Lucent Technologies
Bell Labs Innovations

Convergence Properties of the Nelder-Mead Simplex Algorithm in Low Dimensions

Jeffrey C. Lagarias
James A. Reeds
Margaret H. Wright
Paul E. Wright

Technical Report 96-4-07
Computing Sciences Research Center
Bell Laboratories
Murray Hill. New Jersey 07974

Revised Version
May 1, 1997

http://cm.bell-labs.com/cm/cs/doc/96/4-07.ps.gz

Convergence Properties of the Nelder-Mead Simplex Algorithm in Low Dimensions

Jeffrey C. Lagarias*
James A. Reeds†
Margaret H. Wright‡
Paul E. Wright§

May 1, 1997

Abstract

The Nelder-Mead simplex algorithm, first published in 1965, is an enormously popular direct search method for multidimensional unconstrained minimization. Despite its widespread use, essentially no theoretical results have been proved explicitly for the Nelder-Mead algorithm. This paper presents convergence properties of the Nelder-Mead algorithm applied to strictly convex functions in dimensions 1 and 2. We prove convergence to a minimizer for dimension 1, and various limited convergence results for dimension 2. A counterexample of McKinnon gives a family of strictly convex functions in two dimensions and a set of initial conditions for which the Nelder-Mead algorithm converges to a non-minimizer. It is not yet known whether the Nelder-Mead method can be proved to converge to a minimizer for a more specialized class of convex functions in two dimensions.

1. Introduction

Since its publication in 1965, the Nelder-Mead (NM) "simplex" algorithm [6] has become one of the most widely used methods for nonlinear unconstrained optimization. The Nelder-Mead algorithm should not be confused with the (probably) more famous simplex algorithm of Dantzig for linear programming; both algorithms employ a sequence of simplices, but are otherwise completely different and unrelated—in particular, the NM method is intended for unconstrained optimization.

The Nelder-Mead algorithm is especially popular in the fields of chemistry, chemical engineering, and medicine. The recent book [16], which contains a bibliography with thousands of references, is devoted entirely to the NM method and variations. Two measures of the ubiquity of the NM method are that it appears in the best-selling handbook *Numerical Recipes* [7], where it is called the "amoeba algorithm", and in the Matlab™ optimization toolbox [4].

The NM method attempts to minimize a scalar-valued nonlinear function of n real variables using only function values, without any derivative information (explicit or implicit). The NM method thus falls in the general class of *direct search methods*; for a discussion of these methods, see, for example, [13, 18]. A large subclass of direct search methods, including the NM method, maintain at each step a nondegenerate *simplex*, a geometric figure in n dimensions of nonzero volume that is the convex hull of $n+1$ vertices.

*AT&T Research, Murray Hill, New Jersey 07974 (jcl@research.att.com).
†AT&T Research, Murray Hill, New Jersey 07974 (reeds@research.att.com).
‡Bell Laboratories, Murray Hill, New Jersey 07974 (mhw@research.bell-labs.com).
§AT&T Research, Murray Hill, New Jersey 07974 (pew@research.att.com).

Each iteration of a simplex-based direct search method begins with a simplex, specified by its $n+1$ vertices and the associated function values. One or more test points are computed, along with their function values, and the iteration terminates with a new (different) simplex such that the function values at its vertices satisfy some form of descent condition compared to the previous simplex. Among such algorithms, the NM algorithm is particularly parsimonious in function evaluations per iteration, since in practice it typically requires only one or two function evaluations to construct a new simplex. (Several popular direct search methods use n or more function evaluations to obtain a new simplex.) There is a wide array of folklore about the NM method, mostly along the lines that it works well in "small" dimensions and breaks down in "large" dimensions, but very few careful numerical results have been published to support these perceptions. Apart from the discussion in [12], little attention has been paid to a systematic analysis of why the NM algorithm fails or breaks down numerically, as it often does.

Remarkably, there has been no published theoretical analysis explicitly treating the *original* NM algorithm in the more than 30 years since its publication. Essentially no convergence results have been proved, although in 1985 Woods [17] studied a modified[1] Nelder-Mead algorithm applied to a strictly convex function. The few known facts about the original Nelder-Mead algorithm consist mainly of negative results. Woods [17] displayed a nonconvex example in two dimensions for which the Nelder-Mead algorithm converges to a nonminimizing point. Very recently, McKinnon [5] gave a family of strictly convex functions and a starting configuration in two dimensions for which all vertices in the Nelder-Mead method converge to a nonminimizing point.

The theoretical picture for other direct search methods is much clearer. Torczon [13] proved that "pattern search" algorithms converge to a stationary point when applied to a general smooth function in n dimensions. Pattern search methods, including multidirectional search methods [12, 1], maintain uniform linear independence of the simplex edges (i.e., the dihedral angles are uniformly bounded away from zero and π), and require only simple decrease in the best function value at each iteration. Rykov [8, 9, 10] introduced several direct search methods that converge to a minimizer for strictly convex functions. In the methods proposed by Tseng [15], a "fortified descent" condition—stronger than simple descent—is required, along with uniform linear independence of the simplex edges. Depending on a user-specified parameter, Tseng's methods may involve only a small number of function evaluations at any given iteration, and are shown to converge to a stationary point for general smooth functions in n dimensions.

Published convergence analyses of simplex-based direct search methods impose one or both of the following requirements: (i) the edges of the simplex remain uniformly linearly independent at every iteration; (ii) a descent condition stronger than simple decrease is satisfied at every iteration. In general, the Nelder-Mead algorithm fails to have either of these properties; the resulting difficulties in analysis may explain the longstanding lack of convergence results.

Because the Nelder-Mead method is so widely used by practitioners to solve important optimization problems, we believe that its theoretical properties should be understood as fully as possible. This paper presents convergence results in one and two dimensions for the original Nelder-Mead algorithm applied to strictly convex functions with bounded level sets. Our approach is to consider the Nelder-Mead algorithm as a discrete dynamical system whose iterations are "driven" by the function values. Combined with strict convexity of the function, this interpretation implies restrictions on the allowed sequences of Nelder-Mead moves, from which convergence results can be derived. Our main results are:

1. In dimension 1, the Nelder-Mead algorithm converges to a minimizer (Theorem 4.1), and convergence is eventually M-step linear[2] when the reflection parameter $\rho = 1$ (Theorem 4.2).

[1] The modifications in [17] include a contraction acceptance test different from the one given in the Nelder-Mead paper and a "relative decrease" condition (stronger than simple decrease) for accepting a reflection step. Woods did not give any conditions under which the iterates converge to the minimizer.

[2] By M-step linear convergence we mean that there is an integer M, independent of the function being minimized, such that the simplex diameter is reduced by a factor no less than $1/2$ after M iterations.

2. In dimension 2, the function values at all simplex vertices in the standard Nelder-Mead algorithm converge to the same value (Theorem 5.1).

3. In dimension 2, the simplices in the standard Nelder-Mead algorithm have diameters converging to zero (Theorem 5.2).

Note that Result 3 does *not* assert that the simplices converge to a single point $\mathbf{x}_*$. No example is known in which the iterates fail to converge to a single point, but the issue is not settled.

For the case of dimension 1, V. Torczon [14] has recently informed us that some convergence results for the original Nelder-Mead algorithm can be deduced from the results in [13]; see Section 4.4. For dimension 2, our results may appear weak, but the McKinnon example [5] shows that convergence to a minimizer is not guaranteed for general strictly convex functions in dimension 2. Because the smoothest McKinnon example has a point of discontinuity in the fourth derivatives, a logical question is whether or not the NM method converges to a minimizer in two dimensions for a more specialized class of strictly convex functions—in particular, for smooth functions. This remains a challenging open problem. At present there is no function in any dimension greater than 1 for which the original Nelder-Mead algorithm has been proved to converge to a minimizer.

Given all the known inefficiencies and failures of the Nelder-Mead algorithm (see, for example, [12]), one might wonder why it is used *at all*, let alone why it is so extraordinarily popular. We offer three answers. First, in many applications, for example in industrial process control, one simply wants to find parameter values that improve some performance measure; the Nelder-Mead algorithm typically produces significant improvement for the first few iterations. Second, there are important applications where a function evaluation is enormously expensive or time-consuming, but derivatives cannot be calculated. In such problems, a method that requires at least n function evaluations at every iteration (which would be the case if using finite-difference gradient approximations or one of the more popular pattern search methods) is too expensive or too slow. When it succeeds, the NM method tends to require substantially fewer function evaluations than these alternatives, and its relative "best-case efficiency" often outweighs the lack of convergence theory. Third, the NM method is appealing because its steps are easy to explain and simple to program.

In light of weaknesses exposed by the McKinnon counterexample and the analysis here, future work involves developing methods that retain the good features of the Nelder-Mead method, but are more reliable and efficient in theory and practice; see, for example, [2].

The contents of this paper are as follows. Section 2 describes the Nelder-Mead algorithm, and Section 3 gives its general properties. For a strictly convex function with bounded level sets, Section 4 analyzes the Nelder-Mead method in one dimension, and Section 5 presents limited convergence results for the standard Nelder-Mead algorithm in two dimensions. Finally, Section 6 discusses open problems.

2. The Nelder-Mead Algorithm

The Nelder-Mead algorithm [6] was proposed as a method for minimizing a real-valued function $f(\mathbf{x})$ for $\mathbf{x} \in \mathcal{R}^n$. Four scalar parameters must be specified to define a complete Nelder-Mead method: coefficients of *reflection* (ρ), *expansion* (χ), *contraction* (γ), and *shrinkage* (σ). According to the original Nelder-Mead paper, these parameters should satisfy

$$\rho > 0, \quad \chi > 1, \quad \chi > \rho, \quad 0 < \gamma < 1, \quad \text{and} \quad 0 < \sigma < 1. \tag{2.1}$$

(The relation $\chi > \rho$, while not stated explicitly in the original paper, is implicit in the algorithm description and terminology.) The nearly universal choices used in the *standard* NM algorithm are

$$\rho = 1, \quad \chi = 2, \quad \gamma = \tfrac{1}{2}, \quad \text{and} \quad \sigma = \tfrac{1}{2}. \tag{2.2}$$

We assume the general conditions (2.1) for the one-dimensional case, but restrict ourselves to the standard case (2.2) in the two-dimensional analysis.

2.1. Statement of the algorithm

At the beginning of the kth iteration, $k \geq 0$, a nondegenerate simplex Δ_k is given, along with its $n+1$ vertices, each of which is a point in $\mathcal{R}^n$. It is always assumed that iteration k begins by ordering and labeling these vertices as $\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_{n+1}^{(k)}$, such that

$$f_1^{(k)} \leq f_2^{(k)} \leq \cdots \leq f_{n+1}^{(k)}, \tag{2.3}$$

where $f_i^{(k)}$ denotes $f(\mathbf{x}_i^{(k)})$. The kth iteration generates a set of $n+1$ vertices that define a different simplex for the next iteration, so that $\Delta_{k+1} \neq \Delta_k$. Because we seek to minimize f, we refer to $\mathbf{x}_1^{(k)}$ as the *best* point or vertex, to $\mathbf{x}_{n+1}^{(k)}$ as the *worst* point, and to $\mathbf{x}_n^{(k)}$ as the *next-worst* point. Similarly, we refer to $f_{n+1}^{(k)}$ as the worst function value, and so on.

The 1965 paper [6] contains several ambiguities about strictness of inequalities and tie-breaking that have led to differences in interpretation of the Nelder-Mead algorithm. What we shall call "the" Nelder-Mead algorithm (Algorithm NM) includes well-defined tie-breaking rules, given below, and accepts the better of the reflected and expanded points in Step 3 (see the discussion in Section 3.1 about property 4 of the NM method).

A single generic iteration is specified, omitting the superscript k to avoid clutter. The result of each iteration is either: (1) a single new vertex—the *accepted point*—which replaces $\mathbf{x}_{n+1}$ in the set of vertices for the next iteration; or (2) if a shrink is performed, a set of n new points that, together with $\mathbf{x}_1$, form the simplex at the next iteration.

One iteration of Algorithm NM (the Nelder-Mead algorithm).

1. **Order.**
   Order the $n+1$ vertices to satisfy $f(\mathbf{x}_1) \leq f(\mathbf{x}_2) \leq \cdots \leq f(\mathbf{x}_{n+1})$, using the tie-breaking rules given below.
2. **Reflect.**
   Compute the *reflection point* $\mathbf{x}_r$ from
   $$\mathbf{x}_r = \bar{\mathbf{x}} + \rho(\bar{\mathbf{x}} - \mathbf{x}_{n+1}) = (1+\rho)\bar{\mathbf{x}} - \rho\mathbf{x}_{n+1}, \tag{2.4}$$
   where $\bar{\mathbf{x}} = \sum_{i=1}^{n} \mathbf{x}_i/n$ is the centroid of the n best points (all vertices except for $\mathbf{x}_{n+1}$). Evaluate $f_r = f(\mathbf{x}_r)$.
   If $f_1 \leq f_r < f_n$, accept the reflected point $\mathbf{x}_r$ and terminate the iteration.
3. **Expand.**
   If $f_r < f_1$, calculate the *expansion point* $\mathbf{x}_e$:
   $$\mathbf{x}_e = \bar{\mathbf{x}} + \chi(\mathbf{x}_r - \bar{\mathbf{x}}) = \bar{\mathbf{x}} + \rho\chi(\bar{\mathbf{x}} - \mathbf{x}_{n+1}) = (1+\rho\chi)\bar{\mathbf{x}} - \rho\chi\mathbf{x}_{n+1}, \tag{2.5}$$
   and evaluate $f_e = f(\mathbf{x}_e)$. If $f_e < f_r$, accept $\mathbf{x}_e$ and terminate the iteration; otherwise (if $f_e \geq f_r$), accept $\mathbf{x}_r$ and terminate the iteration.
4. **Contract.**
   If $f_r \geq f_n$, perform a *contraction* between $\bar{\mathbf{x}}$ and the better of $\mathbf{x}_{n+1}$ and $\mathbf{x}_r$.
   a. Outside. If $f_n \leq f_r < f_{n+1}$ (i.e., $\mathbf{x}_r$ is strictly better than $\mathbf{x}_{n+1}$), perform an *outside contraction*: calculate
   $$\mathbf{x}_c = \bar{\mathbf{x}} + \gamma(\mathbf{x}_r - \bar{\mathbf{x}}) = \bar{\mathbf{x}} + \gamma\rho(\bar{\mathbf{x}} - \mathbf{x}_{n+1}) = (1+\rho\gamma)\bar{\mathbf{x}} - \rho\gamma\mathbf{x}_{n+1}, \tag{2.6}$$
   and evaluate $f_c = f(\mathbf{x}_c)$. If $f_c \leq f_r$, accept $\mathbf{x}_c$ and terminate the iteration; otherwise, go to Step 5 (perform a shrink).
   b. Inside. If $f_r \geq f_{n+1}$, perform an *inside contraction*: calculate
   $$\mathbf{x}_{cc} = \bar{\mathbf{x}} - \gamma(\bar{\mathbf{x}} - \mathbf{x}_{n+1}) = (1-\gamma)\bar{\mathbf{x}} + \gamma\mathbf{x}_{n+1}, \tag{2.7}$$
   and evaluate $f_{cc} = f(\mathbf{x}_{cc})$. If $f_{cc} < f_{n+1}$, accept $\mathbf{x}_{cc}$ and terminate the iteration; otherwise, go to Step 5 (perform a shrink).

5. **Perform a shrink step.**
Evaluate f at the n points $\mathbf{v}_i = \mathbf{x}_1 + \sigma(\mathbf{x}_i - \mathbf{x}_1)$, $i = 2, \ldots, n+1$. The (unordered) vertices of the simplex at the next iteration consist of $\mathbf{x}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n+1}$.

Figures 1 and 2 show the effects of reflection, expansion, contraction and shrinkage for a simplex in two dimensions (a triangle), using the standard coefficients $\rho = 1$, $\chi = 2$, $\gamma = \frac{1}{2}$, and $\sigma = \frac{1}{2}$. Observe that, except in a shrink, the one new vertex always lies on the (extended) line joining $\bar{x}$ and x_{n+1}. Furthermore, it is visually evident that the simplex shape undergoes a noticeable change during an expansion or contraction with the standard coefficients.

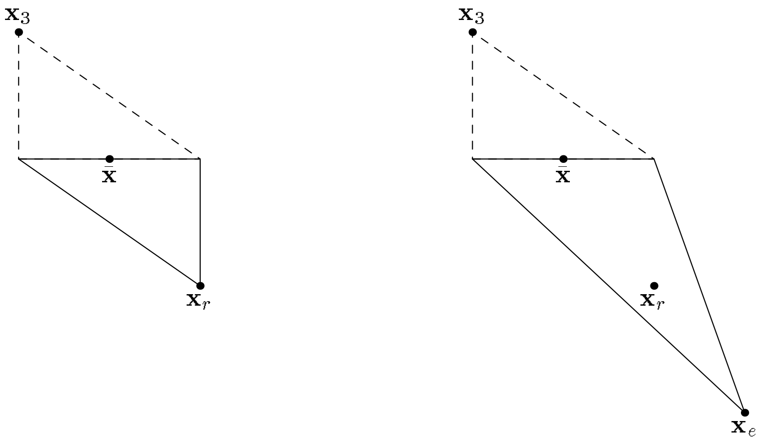


Figure 1: Nelder-Mead simplices after a reflection and an expansion step. The original simplex is shown with a dashed line.

