-8}$ is satisfied. In each time, the initial simplex was chosen as in equation (8) with

$$\mathbf{v}_i^{(0)} = 2 * \text{rand}(n, 1) - 1 \tag{32}$$

for $i = 1, 2, \ldots, n$, and the rate of convergence $\rho(S_0, n)$ was computed. In equation (32), we use the MATLAB function rand$(n, 1)$, which is a vector in $\mathbb{R}^n$ with random entries, chosen from a uniform distribution on the interval (0, 1). Therefore, each $\mathbf{v}_i^{(0)}$ in equation (32) has random entries chosen from a uniform distribution on $(-1, 1)$. We then took the mean of the 10 evaluated rates of convergence. We remark that the results obtained from individual trials are not significantly different, especially for relatively large n. For instance, the 10 computed rates range from 0.9902 to 0.9907 when $n = 30$. It is hoped that the choice of the initial simplex via equation (32) simulates the final stages of the Nelder–Mead method when applied to the function (7).

All the numerical experiments were done on a Dell Precision 420 workstation. We used the MATLAB 12.1 implementation of the Nelder–Mead method FMINSEARCH, which is based on the version given in ref. [10], except the initial simplex and the termination criterion. The results are reported in figure 3, where the solid line corresponds to the rate of convergence $\rho(S_0, n)$. In order to compare the computed $\rho(S_0, n)$ with predicted behavior of the Nelder–Mead method in section 3.4, we also plot L_{oc} (dashed line), L_{ic} (dash-dot line), SL_{oc} (dots), and SL_{ic} (circles).

We see from figure 3 that the numerical results seem to confirm the analysis in Section 3.4, i.e., as the dimension n increases, the convergence of the Nelder–Mead method becomes

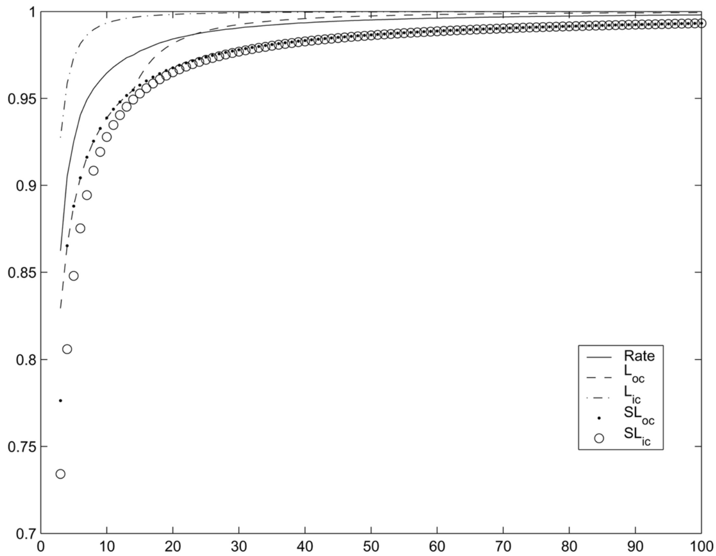

Figure 3. Rate of convergence for $3 \leq n \leq 100$.

slower. In particular, as n becomes larger and larger, the rate of convergence becomes closer and closer to 1 and therefore, the Nelder–Mead method makes less and less progress per iteration (on average) toward the minimizer. We would expect that $\lim_{n\to\infty} \rho(S_0, n) = 1$. It should be noted that when $n = 32$ which is moderately large for a derivative-free optimization method; the rate of convergence is $\rho(S_0, 32) = 0.9912$, already very close to 1.

4. Final remarks

The theoretical and numerical analysis given in Section 3 on the function (7) with the initial simplex (8) reveals an important property of the Nelder–Mead method. It makes less and less progress per iteration (on average) as the dimension n increases. In particular, it may make very tiny progress per iteration when n is large. For more difficult functions and more general initial simplices, we expect the Nelder–Mead method suffers from poorer performance in higher dimensions.

There are issues on the effect of dimensionality on the Nelder–Mead method that remain to be studied. For instance, it would be interesting to generalize the results obtained, in this paper, when a more general initial simplex is used. One difficulty is that it is still open whether the Nelder–Mead simplices converge to the minimizer for the function (7) using a general initial simplex, for $n \geq 2$. A possible approach is to use probabilistic analysis. Another intriguing question which deserves further study is that in ref. [2], it is found that the Nelder–Mead method seems to perform very poorly in certain dimensions (for instance, $n = 24$ and $n = 42$) when applied to the function (7) using certain initial simplices.