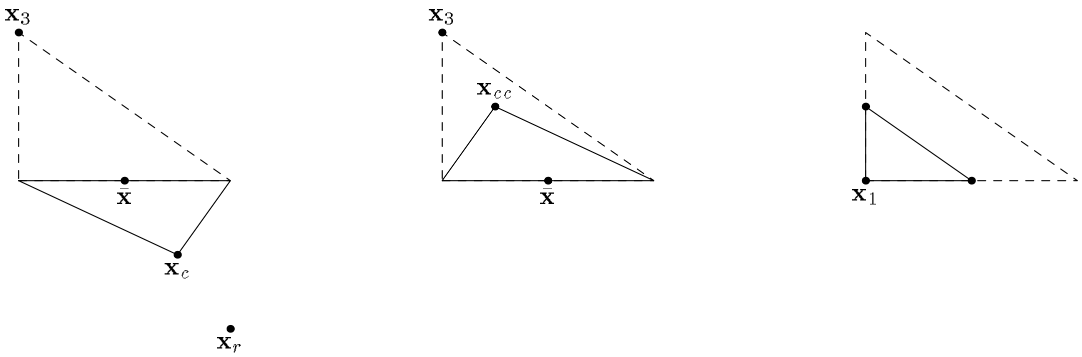


Figure 2: Nelder-Mead simplices after an outside contraction, an inside contraction, and a shrink. The original simplex is shown with a dashed line.

The Nelder-Mead paper [6] did not describe how to order points in the case of equal function values. We adopt the following tie-breaking rules, which assign to the new vertex the highest possible index consistent with the relation $f(\mathbf{x}_1^{(k+1)}) \le f(\mathbf{x}_2^{(k+1)}) \le \cdots \le f(\mathbf{x}_{n+1}^{(k+1)})$.

Non-shrink ordering rule. When a non-shrink step occurs, the worst vertex $\mathbf{x}_{n+1}^{(k)}$ is discarded. The accepted point created during iteration k, denoted by $\mathbf{v}^{(k)}$, becomes a new vertex and takes position $j+1$ in the vertices of Δ_{k+1}, where

$$j = \max_{0 \le \ell \le n} \{\, \ell \mid f(\mathbf{v}^{(k)}) < f(\mathbf{x}_{\ell+1}^{(k)}) \,\};$$

all other vertices retain their relative ordering from iteration k.

Shrink ordering rule. If a shrink step occurs, the only vertex carried over from Δ_k to Δ_{k+1} is $\mathbf{x}_1^{(k)}$. Only one tie-breaking rule is specified, for the case in which $\mathbf{x}_1^{(k)}$ and one or more of the new points are tied as the best point: if $\min\{f(\mathbf{v}_2^{(k)}),\ldots,f(\mathbf{v}_{n+1}^{(k)})\} = f(\mathbf{x}_1^{(k)})$, then $\mathbf{x}_1^{(k+1)} = \mathbf{x}_1^{(k)}$. Beyond this, whatever rule is used to define the original ordering may be applied after a shrink.

We define the *change index* k^* of iteration k as the smallest index of a vertex that differs between iterations k and $k+1$:

$$k^* = \min\{\ i \ \mid \ \mathbf{x}_i^{(k)} \neq \mathbf{x}_i^{(k+1)}\ \}. \tag{2.8}$$

(Tie-breaking rules are needed to define a unique value of k^*.) When Algorithm NM terminates in Step 2, $1 < k^* \le n$; with termination in Step 3, $k^* = 1$; with termination in Step 4, $1 \le k^* \le n+1$; and with termination in Step 5, $k^* = 1$ or 2. A statement that "$\mathbf{x}_j$ changes" means that j is the change index at the relevant iteration.

The rules and definitions given so far imply that, for a non-shrink iteration,

$$\begin{aligned} &f_j^{(k+1)} = f_j^{(k)} \quad \text{and} \quad \mathbf{x}_j^{(k+1)} = \mathbf{x}_j^{(k)}, \quad j < k^*;\\ &f_{k^*}^{(k+1)} < f_{k^*}^{(k)} \quad \text{and} \quad \mathbf{x}_{k^*}^{(k+1)} \neq \mathbf{x}_{k^*}^{(k)};\\ &f_j^{(k+1)} = f_{j-1}^{(k)} \quad \text{and} \quad \mathbf{x}_j^{(k+1)} = \mathbf{x}_{j-1}^{(k)} \quad j > k^*. \end{aligned} \tag{2.9}$$

Thus the vector $(f_1^{(k)},\ldots,f_{n+1}^{(k)})$ strictly lexicographically decreases at each non-shrink iteration.

For illustration, suppose that $n = 4$ and the vertex function values at a non-shrink iteration k are $(1,2,2,3,3)$. If $f(\mathbf{v}^{(k)}) = 2$, the function values at iteration $k+1$ are $(1,2,2,2,3)$, $\mathbf{x}_4^{(k+1)} = \mathbf{v}^{(k)}$, and $k^* = 4$. This example shows that, following a single non-shrink iteration, the worst function value need not strictly decrease; however, the worst function value must strictly decrease after at most $n+1$ consecutive non-shrink iterations.

2.2. Matrix notation

It is convenient to use matrix notation to describe Nelder-Mead iterations. The simplex Δ_k can be represented as an $n \times (n+1)$ matrix whose columns are the vertices:

$$\Delta_k = \begin{pmatrix} \mathbf{x}_1^{(k)} & \cdots & \mathbf{x}_{n+1}^{(k)} \end{pmatrix} = \begin{pmatrix} B_k & \mathbf{x}_{n+1}^{(k)} \end{pmatrix}, \quad \text{where} \quad B_k = \begin{pmatrix} \mathbf{x}_1^{(k)} & \cdots & \mathbf{x}_n^{(k)} \end{pmatrix}.$$

For any simplex Δ_k in $\mathcal{R}^n$, we define M_k as the $n \times n$ matrix whose jth column represents the "edge" of Δ_k between $\mathbf{x}_j^{(k)}$ and $\mathbf{x}_{n+1}^{(k)}$:

$$M_k \equiv \begin{pmatrix} \mathbf{x}_1^{(k)} - \mathbf{x}_{n+1}^{(k)} & \mathbf{x}_2^{(k)} - \mathbf{x}_{n+1}^{(k)} & \cdots & \mathbf{x}_n^{(k)} - \mathbf{x}_{n+1}^{(k)} \end{pmatrix} = B_k - \mathbf{x}_{n+1}^{(k)}\mathbf{e}^T, \tag{2.10}$$

where $\mathbf{e} = (1,1,\ldots,1)^T$. The n-dimensional volume of Δ_k is given by

$$\operatorname{vol}(\Delta_k) = \frac{|\det(M_k)|}{n!}.$$

A simplex Δ_k is *nondegenerate* if M_k is nonsingular, or, equivalently, if $\operatorname{vol}(\Delta_k) > 0$. The volume of the simplex obviously depends only on the coordinates of the vertices, not on their ordering. For future reference, we define the diameter of Δ_k as

$$\operatorname{diam}(\Delta_k) = \max_{i \neq j} \|\mathbf{x}_i^{(k)} - \mathbf{x}_j^{(k)}\|,$$

where $\|\cdot\|$ denotes the two-norm.

During a non-shrink iteration, the function is evaluated only at *trial points* of the form

$$\mathbf{z}^{(k)}(\tau) := \bar{\mathbf{x}}^{(k)} + \tau(\bar{\mathbf{x}}^{(k)} - \mathbf{x}_{n+1}^{(k)}) = (1+\tau)\bar{\mathbf{x}}^{(k)} - \tau\mathbf{x}_{n+1}^{(k)}, \tag{2.11}$$

where the coefficient τ has one of four possible values:

$$\begin{aligned} &\tau = \rho \quad \text{(reflection)}; && \tau = \rho\chi \quad \text{(expansion)};\\ &\tau = \rho\gamma \quad \text{(outside contraction)}; && \tau = -\gamma \quad \text{(inside contraction)}. \end{aligned} \tag{2.12}$$

In a non-shrink step, the single accepted point is one of the trial points, and we let τ_k denote the coefficient associated with the accepted point at iteration k. Thus the new vertex $\mathbf{v}^{(k)}$ produced during iteration k, which will replace $\mathbf{x}_{n+1}^{(k)}$, is given by $\mathbf{v}^{(k)} = \mathbf{z}^{(k)}(\tau_k)$. We sometimes call τ_k the *type* of move for a non-shrink iteration k. During the kth Nelder-Mead iteration, (2.11) shows that each trial point (reflection, expansion, contraction) may be written as

$$\mathbf{z}^{(k)}(\tau) = \Delta_k \mathbf{t}(\tau), \quad \text{where} \quad \mathbf{t}(\tau) = \big(\frac{1+\tau}{n}, \;\ldots,\; \frac{1+\tau}{n}, -\tau\big)^T. \tag{2.13}$$

Following the kth Nelder-Mead iteration, the (unordered) vertices of the next simplex are the columns of $\Delta_k S_k$, where S_k is an $(n+1)\times(n+1)$ matrix given by

$$\begin{pmatrix} I_n & \dfrac{(1+\tau_k)}{n}\mathbf{e} \\ \mathbf{0}^T & -\tau_k \end{pmatrix} \text{ for a step of type } \tau, \text{ and by } \begin{pmatrix} 1 & (1-\sigma)\mathbf{e}^T \\ \mathbf{0} & \sigma I_n \end{pmatrix} \text{ for a shrink step,}$$

where $\mathbf{0}$ is an n-dimensional zero column and I_n is the n-dimensional identity matrix. After being ordered at the start of iteration $k+1$, the vertices of Δ_{k+1} satisfy

$$\Delta_{k+1} = \Delta_k T_k, \quad \text{with} \quad T_k = S_k P_k, \tag{2.14}$$

where P_k is a permutation matrix chosen to enforce the ordering and tie-breaking rules (so that P_k depends on the function values at the vertices).

The updated simplex Δ_{k+1} has a disjoint interior from Δ_k for a reflection, an expansion, or an outside contraction, while $\Delta_{k+1} \subseteq \Delta_k$ for an inside contraction or a shrink.

By the *shape* of a nondegenerate simplex, we mean its equivalence class under similarity, i.e., Δ and $\lambda\Delta$ have the same shape when $\lambda > 0$. The shape of a simplex is determined by its angles, or equivalently by the singular values of the associated matrix M (2.10) after scaling so that Δ has unit volume. The Nelder-Mead method was deliberately designed with the idea that the simplex shapes would "adapt to the features of the local landscape" [6]. The Nelder-Mead moves apparently permit any simplex shape to be approximated—in particular, arbitrarily flat or needle-shaped simplices (as in the McKinnon examples [5]) are possible.

3. Properties of the Nelder-Mead Algorithm

This section establishes various basic properties of the Nelder-Mead method. Although there is a substantial level of folklore about the Nelder-Mead method, almost no proofs have appeared in print, so we include details here.

3.1. General results

The following properties follow immediately from the definition of Algorithm NM.

1. A Nelder-Mead iteration requires one function evaluation when the iteration terminates in Step 2, two function evaluations when termination occurs in Step 3 or Step 4, and $n+2$ function evaluations if a shrink step occurs.

2. The "reflect" step is so named because the reflection point $\mathbf{x}_r$ (2.4) is a (scaled) reflection of the worst point $\mathbf{x}_{n+1}$ around the point $\bar{\mathbf{x}}$ on the line through $\mathbf{x}_{n+1}$ and $\bar{\mathbf{x}}$. It is a genuine reflection on this line when $\rho = 1$, which is the standard choice for the reflection coefficient.

3. For general functions, a shrink step can conceivably lead to an increase in every vertex function value except f_1, i.e., it is possible that $f_i^{(k+1)} > f_i^{(k)}$ for $2 \le i \le n+1$. In addition, observe that with an outside contraction (case 4a), the algorithm takes a shrink step if $f(\mathbf{x}_c) > f(\mathbf{x}_r)$, even though a new point $\mathbf{x}_r$ has already been found that strictly improves over the worst vertex, since $f(\mathbf{x}_r) < f(\mathbf{x}_{n+1})$.

4. In the expand step, the method in the original Nelder-Mead paper accepts $\mathbf{x}_e$ if $f(\mathbf{x}_e) < f_1$, and accepts $\mathbf{x}_r$ otherwise. Standard practice today (which we follow) accepts the better of $\mathbf{x}_r$ and $\mathbf{x}_e$ if both give an improvement over $\mathbf{x}_1$. The proofs of Lemmas 4.6 and 5.2 depend on the rule that the expansion point is accepted only if it is strictly better than the reflection point.

It is commonly (and correctly) assumed that nondegeneracy of the initial simplex Δ_0 implies nondegeneracy of all subsequent Nelder-Mead simplices. We first give an informal indication of why this property holds. By construction, each trial point (2.11) in the Nelder-Mead method lies strictly outside the face defined by the n best vertices, along the line joining the worst vertex to the centroid of that face. If a non-shrink iteration occurs, the worst vertex is replaced by one of the trial points. If a shrink iteration occurs, each current vertex except the best is replaced by a point that lies a fraction of the step to the current best vertex. In either case it is clear from the geometry that the new simplex must be nondegenerate. For completeness, we present a proof of nondegeneracy based on a useful result about the volumes of successive simplices.

Lemma 3.1. (Volume and nondegeneracy of Nelder Mead simplices.)

(1) If the initial simplex Δ_0 is nondegenerate, so are all subsequent Nelder-Mead simplices.

(2) Following a non-shrink step of type τ, $\mathrm{vol}(\Delta_{k+1}) = |\tau|\,\mathrm{vol}(\Delta_k)$.

(3) Following a shrink step at iteration k, $\mathrm{vol}(\Delta_{k+1}) = \sigma^n\,\mathrm{vol}(\Delta_k)$.

Proof. A simplex Δ is nondegenerate if it has nonzero volume, i.e., when $\det(\Delta) \neq 0$. Result (1) will follow immediately from (2) and (3) because $\tau \neq 0$ (see (2.12)) and $\sigma \neq 0$.

When iteration k is a non-shrink, we assume without loss of generality that the worst point is the origin. In this case, it follows from (2.13) that the new vertex is

$$\mathbf{v}^{(k)} = M_k \mathbf{w}, \quad \text{where} \quad \mathbf{w} = \Big(\frac{1+\tau}{n}, \dots, \frac{1+\tau}{n}\Big)^T, \tag{3.1}$$

so that the vertices of Δ_{k+1} consist of the vector $M_k\mathbf{w}$ and the columns of M_k. Since the volume of the new simplex does not depend on the ordering of the vertices, we may assume without affecting the volume that the new vertex is the worst. Applying the form of M in (2.10), we have

$$|\det(M_{k+1})| = |\det(M_k - M_k\mathbf{w}\mathbf{e}^T)| = |\det(M_k)|\,|\det(I - \mathbf{w}\mathbf{e}^T)|.$$

The matrix $I - \mathbf{w}\mathbf{e}^T$ has $n-1$ eigenvalues of unity and one eigenvalue equal to $1 - \mathbf{w}^T\mathbf{e} = -\tau$, so that $\det(I - \mathbf{w}\mathbf{e}^T) = -\tau$. This gives result (2).

If iteration k is a shrink step, each edge of the simplex is multiplied by σ. Thus M_{k+1} is a permutation of σM_k and result (3) for a shrink follows from a standard property of determinants for $n \times n$ matrices. ∎

Lemma 3.1 shows that, in any dimension, a reflection step with $\rho = 1$ preserves volume. The choice $\rho = 1$ is natural geometrically, since a reflection step is then a genuine reflection. A reflected simplex with $\rho = 1$ is necessarily congruent to the original simplex for $n = 1$ and $n = 2$, but this is no longer true for $n \geq 3$.

Note that, although the Nelder-Mead simplices are nondegenerate in exact arithmetic, there is in general no upper bound on $\mathrm{cond}(M_k)$. In fact, the algorithm permits $\mathrm{cond}(M_k)$ to become arbitrarily large, as it does in the McKinnon example [5].

Our next result involves affine-invariance of the Nelder-Mead method when both the simplex and function are transformed appropriately.

Lemma 3.2. (Affine-invariance.) *The Nelder-Mead algorithm is invariant under affine motions of $\mathcal{R}^n$, i.e. under a change of variables $\phi(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ in which A is invertible, in the following sense: when minimizing $f(\mathbf{x})$ starting with simplex Δ_0, the complete sequence of Nelder-Mead steps and function values is the same as when minimizing the function $\tilde{f}(\mathbf{z}) = f(\phi(\mathbf{z}))$ with initial simplex $\tilde{\Delta}_0$ defined by*

$$\tilde{\Delta}_0 = \phi^{-1}(\Delta_0) = A^{-1}(\Delta_0) - A^{-1}\mathbf{b}.$$

Proof. At the vertices of $\tilde{\Delta}_0$, $\tilde{f}(\tilde{\mathbf{x}}_i^{(0)}) = f(\mathbf{x}_i^{(0)})$. We proceed by induction, assuming for simplicity that $\mathbf{b} = 0$. If $\tilde{\Delta}_k = A^{-1}\Delta_k$ and $\tilde{f}(\tilde{\mathbf{x}}_i^{(k)}) = f(\mathbf{x}_i^{(k)})$ for $1 \le i \le n+1$, then relation (2.13) shows that the trial points generated from $\tilde{\Delta}_k$ satisfy $\tilde{\mathbf{z}}(\tau) = A^{-1}\mathbf{z}(\tau)$, which means that $\tilde{f}(\tilde{\mathbf{z}}(\tau)) = f(\mathbf{z}(\tau))$. The matrix T_k of (2.14) will therefore be the same for both Δ_k and $\tilde{\Delta}_k$, so that $\tilde{\Delta}_{k+1} = A^{-1}\Delta_{k+1}$. It follows that $\tilde{f}(\tilde{\mathbf{x}}_i^{(k+1)}) = f(\mathbf{x}_i^{(k+1)})$ for $1 \le i \le n+1$, which completes the induction. A similar argument applies when $\mathbf{b} \ne 0$. ∎

Using Lemma 3.2, we can reduce the study of the Nelder-Mead algorithm for a general strictly convex quadratic function on $\mathcal{R}^n$ to the study of $f(\mathbf{x}) = \|\mathbf{x}\|^2 = x_1^2 + \cdots + x_n^2$.

The next lemma summarizes several straightforward results.

Lemma 3.3. *Let f be a function that is bounded below on $\mathcal{R}^n$. When the Nelder-Mead algorithm is applied to minimize f, starting with a nondegenerate simplex Δ_0, then:*

(1) The sequence $\{f_1^{(k)}\}$ always converges.

(2) At every non-shrink iteration k, $f_i^{(k+1)} \le f_i^{(k)}$ for $1 \le i \le n+1$, with strict inequality for at least one value of i.

(3) If there are only a finite number of shrink iterations, then

- *(i) Each sequence $\{f_i^{(k)}\}$ converges as $k \to \infty$ for $1 \le i \le n+1$;*
- *(ii) $f_i^* \le f_i^{(k)}$ for $1 \le i \le n+1$ and all k, where $f_i^* = \lim_{k\to\infty} f_i^{(k)}$;*
- *(iii) $f_1^* \le f_2^* \le \cdots \le f_{n+1}^*$.*

(4) If there are only a finite number of non-shrink iterations, then all simplex vertices converge to a single point. ∎

We now analyze the Nelder-Mead algorithm in the case when *only non-shrink steps occur*. Torczon [12] observes that shrink steps essentially never happen in practice (she reports only 33 shrink steps in 2.9 million Nelder-Mead iterations on a set of general test problems), and the rarity of shrink steps is confirmed by our own numerical experiments. We show in Lemma 3.5 that no shrink steps are taken when the Nelder-Mead method is applied to a strictly convex function. All of our results that assume no shrink steps can obviously be applied to cases when only a finite number of shrink steps occur.

Assuming that there are no shrink steps, the next lemma gives an important property of the $n+1$ limiting vertex function values whose existence is verified in part (3) of Lemma 3.3.

Lemma 3.4. (Broken convergence.) *Suppose that f is bounded below on $\mathcal{R}^n$, that the Nelder-Mead algorithm is applied to f beginning with a nondegenerate initial simplex Δ_0, and that no shrink steps occur. If there is an integer j, $1 \le j \le n$, for which*

$$f_j^* < f_{j+1}^*, \quad \text{where} \quad f_j^* = \lim_{k\to\infty} f_j^{(k)}, \tag{3.2}$$

then there is an iteration index K such that for all $k \ge K$, the change index satisfies

$$k^* > j, \tag{3.3}$$

i.e., the first j vertices of all simplices remain fixed *after iteration K. (We refer to property (3.2) as* broken convergence *for vertex j.)*

Proof. The lemma is proved by contradiction. By hypothesis (3.2), $f_j^* + \delta = f_{j+1}^*$ for some $\delta > 0$. Pick $\epsilon > 0$ such that $\delta - \epsilon > 0$. Since $f_j^* = \lim_{k\to\infty} f_j^{(k)}$, there exists K such that for all $k \ge K$, $f_j^{(k)} - \epsilon \le f_j^*$. Then, for all $k \ge K$,

$$f_j^{(k)} < f_j^{(k)} - \epsilon + \delta \le f_j^* + \delta = f_{j+1}^*.$$

But, from Lemma 3.3, part (3), for any index ℓ, $f_{j+1}^* \le f_{j+1}^{(\ell)}$. Therefore, for all $k \ge K$ and any ℓ,

$$f_j^{(k)} < f_{j+1}^{(\ell)}. \tag{3.4}$$

But, if $k^* \leq j$ for any $k \geq K$, then, using the third relation in (2.9), it must be true that $f_{j+1}^{(k+1)} = f_j^{(k)}$, which contradicts (3.4). Thus $k^* > j$ for all $k \geq K$. ∎

The following corollary is an immediate consequence of Lemma 3.4.

Corollary 3.1. *Assume that f is bounded below on $\mathcal{R}^n$, the Nelder-Mead algorithm is applied beginning with a nondegenerate initial simplex Δ_0, and no shrink steps occur. If the change index is 1 infinitely often, i.e., the best point changes infinitely many times, then $f_1^* = \cdots = f_{n+1}^*$.* ∎

3.2. Results for strictly convex functions

Without further assumptions, very little more can be said about the Nelder-Mead algorithm, and we henceforth assume that f is strictly convex.

Definition 3.1. (Strict convexity.) *The function f is strictly convex on $\mathcal{R}^n$ if, for every pair of points $\mathbf{y}$, $\mathbf{z}$ with $\mathbf{y} \neq \mathbf{z}$, and every λ satisfying $0 < \lambda < 1$,*

$$f(\lambda \mathbf{y} + (1-\lambda)\mathbf{z}) \;<\; \lambda f(\mathbf{y}) + (1-\lambda) f(\mathbf{z}). \tag{3.5}$$

When f is strictly convex on $\mathcal{R}^n$ and

$$\mathbf{c} = \sum_{i=1}^{\ell} \lambda_i \mathbf{z}_i, \qquad \text{with} \quad 0 < \lambda_i < 1 \;\text{ and }\; \sum_{i=1}^{\ell} \lambda_i = 1,$$

$$\text{then}\;\; f(\mathbf{c}) < \sum_{i=1}^{\ell} \lambda_i f(\mathbf{z}_i) \qquad \text{and hence}\;\; f(\mathbf{c}) < \max\{f(\mathbf{z}_1), \ldots, f(\mathbf{z}_\ell)\}. \tag{3.6}$$

We now use this property to show that, when the Nelder-Mead method is applied to a strictly convex function, shrink steps cannot occur. (This result is mentioned without proof in [12].)

Lemma 3.5. *Assume that f is strictly convex on $\mathcal{R}^n$, and that the Nelder-Mead algorithm is applied to f beginning with a nondegenerate initial simplex Δ_0. Then no shrink steps will be taken.*

Proof. Shrink steps can occur only if the algorithm reaches Step 4 of Algorithm NM and fails to accept the relevant contraction point. When $n = 1$, $f(\bar{\mathbf{x}}) = f_n$. When $n > 1$, application of (3.6) to $\mathbf{x}_1, \ldots, \mathbf{x}_n$ shows that $f(\bar{\mathbf{x}}) < f_n$.

Consider an outside contraction, which is tried if $f_n \leq f_r < f_{n+1}$. Since the contraction coefficient γ satisfies $0 < \gamma < 1$, $\mathbf{x}_c$ as defined by (2.6) is a convex combination of $\bar{\mathbf{x}}$ and the reflection point $\mathbf{x}_r$. Thus, by (3.6),

$$f(\mathbf{x}_c) < \max\{f(\bar{\mathbf{x}}), f_r\}.$$

We know that $f(\bar{\mathbf{x}}) \leq f_n$ and $f_n \leq f_r$, so that $\max\{f(\bar{\mathbf{x}}), f_r\} = f_r$. Hence $f(\mathbf{x}_c) < f_r$, $\mathbf{x}_c$ will be accepted, and a shrink step will not be taken.

A similar argument applies for an inside contraction, since $f_{n+1} \leq f_r$ and $\mathbf{x}_{cc}$ is a convex combination of $\bar{\mathbf{x}}$ and $\mathbf{x}_{n+1}$. ∎

Note that simple convexity of f (for example, f constant) is not sufficient for this result, which depends in the case of an inside contraction on the fact that $f(\mathbf{x}_{cc})$ is *strictly* less than $f(\mathbf{x}_{n+1})$.

By combining the definition of a Nelder-Mead iteration, Lemma 3.4, and a mild further restriction on the reflection and contraction coefficients, we next prove that the limiting worst and next-worst function values are the same. (For $n = 1$, the result holds without the additional restriction; see Lemma 4.4).

Lemma 3.6. *Assume that f is strictly convex on $\mathcal{R}^n$ and bounded below. If, in addition to the properties $\rho > 0$, and $0 < \gamma < 1$, the reflection coefficient ρ and the contraction coefficient γ satisfy $\rho\gamma < 1$, then:*

(1) $f_n^* = f_{n+1}^*$; and

(2) There are infinitely many iterations for which $\mathbf{x}_n^{(k+1)} \neq \mathbf{x}_n^{(k)}$.

Proof. The proof is by contradiction. Assume that $f_n^* < f_{n+1}^*$. From Lemma 3.4, this means that there exists an iteration index K such that the change index $k^* = n+1$ for $k \geq K$. Without loss of generality, we may take $K = 0$. Since $k^* = n+1$ for all $k \geq 0$, the best n vertices, which must be distinct, remain constant for all iterations; thus the centroid $\bar{\mathbf{x}}^{(k)} = \bar{\mathbf{x}}$, a constant vector, and $f(\mathbf{x}_n)$ is equal to its limiting value f_n^*. Because f is strictly convex, $f(\bar{\mathbf{x}}) \leq f(\mathbf{x}_n) = f_n^*$. (This inequality is strict if $n > 1$.)

The change index will be $n+1$ at every iteration only if a contraction point is accepted and becomes the new worst point. Therefore, the vertex $\mathbf{x}_{n+1}^{(k+1)}$ satisfies one of the recurrences

$$\mathbf{x}_{n+1}^{(k+1)} = (1+\rho\gamma)\bar{\mathbf{x}} - \rho\gamma\mathbf{x}_{n+1}^{(k)} \quad \text{or} \quad \mathbf{x}_{n+1}^{(k+1)} = (1-\gamma)\bar{\mathbf{x}} + \gamma\mathbf{x}_{n+1}^{(k)}. \tag{3.7}$$

The homogeneous forms of these equations are

$$\mathbf{y}_{n+1}^{(k+1)} = -\rho\gamma\mathbf{y}_{n+1}^{(k)} \quad \text{or} \quad \mathbf{y}_{n+1}^{(k+1)} = \gamma\mathbf{y}_{n+1}^{(k)}. \tag{3.8}$$

Since $0 < \gamma < 1$ and $0 < \rho\gamma < 1$, we have $\lim_{k\to\infty} \mathbf{y}_{n+1}^{(k)} = \mathbf{0}$, so that the solutions of both equations in (3.8) are zero as $k \to \infty$.

Now we need only to find a particular solution to the inhomogeneous forms of (3.7). Both are satisfied by the constant vector $\bar{\mathbf{x}}$, so that their general solutions are given by $\mathbf{x}_{n+1}^{(k)} = \mathbf{y}_{n+1}^{(k)} + \bar{\mathbf{x}}$, where $\mathbf{y}_{n+1}^{(k)}$ satisfies one of the relations (3.8). Since $\lim_{k\to\infty} \mathbf{y}_{n+1}^{(k)} = \mathbf{0}$, it follows that

$$\lim_{k\to\infty} \mathbf{x}_{n+1}^{(k)} = \mathbf{x}_{n+1}^* = \bar{\mathbf{x}}, \qquad \text{with } f_{n+1}^* = f(\bar{\mathbf{x}}).$$

But we know from the beginning of the proof that $f(\bar{\mathbf{x}}) \leq f_n^*$, which means that $f_{n+1}^* \leq f_n^*$. Lemma 3.3, part (3), shows that this can be true only if $f_n^* = f_{n+1}^*$, which gives part (1).

The result of part (2) is immediate because we have already shown a contradiction if there exists K such that $\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_n^{(k)}$ remain constant for $k \geq K$. ∎

In analyzing convergence, we know from Lemma 3.4 that, if broken convergence occurs, there exists an index j such that all vertices $\{\mathbf{x}_i^{(k)}\}$, for $1 \leq i \leq j$, remain constant from some point on. If this happens, the best point $\mathbf{x}_1^{(k)}$ will not be changed, and hence expansion steps cannot occur. (Nor can reflection steps in which a strict improvement is found over f_1.) For this reason, it is interesting to consider a *restricted* Nelder-Mead algorithm in which expansion steps are not taken; the analysis of the restricted algorithm is simpler because both $\mathrm{vol}(\Delta_k)$ and $\mathrm{diam}(\Delta_k)$ are nonincreasing if $\rho \leq 1$. We do not discuss the restricted algorithm further in this paper, but see [3].

In the remainder of this paper we consider strictly convex functions f with bounded level sets. The *level set* $\Gamma_\mu(f)$ is defined as

$$\Gamma_\mu(f) = \{\, \mathbf{x} : f(\mathbf{x}) \leq \mu \,\}. \tag{3.9}$$

A function f has *bounded level sets* if $\Gamma_\mu(f)$ is bounded for every μ; this restriction excludes strictly convex functions like e^{-x}. The point of this restriction is that a strictly convex function with bounded level sets has a unique minimizer $\mathbf{x}_{\min}$.

4. Nelder-Mead in Dimension 1 for Strictly Convex Functions

We analyze the Nelder-Mead algorithm in dimension 1 on strictly convex functions with bounded level sets. The behavior of the Nelder-Mead algorithm in dimension 1 depends nontrivially on the values of the reflection coefficient ρ, the expansion coefficient χ, and the contraction coefficient γ. (The shrink coefficient σ is irrelevant because shrink steps cannot occur for a strictly convex function; see Lemma 3.5.) We show that convergence to $x_{\min}$ always occurs as long as $\rho\chi \geq 1$ (Theorem 4.1), and that convergence is M-step linear when $\rho = 1$ (Theorem 4.2). The algorithm does not always converge to the minimizer $x_{\min}$ if $\rho\chi < 1$. An interesting feature of the analysis is that M-step linear convergence can be guaranteed even though infinitely many expansion steps may occur.

4.1. Special properties in one dimension

In one dimension, the "next-worst" and the "best" vertices are the same point, which means that the centroid $\bar{x}^{(k)}$ is equal to $x_1^{(k)}$ at every iteration. A Nelder-Mead simplex is a line segment, so that, given iteration k of type τ_k,

$$\operatorname{diam}(\Delta_{k+1}) = |\tau_k| \operatorname{diam}(\Delta_k). \tag{4.1}$$

Thus, in the special case of the standard parameters $\rho = 1$ and $\chi = 2$, a reflection step retains the same diameter and an expansion step doubles the diameter of the simplex. To deal with different orderings of the endpoints, we use the notation $\operatorname{int}(y, z)$ to denote the open interval with endpoints y and z (even if $y > z$), with analogous notation for closed or semi-open intervals.

The following lemma summarizes three important properties, to be used repeatedly, of strictly convex functions in $\mathcal{R}^1$ with bounded level sets.

Lemma 4.1. *Let f be a strictly convex function on $\mathcal{R}^1$ with a unique minimizer $x_{\min}$.*

(1) Let y_1, y_2, and y_3 be three distinct points such that $y_2 \in \operatorname{int}(y_1, y_3)$. Then

$$f(y_1) \geq f(y_2) \quad \text{and} \quad f(y_2) \leq f(y_3) \implies x_{\min} \in \operatorname{int}(y_1, y_3).$$

(2) If $x_{\min} \in \operatorname{int}[y_1, y_2]$, then $f(y_2 + \xi_2(y_1 - y_2)) > f(y_2 + \xi_1(y_1 - y_2))$ if $\xi_2 > \xi_1 \geq 1$.

(3) f is continuous. ∎

A special property of the one-dimensional case is that a Nelder-Mead iteration can never terminate in Step 2 of Algorithm NM (see Section 2): either a contraction will be taken (Step 4), or an expansion step will be tried (Step 3). Using the rule in Step 3 that we must accept the better of the reflection and expansion points, a reflection step will be taken only if $f_r < f_1$ and $f_e \geq f_r$.

4.2. Convergence to the minimizer

We first consider general Nelder-Mead parameters satisfying (2.1) and show that the condition $\rho\chi \geq 1$ is *necessary* for the global convergence of the algorithm to $x_{\min}$. If $\rho\chi < 1$, the so-called "expand" step actually *reduces* the simplex diameter, and the endpoints of the Nelder-Mead interval can move a distance of at most $\operatorname{diam}(\Delta_0)/(1 - \rho\chi)$ from the initial vertex $x_1^{(0)}$. Thus convergence to $x_{\min}$ will not occur whenever

$$\rho\chi < 1 \quad \text{and} \quad |x_{\min} - x_1^{(0)}| > \operatorname{diam}(\Delta_0)/(1 - \rho\chi).$$

We next show the general result that the condition $\rho\chi \geq 1$, combined with the requirements (2.1), is sufficient for global convergence to $x_{\min}$ of the Nelder-Mead algorithm in one dimension.