The nonconvergence examples found in ref. [11] and the inefficiency of the Nelder–Mead method in higher dimensions discovered in refs. [2,16,18,22] and this paper highlights the need to develop methods that retain the positive features of this method and remove its weakness. Several convergent variants of the Nelder–Mead method have been developed recently [see, for instance, refs. 8,15,19]. Some encouraging numerical results are reported in ref. [15].

An interesting future research direction is to develop convergent direct search methods which are efficient in higher dimensions.

Acknowledgements

We are very grateful to the referees for their insightful and constructive comments and suggestions, which helped us to improve the content and presentation of the paper. We also thank David Byatt for providing us his thesis.

Lixing Han's research was supported in part by a UM-Flint 2002 summer research fellowship. The work of this Michael Neumann was supported in part by NSF Grant No. DMS9973247.

References

[1] Barton, R.R. and Ivey, J.S., 1996, Nelder–Mead simplex modifications for simulation optimization. *Management Science*, **42**, 954–972.
[2] Byatt, D., 2000, Convergent variants of the Nelder–Mead algorithm. Master's thesis, University of Canterbury, Christchurch, New Zealand.
[3] Dennis, J.E., Jr. and Woods, D.J., 1987, Optimization on microcomputers: The Nelder–Mead simplex algorithm. In: A. Wouk (Ed.) *New Computing Environments: Microcomputers in Large-Scale Scientific Computing* (Philadelphia: SIAM), pp. 116–122.
[4] Dai, Y.H. and Fletcher, R., 2003, *On the Asymptotic Behaviour of Some New Gradient Methods*. University of Dundee Report NA/212.
[5] Han, L., 2000, Algorithms for unconstrained optimization. PhD thesis, University of Connecticut.
[6] Han, L., Neumann, M. and Xu, J., 2003, On the roots of certain polynomials arising from the analysis of the Nelder–Mead simplex method. *Linear Algebra and Applications*, **363**, 109–124.
[7] Kelley, C.T., 1999, *Iterative Methods for Optimization* (Philadelphia: SIAM Publications).
[8] Kelley, C.T., 2000, Detection and remediation of stagnation in the Nelder–Mead algorithm using a sufficient decrease condition. *SIAM Journal on Optimization*, **10**, 43–55.
[9] Kolda, T.G., Lewis, R.M. and Torczon, V., 2003, Optimization by direct search: new perspectives on some classical and modern methods. *SIAM Review*, **45**, 385–482.
[10] Lagarias, J.C., Reeds, J.A., Wright, M.H. and Wright, P., 1998, Convergence properties of the Nelder–Mead simplex algorithm in low dimensions. *SIAM Journal on Optimization*, **9**, 112–147.
[11] McKinnon, K.I.M., 1998, Convergence of the Nelder–Mead simplex method to a nonstationary point. *SIAM Journal on Optimization*, **9**, 148–158.
[12] Nelder, J.A. and Mead, R., 1965, A simplex method for function minimization. *Computer Journal*, **7**, 308–313.
[13] Parkinson, J.M. and Hutchinson, D., 1972, An investigation into the efficiency of the variants on the simplex method. In: F.A. Lootsma (Ed.) *Numerical Methods for Non-Linear Optimization* (London and New York: Academic Press), pp. 115–135.
[14] Powell, M.J.D., 1986, How bad are the BFGS and DFP methods when the objective function is quadratic? *Mathematical Programming*, **34**, 34–47.
[15] Price, C.J., Coope, I.D. and Byatt, D., 2002, A convergent variant of the Nelder–Mead algorithm. *Journal of Optimization Theory and Applications*, **113**, 5–19.
[16] Rowan, T., 1990, Functional stability analysis of numerical algorithms. PhD thesis, University of Texas at Austin.
[17] Spendley, W., Hext, G.R. and Himsworth, F.R., 1962, Sequential application of simplex designs in optimization and evolutionary operation. *Technometrics*, **4**, 441–461.
[18] Torczon, V., 1989, Multi-directional search: a direct search algorithm for parallel machines. PhD thesis, Rice University, TX.
[19] Tseng, P., 2000, Fortified-descent simplicial search method: a general approach. *SIAM Journal on Optimization*, **10**, 269–288.
[20] Walters, F.H., Parker, L.R., Morgan, S.L. and Deming, S.N., 1991, *Sequential Simplex Optimization* (Boca Raton, FL: CRC Press).
[21] Woods, D.J., 1985, An iterative approach for solving multi-objective optimization problems. PhD thesis, Rice University, TX.
[22] Wright, M.H., 1996, Direct search methods: Once scorned, now respectable. In: D.F. Griffiths and G.A. Watson (Eds.) *Numerical Analysis 1995: Proceedings of the 1995 Dundee Biennial Conference in Numerical Analysis* (Harlow, UK: Addison Wesley Longman), pp. 191–208.