Theorem 4.1. (Convergence of one-dimensional Nelder-Mead method.) *Let f be a strictly convex function on $\mathcal{R}^1$ with bounded level sets. Assume that the Nelder-Mead algorithm is applied to f with parameters satisfying $\rho > 0$, $\chi > 1$, $\chi > \rho$, $\rho\chi \geq 1$, and $0 < \gamma < 1$, beginning with a nondegenerate initial simplex Δ_0. Then both endpoints of the Nelder-Mead interval converge to $x_{\min}$.*

The proof of this theorem depends on several intermediate lemmas. First we show that the Nelder-Mead algorithm finds, within a finite number of iterations, an "interval of uncertainty" in which the minimizer must lie.

Lemma 4.2. (Bracketing of $x_{\min}$.) *Let f be a strictly convex function on $\mathcal{R}^1$ with bounded level sets. Assume that the Nelder-Mead algorithm is applied to f beginning with a nondegenerate initial simplex Δ_0, and that the reflection and expansion coefficients satisfy $\rho > 0$, $\chi > 1$, $\chi > \rho$, and $\rho\chi \geq 1$. Then there is a smallest integer K satisfying*

$$K \leq \frac{|x_{\min} - x_1^{(0)}|}{\operatorname{diam}(\Delta_0)}, \quad \textit{such that} \quad f_2^{(K)} \geq f_1^{(K)} \ \textit{and} \ f_1^{(K)} \leq f_e^{(K)}. \tag{4.2}$$

In this case, $x_{\min} \in \operatorname{int}(x_2^{(K)}, x_e^{(K)})$ and we say that $x_{\min}$ is bracketed *by $x_2^{(K)}$ and $x_e^{(K)}$.*

Proof. To reduce clutter, we drop the superscript k and use a prime to denote quantities associated with iteration $k+1$. By definition, $f_2 \ge f_1$, so that the first inequality in the "up–down–up" relation involving f in (4.2) holds automatically for every Nelder-Mead interval. There are two possibilities.

(i) If $f_1 \le f_e$, the "up–down–up" pattern of f from (4.2) holds at the current iteration.

(ii) If $f_1 > f_e$, we know from strict convexity that $f_r < f_1$, and the expansion point is accepted. At the next iteration, $x_2' = x_1$ and $x_1' = x_e$. There are two cases to consider.

First, suppose that $x_{\min}$ lies in $\operatorname{int}(x_2', x_1'] = \operatorname{int}(x_1, x_e]$. Using result (2) of Lemma 4.1, both $f(x_r')$ and $f(x_e')$ must be strictly larger than $f(x_1')$. Hence the "up–down–up" pattern of (4.2) holds at the next iteration.

Alternatively, suppose that $x_{\min}$ lies "beyond" x_e, i.e., beyond x_1'. Then

$$|x_{\min} - x_1'| = |x_{\min} - x_1| - \operatorname{diam}(\Delta').$$

It follows from (4.1) and the inequality $\rho\chi \ge 1$ that $\operatorname{diam}(\Delta') = \rho\chi \operatorname{diam}(\Delta) \ge \operatorname{diam}(\Delta)$. Thus the distance from $x_{\min}$ to the current best point is reduced by an amount bounded below by Δ_0, the diameter of the initial interval. This gives the upper bound on K of (4.2). ∎

The next result shows that, once $x_{\min}$ lies in a specified interval defined by the current Nelder-Mead interval and a number depending only on the reflection, expansion, and contraction coefficients, it lies in an analogous interval at all subsequent iterations.

Lemma 4.3. *Let f be a strictly convex function on $\mathcal{R}^1$ with bounded level sets. Assume that the Nelder-Mead algorithm with parameters satisfying $\rho > 0$, $\chi > 1$, $\chi > \rho$, $\rho\chi \ge 1$, and $0 < \gamma < 1$, is applied to f beginning with a nondegenerate initial simplex. We define N_{NM} as*

$$N_{NM} = \max\Big(\frac{1}{\rho\gamma},\ \frac{\rho}{\gamma},\ \rho\chi,\ \chi - 1 \Big), \tag{4.3}$$

and we say that the proximity property *holds at iteration k if*

$$x_{\min} \in \operatorname{int}\Big(\, x_2^{(k)},\ \ x_1^{(k)} + N_{NM}(x_1^{(k)} - x_2^{(k)}) \Big]. \tag{4.4}$$

Then, if the proximity property holds at iteration k, it holds at iteration $k+1$.

Proof. To reduce clutter, we omit the index k and use a prime to denote quantities associated with iteration $k+1$. The proof considers all possible cases for location of $x_{\min}$ in the interval defined by (4.4). We have either $x_2 < x_1 < x_r < x_e$ or $x_e < x_r < x_1 < x_2$.

Case 1. $x_{\min} \in \operatorname{int}(x_2, x_1]$.

Lemma 4.1, part (2), implies that $f_r > f_1$, which means that a contraction step will be taken.

1a. If $f_r \ge f_2$, an inside contraction will occur, with $x_{cc} = x_1 - \gamma(x_1 - x_2)$. Strict convexity implies that $f_{cc} < f_2$.

(i) If $f_{cc} \ge f_1$, $x_{\min}$ lies in $\operatorname{int}(x_{cc}, x_1]$. The next Nelder-Mead interval is given by $x_2' = x_{cc}$ and $x_1' = x_1$, which means that $x_{\min} \in \operatorname{int}(x_2', x_1']$, and the proximity property holds at the next iteration.

(ii) If $f_{cc} < f_1$, the next Nelder-Mead interval is $x_2' = x_1$ and $x_1' = x_{cc}$. We also know that $x_{\min} \ne x_1$, so that $x_{\min} \in \operatorname{int}(x_2, x_1) = \operatorname{int}(x_2, x_2')$. To check whether (4.4) holds, we express x_2 in terms of the new Nelder-Mead interval as $x_2 = x_1' + \xi(x_1' - x_2')$. Using the definition of x_{cc} gives

$$x_2 = x_{cc} + \xi(x_{cc} - x_1) = x_1 + \gamma(x_2 - x_1) + \xi\gamma(x_2 - x_1), \quad \text{so that} \quad \xi = 1/\gamma - 1.$$

For $\rho > 1$, we have $1/\gamma - 1 < \rho/\gamma \le N_{NM}$, while for $0 < \rho \le 1$ we have $1/\gamma - 1 < 1/(\rho\gamma) \le N_{NM}$, so that the proximity property (4.4) holds at the next iteration.

1b. If $f_r < f_2$, an outside contraction will occur, with $x_c = x_1 + \rho\gamma(x_1 - x_2)$. Since $x_{\min} \in \operatorname{int}(x_2, x_1]$, part (2) of Lemma 4.1 implies that $f_c > f_1$. The new Nelder-Mead interval is given by $x_2' = x_c$ and $x_1' = x_1$, and the interval of uncertainty remains $\operatorname{int}(x_2, x_1']$. Expressing x_2 as $x_1' + \xi(x_1' - x_2')$ gives

$$x_2 = x_1 + \xi(x_1 - x_c) = x_1 - \xi\rho\gamma(x_1 - x_2), \quad \text{so that} \quad \xi = 1/\rho\gamma \le N_{NM},$$

and (4.4) holds at the next iteration.

Case 2. $x_{\min} \in \operatorname{int}(x_1, x_r]$.

2a. If $f_r < f_1$, we try the expansion step x_e. Part (2) of Lemma 4.1 implies that $f_e > f_r$, which means that the reflection step is accepted, and the new Nelder-Mead interval is $x_2' = x_1$ and $x_1' = x_r$. Then $x_{\min} \in \operatorname{int}(x_2', x_1']$, and (4.4) holds at the next iteration.

2b. If $f_r \geq f_2$, an inside contraction will be taken, $x_{cc} = x_1 - \gamma(x_1 - x_2)$. We also know that $x_{\min} \neq x_r$, so that $x_{\min} \in \operatorname{int}(x_1, x_r)$. Part (2) of Lemma 4.1 implies that $f_{cc} > f_1$, and the next Nelder-Mead interval is $x_2' = x_{cc}$ and $x_1' = x_1$, with $x_{\min} \in \operatorname{int}(x_1', x_r)$. We express x_r as $x_1' + \xi(x_1' - x_2')$, which gives

$$x_r = x_1 + \rho(x_1 - x_2) = x_1 + \xi(x_1 - x_{cc}) = x_1 + \xi\gamma(x_1 - x_2), \quad \text{so that} \quad \xi = \rho/\gamma \leq N_{NM},$$

and (4.4) holds at the next iteration.

2c. If $f_r \geq f_1$ and $f_r < f_2$, an outside contraction will be taken, $x_c = x_1 + \rho\gamma(x_1 - x_2)$. We also know that $x_{\min} \neq x_r$, so that $x_{\min} \in \operatorname{int}(x_1, x_r)$.

(i) If $f_c > f_1$, the new Nelder-Mead interval is $x_2' = x_c$ and $x_1' = x_1$. Because $f_c > f_1$, $x_{\min} \in \operatorname{int}(x_1, x_c) = \operatorname{int}(x_2', x_1')$, and (4.4) holds at the next iteration.

(ii) If $f_c < f_1$, the new Nelder-Mead interval is $x_2' = x_1$ and $x_1' = x_c$, and $x_{\min} \neq x_1$. The interval of uncertainty remains $\operatorname{int}(x_1, x_r) = \operatorname{int}(x_2', x_r)$. We thus write x_r as $x_1' + \xi(x_1' - x_2')$:

$$x_r = x_1 + \rho(x_1 - x_2) = x_c + \xi(x_c - x_1) = x_1 + \rho\gamma(x_1 - x_2) + \xi\rho\gamma(x_1 - x_2), \quad \text{so that} \quad \xi = 1/\gamma - 1 < N_{NM},$$

and (4.4) holds at the next iteration.

Case 3. $x_{\min} \in \operatorname{int}(x_r, x_e]$.

3a. If $f_e \geq f_r$, the new Nelder-Mead interval is $x_2' = x_1$ and $x_1' = x_r$; furthermore, $x_{\min} \neq x_e$ and $x_{\min} \in \operatorname{int}(x_1', x_e)$. Expressing x_e as $x_1' + \xi(x_1' - x_2')$ gives

$$x_e = x_1 + \rho\chi(x_1 - x_2) = x_r + \xi(x_r - x_1) = x_1 + \rho(x_1 - x_2) + \xi\rho(x_1 - x_2), \quad \text{so that} \quad \xi = \chi - 1.$$

Since $\xi \leq N_{NM}$, (4.4) holds at the next iteration.

3b. If $f_e < f_r$, we accept x_e. The new Nelder-Mead interval is $x_2' = x_1$ and $x_1' = x_e$. Since x_r lies between x_1 and x_e, $x_{\min} \in \operatorname{int}(x_2', x_1')$ and (4.4) holds at the next iteration.

Case 4. $x_{\min} \in \operatorname{int}(x_e,\ x_1 + N_{NM}(x_1 - x_2)]$.

Case 4 can happen only if $N_{NM} > \rho\chi$, since $x_e = x_1 + \rho\chi(x_1 - x_2)$. Thus it must be true that $f_1 > f_r > f_e$, and the expansion point will be accepted. The new Nelder-Mead interval is defined by $x_2' = x_1$ and $x_1' = x_e$. Writing $x_1 + N_{NM}(x_1 - x_2)$ as $x_e + \xi(x_e - x_1)$ gives

$$x_1 + N_{NM}(x_1 - x_2) = x_1 + \rho\chi(x_1 - x_2) + \xi\rho\chi(x_1 - x_2), \ \text{so that} \ \xi = (N_{NM} - \rho\chi)/\rho\chi.$$

Since $\rho\chi \geq 1$, $\xi < N_{NM}$ and the proximity property holds at the next iteration.

Cases 1–4 are exhaustive, and the lemma is proved. ∎

We prove that the Nelder-Mead simplex diameter converges to zero by first showing that the result of Lemma 3.6 holds, i.e., the function values at the interval endpoints converge to the same value, even when $\rho\gamma \geq 1$.

Lemma 4.4. *Let f be a strictly convex function on $\mathcal{R}^1$ with bounded level sets. Assume that the Nelder-Mead algorithm with parameters satisfying $\rho > 0$ and $0 < \gamma < 1$ is applied to f beginning with a nondegenerate initial simplex. Then $f_1^* = f_2^*$.*

Proof. If $\rho\gamma < 1$, the result follows from Lemma 3.6. Hence we assume that $\rho\gamma \geq 1$, which means that $\rho > 1$. The proof is by contradiction, beginning as in the proof of Lemma 3.6. If $f_1^* < f_2^*$, there is an iteration index K such that, for $k \geq K$, every iteration k is a contraction and x_1 does not change. (Without loss of generality, we may take $K = 0$.)

If iteration k is an inside contraction, $\operatorname{diam}(\Delta_{k+1}) = \gamma \operatorname{diam}(\Delta_k) < \operatorname{diam}(\Delta_k)$. If iteration k is an outside contraction, $\operatorname{diam}(\Delta_{k+1}) = \rho\gamma \operatorname{diam}(\Delta_k) \geq \operatorname{diam}(\Delta_k)$. Thus $\lim_{k\to\infty} \operatorname{diam}(\Delta_k) \to 0$ if there are a finite number of outside contractions, and so we need to consider only the case of an infinite number of outside contractions.

Suppose that iteration k is an outside contraction. Then $f_r^{(k)} \ge f_1^{(k)}$, $f_r^{(k)} < f_2^{(k)}$, and the contraction point is $x_c^{(k)} = x_1^{(k)} + \rho\gamma(x_1^{(k)} - x_2^{(k)})$. Since the best point does not change, $f_c^{(k)} \ge f_1^{(k)}$ and $x_2^{(k+1)} = x_c^{(k)}$. By strict convexity, $f_c^{(k)} < f_r^{(k)}$.

Define $z(\xi)$ as

$$z(\xi) \equiv x_1^{(k)} + \xi\big(x_1^{(k)} - x_2^{(k)}\big),$$

so that $x_2^{(k)} = z(-1)$ and $x_r^{(k)} = z(\rho)$. Expressing $f_2^{(k)}$, $f_1^{(k)}$, and $f_c^{(k)}$ in this form, we have

$$f\big(z(-1)\big) \;>\; f\big(z(0)\big) \;\le\; f\big(z(\rho\gamma)\big) = f_2^{(k+1)}, \quad \text{so that} \quad x_{\min} \in \operatorname{int}\big(\,z(-1),\, z(\rho\gamma)\,\big). \tag{4.5}$$

The relation $f(z(-1)) = f_2^{(k)} > f_2^{(k+1)}$ and result (2) of Lemma 4.1 then imply that

$$f(z(\xi)) > f_2^{(k+1)} \quad \text{if} \quad \xi \le -1. \tag{4.6}$$

The next reflection point $x_r^{(k+1)}$ is given by

$$x_r^{(k+1)} = x_1^{(k)} + \rho(x_1^{(k)} - x_2^{(k+1)}) = x_1^{(k)} - \rho^2\gamma(x_1^{(k)} - x_2^{(k)}) = z(-\rho^2\gamma).$$

Since $\rho\gamma \ge 1$ and $\rho > 1$, we have $\rho^2\gamma > 1$, and we conclude from (4.6) that $f_r^{(k+1)}$ strictly exceeds $f_2^{(k+1)}$. Iteration $k+1$ must therefore be an *inside* contraction, with

$$x_{cc}^{(k+1)} = x_1^{(k+1)} + \gamma(x_1^{(k+1)} - x_2^{(k+1)}) = x_1^{(k)} + \rho\gamma^2(x_1^{(k)} - x_2^{(k)}) = z(\rho\gamma^2).$$

Because x_1 does not change, $x_2^{(k+2)} = x_{cc}^{(k+1)}$ and the reflection point at iteration $k+2$ is given by

$$x_r^{(k+2)} = x_1^{(k)} + \rho(x_1^{(k)} - x_2^{(k+2)}) = x_1^{(k)} - \rho^2\gamma^2(x_1^{(k)} - x_2^{(k)}) = z(-\rho^2\gamma^2).$$

Since $\rho^2\gamma^2 \ge 1$, (4.6) again implies that the value of f at $x_r^{(k+2)}$ exceeds $f_2^{(k+2)}$, and iteration $k+2$ must be an inside contraction. Continuing, if iteration k is an outside contraction followed by j inside contractions, the (rejected) reflection point at iteration $k+j$ is $z(-\rho^2\gamma^j)$ and the (accepted) contraction point is $z(\rho\gamma^{j+1})$.

Because of (4.6), iteration $k+j$ must be an inside contraction as long as $\rho^2\gamma^j \ge 1$. Let c^* denote the smallest integer such that $\rho^2\gamma^{c^*} < 1$; note that $c^* > 2$. It follows that the sequence of contractions divides into blocks, where the jth block consists of a single outside contraction followed by some number c_j of inside contractions, with $c_j \ge c^*$ in each case. Letting k_j denote the iteration index at the start of the jth such block, we have

$$\operatorname{diam}(\Delta_{k_j}) = \rho\gamma^{c_j}\operatorname{diam}(\Delta_{k_{j-1}}) \le \theta\operatorname{diam}(\Delta_{k_{j-1}}), \quad \text{with} \quad \theta = \rho\gamma^{c^*} < 1.$$

The simplex of largest diameter within each block occurs after the outside contraction, and has diameter $\rho\gamma\operatorname{diam}(\Delta_{k_j})$. Thus we have

$$\lim_{k\to\infty}\operatorname{diam}(\Delta_k) \to 0, \quad \lim_{k\to\infty} x_2^{(k)} = x_1^{(k)}, \quad \text{and} \quad f_2^* = f_1^*,$$

contradicting our assumption that $f_1^* < f_2^*$ and giving the desired result. ∎

We next show that in all cases the simplex diameter converges to zero, i.e., the simplex shrinks to a point.

Lemma 4.5. *Let f be a strictly convex function on $\mathcal{R}^1$ with bounded level sets. Assume that the Nelder-Mead algorithm with parameters satisfying $\rho > 0$ and $0 < \gamma < 1$ is applied to f beginning with a nondegenerate initial simplex Δ_0. Then $\lim_{k\to\infty}\operatorname{diam}(\Delta_k) = 0$.*

Proof. Lemma 4.4 shows that $f_1^* = f_2^*$. If $f_1^* = f_{\min}$, this function value is assumed at exactly one point, $x_{\min}$, and the desired result is immediate. If $f_1^* > f_{\min}$, we know from strict convexity that f takes the value f_1^* at exactly two distinct points, denoted by x_1^* and x_2^*, with $x_1^* < x_{\min} < x_2^*$. The vertex function values converge from above to their limits and f is continuous. Thus for any $\epsilon > 0$ there is an iteration index $\widetilde{K}$ such that, for $k \ge \widetilde{K}$, $x_1^{(k)}$ and $x_2^{(k)}$ are confined to $\mathcal{I}_1^\epsilon \cup \mathcal{I}_2^\epsilon$, where

$$\mathcal{I}_1^\epsilon = [x_1^* - \epsilon,\, x_1^*] \quad \text{and} \quad \mathcal{I}_2^\epsilon = [x_2^*,\, x_2^* + \epsilon]. \tag{4.7}$$

There are two cases to consider.

Case 1. Both endpoints $x_1^{(k)}$ and $x_2^{(k)}$ lie in the same interval for infinitely many iterations, i.e., for one of $j = 1, 2$, the relation

$$x_1^{(k)} \in \mathcal{I}_j^\epsilon \quad \text{and} \quad x_2^{(k)} \in \mathcal{I}_j^\epsilon \tag{4.8}$$

holds for infinitely many k.

In this case we assert that both endpoints remain in one of these intervals for *all* sufficiently large k. This result is proved by contradiction: assume that for any $\epsilon > 0$ and iteration K_1 where (4.8) holds, there is a later iteration K_2 at which $x_1^{(K_2)}$ and $x_2^{(K_2)}$ are in different intervals. Then, since $\operatorname{diam}(\Delta_{K_1}) \le \epsilon$ and $\operatorname{diam}(\Delta_{K_2}) \ge x_2^* - x_1^*$, we may pick ϵ so small that $\operatorname{diam}(\Delta_{K_2}) > \max(1, \rho\chi) \operatorname{diam}(\Delta_{K_1})$. The simplex diameter can be increased only by reflection, expansion, or outside contraction, and the maximum factor by which the diameter can increase in a single iteration is $\rho\chi$. If $x_1^{(K_1)}$ and $x_2^{(K_1)}$ are both in $\mathcal{I}_1^\epsilon$, then strict convexity implies that any reflection, expansion, or outside contraction must move toward $\mathcal{I}_2^\epsilon$ (and vice versa if the two vertices lie in $\mathcal{I}_2^\epsilon$). But if ϵ is small enough so that $\epsilon\rho\chi < x_2^* - x_1^*$, then some trial point between iterations K_1 and K_2 must lie in the open interval (x_1^*, x_2^*), and by strict convexity its associated function value is less than f_1^*, a contradiction. We conclude that, since the Nelder-Mead endpoints $x_1^{(k)}$ and $x_2^{(k)}$ are in $\mathcal{I}_j^\epsilon$ for all sufficiently large k, and since $f_2^{(k)} \to f_1^{(k)} \to f_1^*$, both endpoints must converge to the point x_j^*, and $\operatorname{diam}(\Delta_k) \to 0$.

Case 2. Both endpoints $x_1^{(k)}$ and $x_2^{(k)}$ are in separate intervals $\mathcal{I}_1^\epsilon$ and $\mathcal{I}_2^\epsilon$ for all $k \ge K_1$.

We show by contradiction that this cannot happen because an inside contraction eventually occurs that generates a point inside (x_1^*, x_2^*). Let x_r^* denote the reflection point for the Nelder-Mead interval $[x_1^*, x_2^*]$, where either point may be taken as the "best" point; we know from strict convexity that $f(x_r^*) > f_1^*$, with $f_r^* = f_1^* + \delta_r$ for some $\delta_r > 0$. Because f is continuous and $x_r^{(k)}$ is a continuous function of $x_1^{(k)}$ and $x_2^{(k)}$, it follows that, given any $\delta > 0$, eventually $f_1^{(k)}$, $f_2^{(k)}$, and $f_r^{(k)}$ are within δ of their limiting values. Thus, for sufficiently large k, $f_r^{(k)} > f_2^{(k)} \ge f_1^{(k)}$ and an inside contraction will be taken.

Since $x_1^{(k)}$ and $x_2^{(k)}$ are in different intervals, the inside contraction point $x_{cc}^{(k)}$ satisfies

$$x_1^* - \epsilon + \gamma\big(x_2^* - (x_1^* - \epsilon)\big) \;\le\; x_{cc}^{(k)} \;\le\; x_2^* + \epsilon + \gamma\big(x_1^* - (x_2^* + \epsilon)\big).$$

If ϵ is small enough, namely $\epsilon < \gamma(x_2^* - x_1^*)/(1-\gamma)$, then

$$x_1^* \;<\; x_1^* + \gamma(x_2^* - x_1^*) - (1-\gamma)\epsilon \;\le\; x_{cc}^{(k)} \;\le\; x_2^* - \gamma(x_2^* - x_1^*) + (1-\gamma)\epsilon < x_2^*,$$

i.e., $x_{cc}^{(k)}$ lies in the open interval (x_1^*, x_2^*) and $f(x_{cc}^{(k)}) < f_1^*$, a contradiction. ∎

We now combine these lemmas to prove Theorem 4.1.

Proof of Theorem 4.1 (Convergence of Nelder-Mead in one dimension.) Lemma 4.2 shows that $x_{\min}$ is eventually bracketed by the worst vertex and the expansion point, i.e., for some iteration K,

$$x_{\min} \in \operatorname{int}(x_2^{(K)}, x_1^{(K)} + \rho\chi(x_1^{(K)} - x_2^{(K)}).$$

Since the constant N_{NM} of (4.3) satisfies $N_{NM} \ge \rho\chi$, Lemma 4.3 shows that, for all $k \ge K$, $x_{\min}$ satisfies the proximity property (4.4),

$$x_{\min} \in \operatorname{int}(x_2^{(k)}, x_1^{(k)} + N_{NM}(x_1^{(k)} - x_2^{(k)}),$$

which implies that

$$|x_{\min} - x_1^{(k)}| \le N_{NM} \operatorname{diam}(\Delta_k). \tag{4.9}$$

Lemma 4.5 shows that $\operatorname{diam}(\Delta_k) \to 0$. Combined with (4.9), this gives the desired result. ∎

4.3. Linear convergence with $\rho = 1$

When the reflection coefficient is the standard choice $\rho = 1$, the Nelder-Mead method not only converges to the minimizer, but its convergence rate is eventually M-step linear, i.e., the distance from the best vertex to the optimal point decreases every M steps by at least a fixed multiplicative constant less than one. This result follows from analyzing the special structure of permitted Nelder-Mead move sequences.

Theorem 4.2. (Linear convergence of Nelder-Mead method for $n = 1$ with $\rho = 1$.) *Let f be a strictly convex function on $\mathcal{R}^1$ with bounded level sets. Assume that the Nelder-Mead algorithm with reflection coefficient $\rho = 1$, and expansion and contraction coefficients satisfying $\chi > 1$ and $0 < \gamma < 1$, is applied to f beginning with a nondegenerate initial simplex Δ_0. Then there is an integer M depending only on χ and γ such that*

$$\operatorname{diam}(\Delta_{k+M}) \le \tfrac{1}{2}\operatorname{diam}(\Delta_k) \quad \textit{for all} \quad k \ge K,$$

where K is the iteration index defined in Lemma 4.2.

As the first step in proving this theorem, we obtain two results unique to dimension 1 about sequences of Nelder-Mead iterations.

Lemma 4.6. *Let f be a strictly convex function on $\mathcal{R}^1$ with bounded level sets, and assume that the Nelder-Mead algorithm with parameters $\rho = 1$, $\chi > 1$, and $0 < \gamma < 1$, is applied to f beginning with a nondegenerate initial simplex. Then*

(1) the number of consecutive reflections is bounded by $r^* = \lceil \chi - 1 \rceil$;

(2) the iteration immediately following a reflection cannot be an expansion.

Proof. For any iteration k, define $z^{(k)}(\xi)$ as

$$z^{(k)}(\xi) \equiv x_1^{(k)} + \xi\big(x_1^{(k)} - x_2^{(k)}\big), \tag{4.10}$$

so that $x_2^{(k)} = z^{(k)}(-1)$, $x_r^{(k)} = z^{(k)}(1)$, and $x_e^{(k)} = z^{(k)}(\chi)$.

If iteration k is a reflection,

$$f_r^{(k)} < f_1^{(k)}, \quad f_e^{(k)} \ge f_r^{(k)}, \quad x_1^{(k+1)} = x_r^{(k)}, \quad \text{and} \quad x_2^{(k+1)} = x_1^{(k)}. \tag{4.11}$$

Applying Lemma 4.1 to the first two relations in (4.11), we see that $x_{\min} \in \operatorname{int}(x_1^{(k)}, x_e^{(k)})$ and

$$f\big(z^{(k)}(\xi)\big) \ge f_1^{(k+1)} \quad \text{if} \quad \xi \ge \chi. \tag{4.12}$$

Starting with iteration k, the (potential) ℓth consecutive reflection point is given by

$$x_r^{(k+\ell-1)} = x_1^{(k)} + \ell(x_1^{(k)} - x_2^{(k)}) = z^{(k)}(\ell), \tag{4.13}$$

which can be accepted only if its function value is strictly less than $f(x_1^{(k+\ell-1)})$. Strict convexity and (4.12) show that any point $z^{(k)}(\xi)$ with $\xi \ge \chi$ cannot be an accepted reflection point. Thus the number of consecutive reflections is bounded by the integer r^* satisfying

$$r^* < \chi \ \text{ and } \ r^* + 1 \ge \chi, \quad \text{i.e.,} \quad r^* = \lceil \chi - 1 \rceil.$$

This completes the proof of (1).

If iteration k is a reflection, the expansion point at iteration $k+1$ is given by

$$x_e^{(k+1)} = x_1^{(k+1)} + \chi(x_1^{(k+1)} - x_2^{(k+1)}) = x_1^{(k)} + (1+\chi)(x_1^{(k)} - x_2^{(k)}) = z^{(k)}(1+\chi).$$

Relation (4.12) implies that the function value at $x_e^{(k+1)}$ exceeds $f_1^{(k+1)}$, so that $x_e^{(k+1)}$ will not be accepted. This proves result (2), and shows that the iteration immediately following a successful reflection must be either a reflection or a contraction. ∎

Note that $r^* = 1$ whenever the expansion coefficient $\chi \le 2$; thus there cannot be two consecutive reflections with the standard Nelder-Mead coefficients (2.2) for $n = 1$.

As a corollary, we show that a contraction must occur no later than iteration $K + r^*$, where K is the first iteration at which the minimizer is bracketed by x_2 and the expansion point (Lemma 4.2).

Corollary 4.1. *Let f be a strictly convex function on $\mathcal{R}^1$ with bounded level sets. Assume that the Nelder-Mead algorithm with $\rho = 1$ is applied to f beginning with a nondegenerate initial simplex Δ_0, and let K denote the iteration index defined by Lemma 4.2 at which, for the first time, $f_1^{(K)} \le f_e^{(K)}$. Then a contraction must occur no later than iteration $K + r^*$.*

Proof. There are two cases. If $f_r^{(K)} \geq f_1^{(K)}$, iteration K is a contraction, and the result is immediate. Otherwise, if $f_r^{(K)} < f_1^{(K)}$, iteration K is a reflection. Lemma 4.6 shows that there cannot be more than r^* consecutive reflections, and any sequence of consecutive reflections ends with a contraction. Hence a contraction must occur no later than iteration $K + r^*$. ∎

The next lemma derives a bound on the number of consecutive expansions immediately following a contraction.

Lemma 4.7. (Bounded consecutive expansions.) *Let f be a strictly convex function on $\mathcal{R}^1$ with bounded level sets. Assume that the Nelder-Mead algorithm with $\rho = 1$, $\chi > 1$, and $0 < \gamma < 1$, is applied to f beginning with a nondegenerate initial simplex Δ_0. If iteration k is a contraction, then for all subsequent iterations there can be no more than j^* consecutive expansion steps, where j^* is the largest integer satisfying*

$$\chi + \chi^2 + \cdots + \chi^{j^*} < N_{\scriptscriptstyle NM} = \max(\chi, 1/\gamma). \tag{4.14}$$

(This value of $N_{\scriptscriptstyle NM}$ is equal to its general definition (4.3) when $\rho = 1$.)

Proof. Since iteration k is a contraction, $x_{\min} \in \operatorname{int}(x_2^{(k)},\ x_r^{(k)})$. Thus the proximity property (4.4) is satisfied at iteration k, and, by Lemma 4.3, for all subsequent iterations. The first expansion in a sequence of consecutive expansions must immediately follow a contraction (see result (2) of Lemma 4.6), and strict convexity imposes a bound on the number of subsequent consecutive expansions.

Using the notation of (4.10), we consider inequalities that apply to the best function value $f_1^{(k+1)}$ at the next iteration, which is (possibly) the first expansion step in a sequence of consecutive expansions.

Case 1. If $f_r^{(k)} < f_2^{(k)}$, iteration k is an outside contraction with $x_c^{(k)} = x_1^{(k)} + \gamma(x_1^{(k)} - x_2^{(k)})$.

(i) If $f_c^{(k)} \geq f_1^{(k)}$, the next Nelder-Mead interval is defined by $x_2^{(k+1)} = x_c^{(k)}$ and $x_1^{(k+1)} = x_1^{(k)}$, and $x_{\min} \in \operatorname{int}(x_2^{(k)}, x_2^{(k+1)})$. (The tie-breaking rule in Section 2 is invoked if $f_c^{(k)} = f_1^{(k)}$.) If an expansion occurs, the interval will expand toward $x_2^{(k)}$, which satisfies

$$x_2^{(k)} = x_1^{(k+1)} + \left(x_1^{k+1} - x_2^{(k+1)}\right)/\gamma = z^{(k+1)}(1/\gamma), \quad \text{with} \quad f_2^{(k)} > f_1^{(k+1)}. \tag{4.15}$$

(ii) If $f_c^{(k)} < f_1^{(k)}$, the next Nelder-Mead interval is defined by $x_2^{(k+1)} = x_1^{(k)}$ and $x_1^{(k+1)} = x_c^{(k)}$, and $x_{\min} \in \operatorname{int}(x_1^{(k+1)}, x_r^{(k)})$. Any expansion will be toward $x_r^{(k)}$, which satisfies

$$x_r^{(k)} = x_1^{(k+1)} + (1/\gamma - 1)\left(x_1^{(k+1)} - x_2^{(k+1)}\right) = z^{(k+1)}(1/\gamma - 1), \quad \text{with} \quad f_r^{(k)} > f_1^{(k+1)}. \tag{4.16}$$

Case 2. If $f_r^{(k)} \geq f_2^{(k)}$, iteration k is an inside contraction with $x_{cc}^{(k)} = x_1^{(k)} - \gamma(x_1^{(k)} - x_2^{(k)})$.

(i) If $f_{cc}^{(k)} \geq f_1^{(k)}$, the next Nelder-Mead interval is defined by $x_2^{(k+1)} = x_{cc}^{(k)}$ and $x_1^{(k+1)} = x_1^{(k)}$, and $x_{\min} \in \operatorname{int}(x_2^{(k+1)}, x_r^{(k)})$. (The tie-breaking rule in Section 2 is invoked if $f_{cc}^{(k)} = f_1^{(k)}$.) If an expansion occurs, the interval will expand toward $x_r^{(k)}$, which satisfies

$$x_r^{(k)} = x_1^{(k+1)} + \left(x_1^{(k+1)} - x_2^{(k+1)}\right)/\gamma = z^{(k+1)}(1/\gamma), \quad \text{with} \quad f_r^{(k)} > f_1^{(k+1)}. \tag{4.17}$$

(ii) If $f_{cc}^{(k)} < f_1^{(k)}$, the next Nelder-Mead interval is defined by $x_2^{(k+1)} = x_1^{(k)}$ and $x_1^{(k+1)} = x_{cc}^{(k)}$, and $x_{\min} \in \operatorname{int}(x_2^{(k)}, x_1^{(k+1)})$. Any expansion will be toward $x_2^{(k)}$, which satisfies

$$x_2^{(k)} = x_1^{(k+1)} + (1/\gamma - 1)\left(x_1^{k+1} - x_2^{(k+1)}\right) = z^{(k+1)}(1/\gamma - 1), \quad \text{with} \quad f_2^{(k)} > f_1^{(k+1)}. \tag{4.18}$$

For each of the four cases 1(i)–2(ii), the value of f at $z^{(k+1)}(\xi)$ exceeds $f_1^{(k+1)}$ for some ξ that is equal to or bounded above by $N_{\scriptscriptstyle NM}$. Applying result (2) of Lemma 4.1 to the interval in which $x_{\min}$ lies and the corresponding expression from (4.15)–(4.18), we conclude that, if a sequence of consecutive expansions *begins* at iteration $k + 1$, then

$$f(z^{(k+1)}(\xi)) > f(x_1^{(k+1)}) \quad \text{whenever} \quad \xi \geq N_{\scriptscriptstyle NM}. \tag{4.19}$$

The remainder of the proof is similar to that of Lemma 4.6. Starting with iteration $k+1$, the (potential) jth consecutive expansion point is given by

$$x_e^{(k+j)} = z^{(k+1)}\big(\chi + \chi^2 + \cdots + \chi^j\big). \tag{4.20}$$

This point can be accepted only if its function value is strictly less than $f(x_1^{(k+j)})$, which strictly decreases after each accepted expansion. Relations (4.19) and (4.20) together show that $x_e^{(k+j)}$ can be accepted only if

$$\chi + \chi^2 + \cdots + \chi^j \;<\; N_{NM},$$

which holds automatically for $j = 0$ because $N_{NM} > 1$. The value of j is therefore bounded above by j^* defined in (4.14). ∎

For the standard expansion coefficient $\chi = 2$, the value of N_{NM} is $\max(2, 1/\gamma)$ and the values of j^* for several ranges of γ are:

$$j^* = 0 \text{ when } \tfrac{1}{2} \le \gamma < 1; \qquad j^* = 1 \text{ when } \tfrac{1}{6} \le \gamma < \tfrac{1}{2}; \qquad j^* = 2 \text{ when } \tfrac{1}{14} \le \gamma < \tfrac{1}{6}.$$

In the "standard" Nelder-Mead algorithm with contraction coefficient $\gamma = \frac{1}{2}$, the zero value of j^* means that *no* expansion steps can occur once the minimizer is bracketed by the worst point and the reflection point at any iteration.

Our results so far have shown that the minimizer is bracketed at iteration K (Lemma 4.2), and that a contraction must occur no later than iteration $K + r^*$ (Lemmas 4.6 and Corollary 4.1). Thereafter, no more than j^* consecutive expansions can occur (Lemma 4.7), and any sequence of consecutive expansions must end with either a contraction alone or a sequence of at most r^* consecutive reflections followed by a contraction (see Lemma 4.6). These constraints on legal iteration sequences following a contraction can be represented by a directed graph with four states: expansion, reflection, and the two forms of contraction. Each state is labeled by the absolute value of its move type, so that an inside contraction is labeled "γ", an outside contraction is labeled "$\rho\gamma$", a reflection is labeled "ρ", and an expansion is labeled "$\rho\chi$". According to (4.1), the simplex diameter is multiplied by ρ for a reflection, $\rho\chi$ for an expansion, $\rho\gamma$ for an outside contraction, and γ for an inside contraction. Starting in the contraction state with initial diameter 1, the diameter of the Nelder-Mead interval after any sequence of moves is thus the product of the state labels encountered.

Figure 3 shows the graph corresponding to $\rho = 1$, $\chi = 2$, and any contraction coefficient satisfying $\frac{1}{14} \le \gamma < \frac{1}{6}$. For these coefficients, at most two consecutive expansion steps can occur ($j^* = 2$), and at most one consecutive reflection ($r^* = 1$). (Because $\rho = 1$, we have not distinguished between inside and outside contractions.)

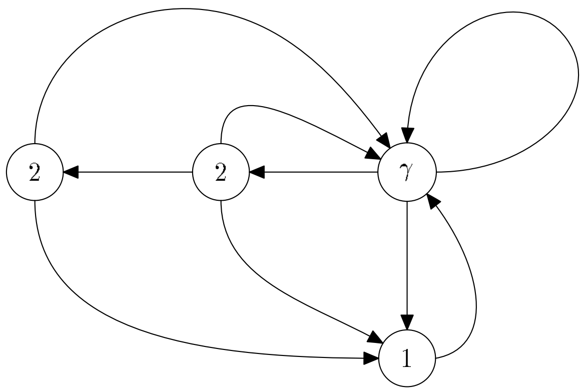


Figure 3: Directed graph depicting legal Nelder-Mead moves for $\rho = 1$, $\chi = 2$, and $\frac{1}{14} \le \gamma < \frac{1}{6}$.

We now examine the effects of valid Nelder-Mead move sequences on the simplex diameter.

Lemma 4.8. *Let f be a strictly convex function on $\mathcal{R}^1$ with bounded level sets. Assume that the Nelder-Mead algorithm with $\rho = 1$, $\chi > 1$, and $0 < \gamma < 1$, is applied to f beginning with a nondegenerate initial simplex Δ_0. Let Δ denote the simplex immediately following any contraction, and Δ' the simplex immediately following the next contraction. Then there exists a value φ depending only on χ and γ such that* $\mathrm{diam}(\Delta') \le \varphi\, \mathrm{diam}(\Delta)$, *where $\varphi < 1$.*

Proof. Lemma 4.7 shows that the number of consecutive expansions between any two contractions cannot exceed j^*. Since $N_{NM} = \max(1/\gamma, \chi)$ and reflection does not change the diameter, the worst-case growth occurs when j^* expansions are followed by a contraction, which corresponds to $\varphi = \chi^{j^*}\gamma$. If $j^* = 0$, $\varphi = \gamma$ and is consequently less than 1. If $N_{NM} = \chi$, j^* must be zero. In the remaining case when $N_{NM} = 1/\gamma$ and $j^* > 0$, (4.14) may be written as

$$\gamma(\chi + \cdots + \chi^{j^*}) < 1, \quad \text{which implies that} \quad \varphi = \gamma\chi^{j^*} < 1, \quad \text{the desired result.} \quad \blacksquare$$

Combining these results about the structure and effects of valid NM move sequences, we prove M-step linear convergence when $\rho = 1$.

Completion of the proof of Theorem 4.2. The first contraction in the Nelder-Mead method can occur no later than iteration $K + r^*$. Lemmas 4.6 and 4.7 show that any minimal cycle (i.e., a cycle that does not pass through any vertex twice) thereafter has length at most $j^* + r^* + 1$; Lemma 4.8 shows that the product of state labels over any cycle cannot exceed φ. For any integer m, a path of length $m(j^* + r^* + 1)$ must contain at least m minimal cycles. Given any such path, we can remove minimal cycles until at most $j^* + r^*$ edges are left over. Consequently, the simplex diameter at the end of the associated sequence of Nelder-Mead iterations must be multiplied by a factor no larger than $\chi^{j^*+r^*}\varphi^m$. If we choose m as the smallest value such that

$$\chi^{j^*+r^*}\varphi^m \le \tfrac{1}{2}, \quad \text{then } M = m(j^* + r^* + 1) \text{ satisfies } \mathrm{diam}(\Delta_{k+M}) \le \tfrac{1}{2}\,\mathrm{diam}(\Delta_k),$$

which gives the desired result. $\blacksquare$

M-step linear convergence can also be proved for certain ranges of parameter values with $\rho \ne 1$ by imposing restrictions that guarantee, for example, that $j^* = 0$ and $r^* = 1$.

4.4. A pattern search method interpretation of Nelder-Mead for $n = 1$

Pattern search methods [13] are direct search methods that presuppose a lattice grid pattern for search points. Torczon [14] has recently informed us that the analysis [13] for pattern search methods can be adapted in dimension 1 to the Nelder-Mead method when

$$\rho = 1 \qquad \text{and } \chi \text{ and } \gamma \text{ are rational.} \tag{4.21}$$

(These restrictions are satisfied for the standard coefficients $\rho = 1$, $\chi = 2$, and $\gamma = \frac{1}{2}$.) The condition $\rho = 1$ is needed to guarantee that, following an outside contraction at iteration k, the reflection point at iteration $k + 1$ is identical to the inside contraction at iteration k (and vice versa). Rationality of χ and γ is needed to retain the lattice structure that underlies pattern search methods. When (4.21) holds and f is once-continuously differentiable, the Nelder-Mead method generates the same sequence of points as a (related) pattern search method with relabeled iterations. Consequently, the results in [13] imply that $\liminf |\nabla f(x_k)| \to 0$, where x_k denotes the best point in Δ_k.

5. Standard Nelder-Mead in Dimension 2 for Strictly Convex Functions

In this section we consider the *standard* Nelder-Mead algorithm—with coefficients $\rho = 1$, $\chi = 2$, and $\gamma = \frac{1}{2}$—applied to a strictly convex function $f(\mathbf{x})$ on $\mathcal{R}^2$ with bounded level sets. The assumption that $\rho = 1$ is essential in our analysis.

We denote the (necessarily unique) minimizer of f by $\mathbf{x}_{\min}$, and let $f_{\min} = f(\mathbf{x}_{\min})$. Note that the level set $\{\mathbf{x} \mid f(\mathbf{x}) \le \mu\}$ is empty if $\mu < f_{\min}$, the single point $\mathbf{x}_{\min}$ if $\mu = f_{\min}$, and a closed convex set if $\mu > f_{\min}$.

5.1. Convergence of vertex function values

Our first result shows that, for the standard Nelder-Mead algorithm, the limiting function values at the vertices are equal.

Theorem 5.1. (Convergence of vertex function values for $n = 2$.) *Let f be a strictly convex function on $\mathcal{R}^2$ with bounded level sets. Assume that the Nelder-Mead algorithm with reflection coefficient $\rho = 1$ and contraction coefficient $\gamma = \frac{1}{2}$ is applied to f beginning with a nondegenerate initial simplex Δ_0. Then the three limiting vertex function values are the same, i.e.,*

$$f_1^* = f_2^* = f_3^*.$$

Proof. Corollary 3.1, which applies in any dimension, gives the result immediately if the best vertex $\mathbf{x}_1^{(k)}$ changes infinitely often. The following lemma treats the only remaining case, in which $\mathbf{x}_1^{(k)}$ eventually becomes constant.

Lemma 5.1. *Let f be a strictly convex function on $\mathcal{R}^2$ with bounded level sets. Assume that the Nelder-Mead algorithm with $\rho = 1$ and $\gamma = \frac{1}{2}$ is applied to f beginning with a nondegenerate initial simplex Δ_0. If the best vertex $\mathbf{x}_1^{(k)}$ is constant for all k, then the simplices Δ_k converge to the point $\mathbf{x}_1^{(0)}$ as $k \to \infty$.*

Proof. Without loss of generality, the (constant) best vertex $\mathbf{x}_1$ may be taken as the origin. The proof that $\mathbf{x}_2$ and $\mathbf{x}_3$ converge to the origin has four elements: (i) a matrix recursion that defines the Nelder-Mead vertices at the infinite subsequence of iterations when $\mathbf{x}_2$ changes; (ii) a special norm that measures progress toward the origin; (iii) bounds on this norm obtained from the singular values of a matrix constrained to a subspace; and (iv) the illegality of certain patterns of Nelder-Mead move types in the iteration subsequence.

(i) The matrix recursion. We know from Lemma 3.6 that the next-worst vertex $\mathbf{x}_2^{(k)}$ must change infinitely often. There is thus a subsequence of iterations $\{k_\ell\}$, $\ell = 0, 1, \ldots$, with $k_0 = 0$, where $\mathbf{x}_2$ changes, i.e.,

$$\mathbf{x}_2^{(k_{\ell+1})} \neq \mathbf{x}_2^{(k_\ell)} \quad \text{and} \quad \mathbf{x}_2^{(i)} = \mathbf{x}_2^{(i-1)}, \quad i = k_\ell + 1, \ldots, k_{\ell+1} - 1.$$

We then define new sequences $\tilde{\mathbf{x}}_2$ and $\tilde{\mathbf{x}}_3$ from

$$\tilde{\mathbf{x}}_2^{(\ell)} = \mathbf{x}_2^{(k_\ell)} \quad \text{and} \quad \tilde{\mathbf{x}}_3^{(\ell)} = \mathbf{x}_3^{(k_\ell)}. \tag{5.1}$$

Because $\mathbf{x}_1$ is constant and $\mathbf{x}_2$ changes at iteration k_ℓ, $\mathbf{x}_3$ thereupon becomes the "old" $\mathbf{x}_2$, i.e.,

$$\tilde{\mathbf{x}}_3^{(\ell)} = \tilde{\mathbf{x}}_2^{(\ell-1)}. \tag{5.2}$$

For each iteration strictly between k_ℓ and $k_{\ell+1}$, only $\mathbf{x}_3$ changes, so that

$$\mathbf{x}_3^{(i)} = \tfrac{1}{2}\mathbf{x}_2^{(i-1)} + \tau_{i-1}\big(\tfrac{1}{2}\mathbf{x}_2^{(i-1)} - \mathbf{x}_3^{(i-1)}\big) \quad \text{for} \quad i = k_\ell + 1, \ldots, k_{\ell+1} - 1, \tag{5.3}$$

where τ_i is the type of iteration i. Note that any iteration in which only $\mathbf{x}_3$ changes must be a contraction, so that τ_i is necessarily $\pm\frac{1}{2}$ when $k_\ell < i < k_{\ell+1}$; the value of τ_{k_ℓ}, however, can be 1 or $\pm\frac{1}{2}$. Since only $\mathbf{x}_3$ is changing between iterations k_ℓ and $k_{\ell+1}$, relation (5.3) implies that

$$\mathbf{x}_3^{(k_\ell+j)} = \tfrac{1}{2}\mathbf{x}_2^{(k_\ell)} + (-1)^{j-1} \prod_{i=0}^{j-1} \tau_{k_\ell+i} \,\big(\tfrac{1}{2}\mathbf{x}_2^{(k_\ell)} - \mathbf{x}_3^{(k_\ell)}\big) \quad \text{for} \quad j = 1, \ldots, k_{\ell+1} - k_\ell - 1. \tag{5.4}$$

Using (5.1), (5.2) and (5.4), we obtain an expression representing $\tilde{\mathbf{x}}_2^{(\ell+1)}$ entirely in terms of $\tilde{\mathbf{x}}_2^{(\ell)}$ and $\tilde{\mathbf{x}}_2^{(\ell-1)}$:

$$\tilde{\mathbf{x}}_2^{(\ell+1)} = \tfrac{1}{2}\tilde{\mathbf{x}}_2^{(\ell)} + \tilde{\tau}_\ell\big(\tfrac{1}{2}\tilde{\mathbf{x}}_2^{(\ell)} - \tilde{\mathbf{x}}_2^{(\ell-1)}\big), \tag{5.5}$$

where

$$\tilde{\tau}_\ell = (-1)^{\tilde{\ell}} \prod_{i=0}^{\tilde{\ell}} \tau_{k_\ell+i}, \quad \text{with} \quad \tilde{\ell} = k_{\ell+1} - k_\ell - 1.$$

Because reflections cannot occur between iterations k_ℓ and $k_{\ell+1}$, we know that $|\tilde{\tau}_\ell| \le \frac{1}{2}$ or $\tilde{\tau}_\ell = 1$. (The latter happens only when iterations k_ℓ and $k_{\ell+1}$ are consecutive).

Using matrix notation, we have

$$\tilde{\mathbf{x}}_2^{(\ell)} = \begin{pmatrix} \tilde{x}_{21}^{(\ell)} \\ \tilde{x}_{22}^{(\ell)} \end{pmatrix} = \begin{pmatrix} u_\ell \\ v_\ell \end{pmatrix}, \quad \text{so that, using (5.1),} \quad \tilde{\mathbf{x}}_3^{(\ell)} = \tilde{\mathbf{x}}_2^{(\ell-1)} = \begin{pmatrix} u_{\ell-1} \\ v_{\ell-1} \end{pmatrix}. \tag{5.6}$$

The Nelder-Mead update embodied in (5.5) can be written as a matrix recursion in u and v:

$$\begin{pmatrix} u_{\ell+1} & v_{\ell+1} \\ u_\ell & v_\ell \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(1+\tilde{\tau}_\ell) & -\tilde{\tau}_\ell \\ 1 & 0 \end{pmatrix} \begin{pmatrix} u_\ell & v_\ell \\ u_{\ell-1} & v_{\ell-1} \end{pmatrix}. \tag{5.7}$$

Define $\mathbf{u}_\ell$ and $\mathbf{v}_\ell$ by

$$\mathbf{u}_\ell = \begin{pmatrix} u_\ell \\ u_{\ell-1} \end{pmatrix} \quad \text{and} \quad \mathbf{v}_\ell = \begin{pmatrix} v_\ell \\ v_{\ell-1} \end{pmatrix},$$

so that $\mathbf{u}_\ell$ contains the x-coordinates of the current second-worst and worst vertices, $\tilde{\mathbf{x}}_2^{(\ell)}$ and $\tilde{\mathbf{x}}_3^{(\ell)}$, and $\mathbf{v}_\ell$ contains their y coordinates. The desired conclusion of Lemma 5.1 follows if we can show that

$$\lim_{\ell\to\infty} \mathbf{u}_\ell = 0 \quad \text{and} \quad \lim_{\ell\to\infty} \mathbf{v}_\ell = 0. \tag{5.8}$$

We shall prove only the first relation in (5.8); the proof of the second is similar.

(ii) Measuring progress toward the origin. To prove convergence of $\mathbf{u}_\ell$ to the origin, it might appear that we could simply apply norm inequalities to the matrix equation (5.7). Unfortunately, the two-norm of the matrix in (5.7) exceeds one for all valid $\tilde{\tau}_\ell$, which means that $\|\mathbf{u}_{\ell+1}\|$ can be larger than $\|\mathbf{u}_\ell\|$. Hence we need to find a suitable nonincreasing size measure associated with each Nelder-Mead iteration (5.7).

Such a size measure is given by a positive definite quadratic function Q of two scalar arguments (or, equivalently, of a 2-vector):

$$Q(a,b) = 2(a^2 - ab + b^2) = a^2 + b^2 + (a-b)^2. \tag{5.9}$$

Evaluating $Q(\mathbf{u}_{\ell+1})$ using (5.7) gives

$$Q(\mathbf{u}_{\ell+1}) = (\tfrac{3}{2} + \tfrac{1}{2}\tilde{\tau}_\ell^2)u_\ell^2 - 2\tilde{\tau}_\ell^2 u_\ell u_{\ell-1} + 2\tilde{\tau}_\ell^2 u_{\ell-1}^2.$$

After substitution and manipulation, we obtain

$$Q(\mathbf{u}_\ell) - Q(\mathbf{u}_{\ell+1}) \;=\; 2(1-\tilde{\tau}_\ell^2)(\tfrac{1}{2}u_\ell - u_{\ell-1})^2, \tag{5.10}$$

which shows that

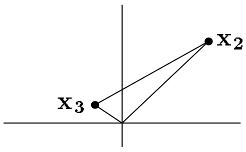
$$Q(\mathbf{u}_{\ell+1}) \;\le\; Q(\mathbf{u}_\ell) \quad \text{when} \quad -1 \le \tilde{\tau}_\ell \le 1. \tag{5.11}$$

It follows that Q is, as desired, a size measure that is nonincreasing for all valid values of $\tilde{\tau}_\ell$. Furthermore, because Q is positive definite, we can prove that $\mathbf{u}_\ell \to 0$ by showing that $Q(\mathbf{u}_\ell) \to 0$.

An obvious and appealing geometric interpretation of Q in terms of the Nelder-Mead simplices is that the quantity $Q(\mathbf{u}_\ell) + Q(\mathbf{v}_\ell)$ is the sum of the squared side lengths of the Nelder-Mead triangle at iteration k_ℓ, with vertices at the origin, $\tilde{\mathbf{x}}_2^{(\ell)}$, and $\tilde{\mathbf{x}}_3^{(\ell)}$. Relation (5.11) indicates that, after a reflection or contraction in which $\mathbf{x}_2$ changes, the sum of the squared side lengths of the new Nelder-Mead triangle cannot increase, even though $\|\mathbf{u}_{\ell+1}\|$ may be larger. Figure 4 depicts an example in which, after an outside contraction, both $\|\mathbf{u}_{\ell+1}\|$ and $\|\mathbf{v}_{\ell+1}\|$ increase. Nonetheless, the sum of the squared triangle side lengths is reduced.

(iii) Singular values in a subspace. To obtain worst-case bounds on the size of Q, it is convenient to interpret Q as the two-norm of a specially structured 3-vector derived from $\mathbf{u}_\ell$. Within the context of a Nelder-Mead iteration (5.6), we use the notation

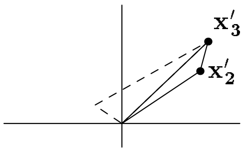
$$\boldsymbol{\xi}_\ell = \begin{pmatrix} u_\ell \\ u_{\ell-1} \\ u_\ell - u_{\ell-1} \end{pmatrix}, \quad \text{so that} \quad Q(\mathbf{u}_\ell) = \|\boldsymbol{\xi}_\ell\|^2. \tag{5.12}$$



sum of squared sides = 3.895

$x_{21}^2 + x_{31}^2 = 0.9925$

$x_{22}^2 + x_{32}^2 = 0.85$



sum of squared sides = 2.9003

$(x'_{21})^2 + (x'_{31})^2 = 1.646$

$(x'_{22})^2 + (x'_{32})^2 = 1.1406$

Figure 4: A triangle and its outside contraction.

The structure of $\boldsymbol{\xi}$ (5.12) can be formalized by observing that it lies in the two-dimensional null space of the vector $(1, -1, -1)$. Let Z denote the following 3×2 matrix whose columns form a (nonunique) orthonormal basis for this null space:

$$Z = \begin{pmatrix} z_1 & z_2 \end{pmatrix}, \quad \text{where} \quad z_1 = \frac{1}{\sqrt{6}} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad z_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix},$$

and let $\mathbf{q}_\ell$ denote the unique 2-vector satisfying

$$\boldsymbol{\xi}_\ell = Z\mathbf{q}_\ell = \begin{pmatrix} u_\ell \\ u_{\ell-1} \\ u_\ell - u_{\ell-1} \end{pmatrix}. \tag{5.13}$$

Since $Z^T Z = I$, we have

$$\|\boldsymbol{\xi}_\ell\| = \|\mathbf{q}_\ell\| \quad \text{and} \quad Q(\mathbf{u}_\ell) = \|\boldsymbol{\xi}_\ell\|^2 = \|\mathbf{q}_\ell\|^2, \tag{5.14}$$

so that we may use $\|\mathbf{q}_\ell\|$ to measure Q.

The Nelder-Mead move (5.7) can be written in terms of a 3×3 matrix M_ℓ applied to $\boldsymbol{\xi}_\ell$:

$$\boldsymbol{\xi}_{\ell+1} = M_\ell \boldsymbol{\xi}_\ell, \quad \text{where} \quad M_\ell = \begin{pmatrix} \frac{1}{2}(1+\tilde{\tau}_\ell) & -\tilde{\tau}_\ell & 0 \\ 1 & 0 & 0 \\ -\frac{1}{2} & -\frac{1}{2}\tilde{\tau}_\ell & \frac{1}{2}\tilde{\tau}_\ell \end{pmatrix}. \tag{5.15}$$

As we have already shown, the special structure of the vector $\boldsymbol{\xi}_\ell$ constrains the effects of the transformation M_ℓ to a subspace. To analyze these effects, note that, by construction of M_ℓ, its application to any vector in the column space of Z produces a vector in the same column space, i.e.,

$$M_\ell Z = Z W_\ell, \quad \text{where} \quad W_\ell = Z^T M_\ell Z. \tag{5.16}$$

A single Nelder-Mead move (5.7) is thus given by

$$\boldsymbol{\xi}_{\ell+1} = M_\ell \boldsymbol{\xi}_\ell = M_\ell Z \mathbf{q}_\ell = Z W_\ell \mathbf{q}_\ell,$$

so that, using (5.14),

$$Q(\mathbf{u}_{\ell+1}) = \|\boldsymbol{\xi}_{\ell+1}\|^2 = \|W_\ell \mathbf{q}_\ell\|^2,$$

and we may deduce information about the behavior of Q from the structure of the 2×2 matrix W.

Direct calculation shows that, for any $\tilde{\tau}_\ell$, W_ℓ is the product of an orthonormal matrix $\tilde{Z}$ and a diagonal matrix:

$$W_\ell = \tilde{Z}\Sigma_\ell, \quad \text{where} \quad \tilde{Z} = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} \quad \text{and} \quad \Sigma_\ell = \begin{pmatrix} 1 & 0 \\ 0 & -\tilde{\tau}_\ell \end{pmatrix}, \tag{5.17}$$

with $\tilde{Z}$ representing a rotation through 60 degrees. The form (5.17), analogous to the singular value decomposition apart from the possibly negative diagonal element of Σ_ℓ, reveals that the extreme values of $||W_\ell \mathbf{q}_\ell||$ are

$$\max_{||\mathbf{q}||=1} ||W_\ell \mathbf{q}|| = 1 \quad \text{when} \quad \mathbf{q} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \text{and}$$
$$\min_{||\mathbf{q}||=1} ||W_\ell \mathbf{q}|| = |\tilde{\tau}_\ell| \quad \text{when} \quad \mathbf{q} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{5.18}$$

For a reflection ($\tilde{\tau}_\ell = 1$), the value of Q is unchanged for all $\mathbf{q}$, and hence for all $\mathbf{u}$. When $|\tilde{\tau}_\ell| = \frac{1}{2}$, relationship (5.13) indicates how the extremes of (5.18) map into $\mathbf{u}$-space. The value of Q remains constant, i.e., $Q(\mathbf{u}_{\ell+1}) = Q(\mathbf{u}_\ell)$, only when $\mathbf{u}_\ell$ has the form $(2\alpha, \alpha)$ for some nonzero α; this can also be seen directly in (5.10). The maximum reduction in Q, by a factor of $\tilde{\tau}_\ell^2$, occurs only when $\mathbf{u}_\ell$ has the form $(0, \alpha)$ for some nonzero α.

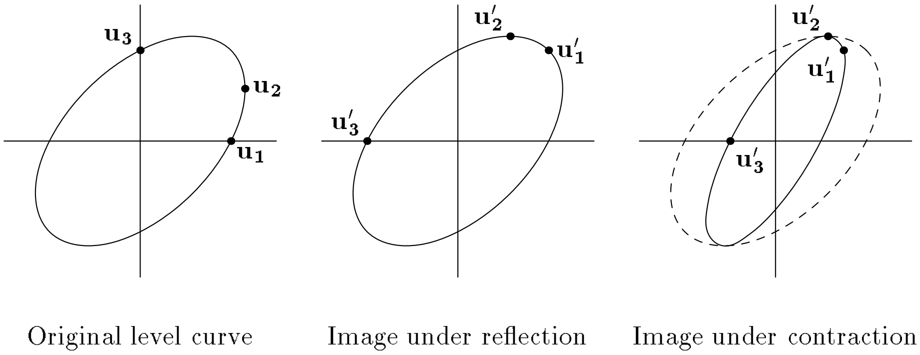


Figure 5: The effects of reflection and contraction moves in $\mathbf{u}$-space on a level curve of constant Q.

A geometric interpretation of reflection and contraction moves is depicted in Figure 5. The plane in each case represents $\mathbf{u}$-space. The first figure shows an elliptical level curve of points $(u_\ell, u_{\ell-1})$ for which $Q = 2$; three particular points on the level curve are labeled as $\mathbf{u}_i$. The second figure shows the image of this level curve following the reflection move (5.7) with $\tilde{\tau} = 1$. Points on the level curve are transformed by a reflection to rotated points on the same level curve; the image points of $\mathbf{u}_i$ are labeled as $\mathbf{u}'_i$. The third figure shows the image of the level curve in the first figure after a Nelder-Mead contraction move (5.7) with $\tilde{\tau} = \frac{1}{2}$. The transformed points are not only rotated, but their Q-values are (except for two points) reduced. The points $\mathbf{u}_2 = (2/\sqrt{3},\, 1/\sqrt{3})$ and $\mathbf{u}_3 = (0,\, 1)$ represent the extreme effects of contraction, since $Q(\mathbf{u}'_2) = Q(\mathbf{u}_2)$, and $Q(\mathbf{u}'_3) = \frac{1}{4}Q(\mathbf{u}_3)$.

Our next step is to analyze what can happen to the value of Q following a *sequence* of Nelder-Mead iterations, and to show that even in the worst case Q must eventually be driven to zero. Relation (5.17) implies that, for any vector $\mathbf{q}$,

$$||W_j \mathbf{q}|| \le ||W_k \mathbf{q}|| \quad \text{if} \quad |\tilde{\tau}_j| \le |\tilde{\tau}_k|.$$

In determining upper bounds on Q, we therefore need to consider only the two values $\tilde{\tau}_\ell = 1$ and $\tilde{\tau}_\ell = \frac{1}{2}$ (the latter corresponding to the largest possible value of $|\tilde{\tau}|$ when $\tilde{\tau} \ne 1$).

Using (5.16) repeatedly to move Z to the left, we express a sequence of N Nelder-Mead moves (5.7) starting at iteration ℓ as

$$\boldsymbol{\xi}_{\ell+N} = M_{\ell+N-1} \cdots M_\ell Z \mathbf{q}_\ell = Z W_{\ell+N-1} \cdots W_\ell \mathbf{q}_\ell.$$

Substituting for each W from (5.17), the Euclidean length of $\mathbf{q}_{\ell+N}$ is bounded by:

$$||\mathbf{q}_{\ell+N}|| \le ||\tilde{Z}\Sigma_{\ell+N-1} \cdots \tilde{Z}\Sigma_\ell||\, ||\mathbf{q}_\ell||. \tag{5.19}$$

A relatively straightforward calculation shows that $||\mathbf{q}_{\ell+N}||$ must be strictly smaller than $||\mathbf{q}_\ell||$ after any ofthe move sequences:

$$(c, c) \text{ for } N = 2, \qquad (c, 1, c) \text{ for } N = 3,$$
$$(c, 1, 1, 1, c) \text{ for } N = 5, \qquad (c, 1, 1, 1, 1, c) \text{ for } N = 6, \tag{5.20}$$

where "c" denotes $\tilde{\tau} = \frac{1}{2}$ and "1" denotes $\tilde{\tau} = 1$. For these sequences,

$$\|\mathbf{q}_{\ell+N}\| \;\le\; \beta_{cc}\,\|\mathbf{q}_\ell\|, \quad \text{where } \beta_{cc} \approx 0.7215.$$

(The quantity β_{cc} is the larger root of the quadratic $\lambda^2 + \frac{41}{64}\lambda + \frac{1}{16}$.) Following any of the Nelder-Mead type patterns (5.20), the size measure Q must be decreased by a factor of at least $\beta_{cc}^2 \approx 0.5206$.

(iv) Illegal patterns of Nelder-Mead move types. At this point we add the final element of the proof: certain patterns of Nelder-Mead move types cannot occur in the subsequence (5.1). Recall that a new point can be accepted only when its function value is strictly less than the current worst function value. Now consider five consecutive Nelder-Mead iterations (5.7) of types $(1, 1, \tilde{\tau}_3, 1, 1)$ in which $\mathbf{x}_2$ changes. After such a pattern, the newly accepted vertex is defined by

$$\begin{pmatrix} u_{\ell+5} & v_{\ell+5} \\ u_{\ell+4} & v_{\ell+4} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}^2 \begin{pmatrix} \frac{1}{2}(1+\tilde{\tau}_3) & -\tilde{\tau}_3 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}^2 \begin{pmatrix} u_\ell & v_\ell \\ u_{\ell-1} & v_{\ell-1} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 1 \\ -\tilde{\tau}_3 & \frac{1}{2}(1+\tilde{\tau}_3) \end{pmatrix} \begin{pmatrix} u_\ell & v_\ell \\ u_{\ell-1} & v_{\ell-1} \end{pmatrix}. \tag{5.21}$$

The first row of this relation gives

$$(u_{\ell+5}, v_{\ell+5}) = (u_{\ell-1}, v_{\ell-1}), \quad \text{so that} \quad \tilde{\mathbf{x}}_2^{(\ell+5)} = \tilde{\mathbf{x}}_3^{(\ell)},$$

which implies the impossible result that the newly accepted vertex is the same as the worst vertex in a previous simplex. Hence the type sequence $(1, 1, \tilde{\tau}_3, 1, 1)$ cannot occur.

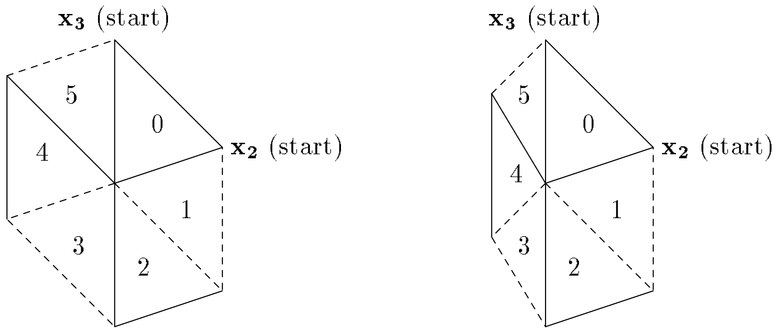


Figure 6: Returning to the original worst point with Nelder-Mead type patterns $(1, 1, 1, 1, 1)$ and $(1, 1, \frac{1}{2}, 1, 1)$.

Figure 6 depicts these unacceptable move sequences geometrically: on the left, five consecutive reflections (illegally) return to the original worst point; on the right, the original worst point is also repeated if there are two reflections, an outside contraction, and two more reflections.

If we eliminate both the norm-reducing patterns (5.20) and the illegal pattern $(1, 1, *, 1, 1)$, only three valid 6-move sequences remain during which Q might stay unchanged:

$$(1, 1, 1, 1, c, 1), \quad (1, c, 1, 1, 1, 1), \quad \text{and} \quad (1, c, 1, 1, c, 1).$$

Examination of these three cases shows immediately that *no legal sequence* of 7 steps exists for which Q can remain constant, since the next move creates either a norm-reducing or illegal pattern. In particular, for all legal sequences of 7 steps it holds that

$$\|\mathbf{q}_{\ell+7}\| \;\le\; \beta_{cc}\|\mathbf{q}_\ell\| \;<\; 0.7216\|\mathbf{q}_\ell\|.$$

We conclude that $\|\mathbf{q}_\ell\| \to 0$, and hence, using (5.14), that $Q(\mathbf{u}_\ell) \to 0$, as desired. This completes the proof of Lemma 5.1. ∎

To finish the proof of Theorem 5.1, we note that, in the case when $\mathbf{x}_1^{(k)}$ eventually becomes constant, the just-completed proof of Lemma 5.1 implies convergence of $\mathbf{x}_2$ and $\mathbf{x}_3$ to $\mathbf{x}_1$, which gives $f_1^* = f_2^* = f_3^*$, as desired. ∎

5.2. Convergence of simplex diameters to zero

Knowing that the vertex function values converge to a common value does not imply that the vertices themselves converge. We next analyze the evolution of the shapes of the triangles Δ_k produced by the Nelder-Mead algorithm on a strictly convex function in $\mathcal{R}^2$. First, we show that they "collapse" to zero volume, i.e., to either a point or a line segment.

Lemma 5.2. (Convergence of simplex volumes to zero.) *Let f be a strictly convex function on $\mathcal{R}^2$ with bounded level sets. Assume that the Nelder-Mead algorithm with reflection coefficient $\rho = 1$, expansion coefficient $\chi = 2$, and contraction coefficient $\gamma = \frac{1}{2}$ is applied to f beginning with a nondegenerate initial simplex Δ_0. Then the simplices $\{\Delta_k\}$ generated by the algorithm satisfy*

$$\lim_{k\to\infty} \mathrm{vol}(\Delta_k) = 0. \tag{5.22}$$

Proof. We know from Theorem 5.1 that the limiting function values at the vertices are equal, say to f^*. If $f^* = f_{\min}$, then by strict convexity this value is assumed at a unique point, in which case the desired result (5.22) follows immediately and the proof is complete. Furthermore, Lemma 5.1 shows that, if the best vertex $\mathbf{x}_1$ eventually becomes constant, then the remaining two vertices converge to $\mathbf{x}_1$, and (5.22) holds in this case also.

In the rest of the proof we assume that $f^* > f_{\min}$ and that $\mathbf{x}_1$ changes infinitely often. Corresponding to f^*, we define the level set $\mathcal{L}_*$ and its boundary Γ_*:

$$\mathcal{L}_* = \{\mathbf{x} \mid f(\mathbf{x}) \le f^*\} \quad \text{and} \quad \Gamma_* = \{\mathbf{x} \mid f(\mathbf{x}) = f^*\}. \tag{5.23}$$

It follows from our assumptions about f that $\mathcal{L}_*$ is nonempty, closed, and strictly convex.

The proof is obtained by contradiction. Suppose that (5.22) does not hold, so that

$$\limsup_{k\to\infty} \mathrm{vol}(\Delta_k) > 0. \tag{5.24}$$

We know that all Nelder-Mead simplices Δ_k lie inside the compact level set $\{\mathbf{x} \mid f(\mathbf{x}) \le f(\mathbf{x}_3^{(0)})\}$, and that all vertex function values converge to f^*. Hence we can extract at least one subsequence $\{k_j\}$ of iterations such that the simplices Δ_{k_j} satisfy

$$\lim_{j\to\infty} \Delta_{k_j} = T, \tag{5.25}$$

where T is a triangle of *nonzero volume* whose vertices all lie on Γ_*.

Next we consider properties of the set $\mathcal{T}_*$ of all triangles T satisfying (5.25) for some subsequence k_j. Since shrink steps cannot occur, a Nelder-Mead iteration on a given triangle is specified by two values: a distinguished (worst) vertex and a move type τ, where τ is one of $(1, 2, \frac{1}{2}, -\frac{1}{2})$. For each sequence k_j satisfying (5.25) with limit triangle T, there is a sequence of pairs of distinguished vertices and move types associated with moving from Δ_{k_j} to the next simplex Δ_{k_j+1}. For any such pair that occurs infinitely often in the sequence of iterations $\{k_j\}$, the vertices of Δ_{k_j+1}, the successor simplices, are a continuous function of the vertices of Δ_{k_j}. Since all limit vertex function values are equal to f^*, so that all limit vertices lie on Γ_*, there is a subsequence $\{k_{j_i}\}$ of $\{k_j\}$ such that

$$\lim_{i\to\infty} \Delta_{k_{j_i}+1} \to \widetilde{T} \in \mathcal{T}_*.$$

We conclude that for every triangle T in $\mathcal{T}_*$, there is a Nelder-Mead move which, applied to T, yields a new triangle $\widetilde{T}$ in $\mathcal{T}_*$. A similar argument shows that every triangle in $\mathcal{T}_*$ is the result of applying a Nelder-Mead move to another triangle in $\mathcal{T}_*$.

Next we consider sequences of possible Nelder-Mead moves among elements of $\mathcal{T}_*$. Observe first that no move of type $-\frac{1}{2}$ (inside contraction) can occur, since the new vertex would lie inside the convex hull of the original three vertices, contradicting the fact that the original three vertices and the new vertex must lie on Γ_*.

The volumes of triangles in $\mathcal{T}_*$ are bounded above because all vertices of such triangles lie on the boundary of $\mathcal{L}_*$. Define

$$V = \sup\{\, \mathrm{vol}(T) \mid T \in \mathcal{T}_* \,\}, \tag{5.26}$$

where $V > 0$ because of assumption (5.24), and choose a triangle T' in $\mathcal{T}_*$ for which

$$\operatorname{vol}(T') > \tfrac{1}{2}V. \tag{5.27}$$

Let V_* be the volume of the level set $\mathcal{L}_*$, and define the integer h_* as

$$h_* = 1 + \left\lceil \frac{V_*}{V} \right\rceil. \tag{5.28}$$

Now consider all sequences of h_* *consecutive* simplices produced by the Nelder-Mead algorithm applied to the initial simplex Δ_0,

$$\Delta_{r+1}, \Delta_{r+2}, \ldots, \Delta_{r+h_*}, \tag{5.29}$$

and define a sequence $\{T_i\}$ of h_* triangles in $\mathcal{T}_*$, ending with the triangle T' of (5.27), by extracting a subsequence $\{m_j\}$ for which

$$\lim \Delta_{m_j+i} = T_i \quad \text{for } i = 1, \ldots, h_*, \quad \text{with} \quad T_i \in \mathcal{T}_* \quad \text{and} \quad T_{h_*} = T'. \tag{5.30}$$

During any sequence of consecutive Nelder-Mead moves of type 1 (reflections), volume is preserved (see Lemma 3.1) and all triangles are disjoint; no triangle can be repeated because of the strict decrease requirement on the vertex function values. Suppose that there is a sequence of consecutive reflections in the set of iterations $m_j + 1, \ldots, m_j + h_*$; then the associated limiting triangles have disjoint interiors, cannot repeat, and lie inside the curve Γ_*. Since the volume enclosed by Γ_* is V_*, there can be at most $h_* - 1$ consecutive reflections (see (5.28)), and it follows that, for some i, the move from T_i to T_{i+1} is not a reflection.

Consider the first predecessor T_i of T_{h_*} in the sequence (5.30) for which $\operatorname{vol}(T_i) \neq \operatorname{vol}(T_{h_*})$. The Nelder-Mead move associated with moving from T_i to T_{i+1} cannot be a contraction; if it were, then

$$\operatorname{vol}(T_i) = 2\operatorname{vol}(T') > V,$$

which is impossible by definition of V (5.26) and T' (5.27). Thus, in order to satisfy (5.30), the move from T_i to T_{i+1} must be an expansion step, i.e., a move of type 2. We now show that this is impossible because of the strict convexity of $\mathcal{L}_*$ and the logic of the Nelder-Mead algorithm.

For the sequence $\{m_j\}$ of (5.30), the function value at the (accepted) expansion point must satisfy $f(\mathbf{x}_e^{(m_j+i)}) \geq f^*$, since the function values at all vertices converge from above to f^*. The reflection point $\mathbf{r}_i$ for T_i is outside T_i and lies strictly inside the triangle T_{i+1}, all of whose vertices lie on the curve Γ_*. (See Figure 7.) Since the level set $\mathcal{L}_*$ is strictly convex, $f(\mathbf{r}_i)$ must be strictly less than f^*, the value of f on Γ_*, and there must be a small open ball around $\mathbf{r}_i$ within which the values of f are strictly less than f^*.

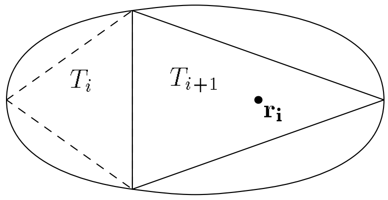


Figure 7: Position of the reflection point $\mathbf{r}_i$ when the vertices of T_i and the expansion point (the new vertex of T_{i+1}) lie on the boundary of a bounded strictly convex set.

The test reflection points $\mathbf{x}_r^{(m_j+i)}$ converge to $\mathbf{r}_i$, and hence eventually $f(\mathbf{x}_r^{(m_j+i)})$ must be strictly less than f^*. It follows that the Nelder-Mead algorithm at step $m_j + i$ could have chosen a new point (the reflection point) with a lower function value than at the expansion point, but failed to do so; this is impossible, since the Nelder-Mead method accepts the better of the reflection and expansion points. (Note that this conclusion would *not* follow for the Nelder-Mead algorithm in the original paper [6], where the expansion point could be chosen as the new vertex even if the value of f at the reflection point were smaller.) Thus we have shown that the assumption $\limsup_{k\to\infty} \operatorname{vol}(\Delta_k) > 0$

leads to a contradiction. This gives the desired result that the Nelder-Mead simplex volumes converge to zero. ∎

Having shown that the simplex volumes converge to zero, we now prove that the *diameters* converge to zero, so that the Nelder-Mead simplices collapse to a point.

Theorem 5.2. (Convergence of simplex diameters to zero.) *Let f be a strictly convex function on $\mathcal{R}^2$ with bounded level sets. Assume that the Nelder-Mead algorithm with reflection coefficient $\rho = 1$, expansion coefficient $\chi = 2$, and contraction coefficient $\gamma = \frac{1}{2}$ is applied to f beginning with a nondegenerate initial simplex Δ_0. Then the simplices $\{\Delta_k\}$ generated by the algorithm satisfy*

$$\lim_{k\to\infty} \operatorname{diam}(\Delta_k) = 0. \tag{5.31}$$

Proof. The proof is by contradiction. Lemma 5.2 shows that $\operatorname{vol}(\Delta_k) \to 0$. Since reflection preserves volume, infinitely many non-reflection steps must occur.

Suppose that the conclusion of the theorem is not true, i.e., that $\operatorname{diam}(\Delta_k)$ does not converge to zero. Then we can find a infinite subsequence $\{k_j\}$ for which the associated simplices Δ_{k_j} have diameters bounded away from zero, so that

$$\operatorname{diam}(\Delta_{k_j}) \geq \alpha > 0. \tag{5.32}$$

For each k_j in this subsequence, consider the sequence of iterations k_j, k_j+1, …, and let k_j' denote the first iteration in this sequence that immediately precedes a non-reflection step. Then the simplex $\Delta_{k_j'}$ is congruent to Δ_{k_j}, so that $\operatorname{diam}(\Delta_{k_j'}) \geq \alpha$, and a non-reflection step occurs when moving from $\Delta_{k_j'}$ to $\Delta_{k_j'+1}$.

Now we define a subsequence k_j'' of k_j' such that:

1. $\Delta_{k_j''}$ converges to a fixed line segment $[\mathbf{v}_0, \mathbf{v}_1]$, with $\mathbf{v}_0 \neq \mathbf{v}_1$ and $\|\mathbf{v}_1 - \mathbf{v}_0\|_2 \geq \alpha$; and
2. each Nelder-Mead step from $\Delta_{k_j''}$ to $\Delta_{k_j''+1}$ has the same combination of distinguished (worst) vertex and move type among the nine possible pairs of three vertices and three non-reflection moves.

Note that the vertices of $\Delta_{k_j''+1}$ are continuous functions of the vertices of $\Delta_{k_j''}$, and that the values of f at all vertices of $\Delta_{k_j''+1}$ must converge monotonically from above to f^*.

The points $\mathbf{v}_0$ and $\mathbf{v}_1$ must lie on the boundary of the strictly convex level set $\mathcal{L}_*$ (5.23). If the vertices of $\Delta_{k_j''}$ converge to three distinct points on the line segment $[\mathbf{v}_0, \mathbf{v}_1]$, strict convexity would imply that the function value at the interior point is strictly less than f^*, which is impossible. Thus two of the three vertices must converge to one of $\mathbf{v}_0$ and $\mathbf{v}_1$, which means that two of the vertices of $\Delta_{k_j''}$ will eventually lie close to one of $\mathbf{v}_0$ or $\mathbf{v}_1$. Without loss of generality we assume that two of the vertices are near $\mathbf{v}_0$ and the remaining vertex is near $\mathbf{v}_1$.

To obtain a contradiction, we show that all non-reflection steps are unacceptable.

(i) An inside contraction applied to $\Delta_{k_j''}$ with distinguished vertex near $\mathbf{v}_0$ produces a (limit) vertex for $\Delta_{k_j''+1}$ at $\frac{3}{4}\mathbf{v}_0 + \frac{1}{4}\mathbf{v}_1$; an inside contraction with distinguished vertex near $\mathbf{v}_1$ produces a limit vertex at $\frac{1}{2}\mathbf{v}_0 + \frac{1}{2}\mathbf{v}_1$. In either case, the limit vertex for $\Delta_{k_j''+1}$ lies strictly between $\mathbf{v}_0$ and $\mathbf{v}_1$, giving a function value smaller than f^*, a contradiction.

(ii) An outside contraction applied to $\Delta_{k_j''}$ with distinguished vertex near $\mathbf{v}_0$ produces a limit vertex for $\Delta_{k_j''+1}$ at $\frac{1}{4}\mathbf{v}_0 + \frac{3}{4}\mathbf{v}_1$, giving a contradiction as in (i). With distinguished vertex near $\mathbf{v}_1$, an outside contraction produces a limit vertex at $-\frac{1}{2}\mathbf{v}_1 + \frac{3}{2}\mathbf{v}_0$. Since $\mathbf{v}_0$ and $\mathbf{v}_1$ lie on the boundary of the strictly convex set $\mathcal{L}_*$, this limit vertex point lies outside the level set and hence has function value greater than f^*. This contradicts the fact that the associated vertex function values in $\Delta_{k_j''+1}$ must converge to f^*.

(iii) An expansion step with distinguished vertex near $\mathbf{v}_0$ produces a limit vertex for $\Delta_{k_j''+1}$ at $3\mathbf{v}_1 - 2\mathbf{v}_0$, and an expansion step with distinguished vertex near $\mathbf{v}_1$ produces a limit vertex at $3\mathbf{v}_0 - 2\mathbf{v}_1$. In both cases, the limit vertex lies outside $\mathcal{L}_*$. This means that its function value exceeds f^*, giving a contradiction.

Since a contradiction arises from applying every possible non-reflection move to the simplex $\Delta_{k_j''}$, the sequence k_j of (5.32) cannot exist. Thus we have shown that $\lim \operatorname{diam}(\Delta_k) \to 0$, namely that each Nelder-Mead simplex eventually collapses to a point. ∎

Note that this theorem does *not* imply that the sequence of simplices $\{\Delta_k\}$ converges to a limit point $\mathbf{x}_*$. We do know, however, that all vertices converge to $\mathbf{x}_1$ if this vertex remains constant (see Lemma 5.1); this situation occurs in the McKinnon examples [5].

6. Conclusions and Open Questions

In dimension 1, the generic Nelder-Mead method converges to the minimizer of a strictly convex function with bounded level sets if and only if the expansion step is a genuine expansion (i.e., if $\rho\chi \geq 1$). It is interesting that, apart from this further requirement, the conditions (2.1) given in the original Nelder-Mead paper suffice to ensure convergence in one dimension. The behavior of the algorithm in dimension 1 can nonetheless be very complicated; for example, there can be an infinite number of expansions even when convergence is M-step linear (Theorem 4.2).

In two dimensions, the behavior of even the standard Nelder-Mead method (with $\rho = 1$, $\chi = 2$, and $\gamma = \frac{1}{2}$) is more difficult to analyze for two reasons:

1. The space of simplex shapes is not compact, where the *shape* of a simplex is its similarity class; see the discussion at the end of Section 2. It appears that the Nelder-Mead moves are dense in this space, i.e., any simplex can be transformed by some sequence of Nelder-Mead moves to be arbitrarily close to any other simplex shape; this property reflects the intent expressed by Nelder and Mead [6] that the simplex shape should "adapt itself to the local landscape". This contrasts strongly with the nature of many pattern search methods [13], in which the simplex shapes remain constant.

2. The presence of the expansion step means that $\operatorname{vol}(\Delta)$ is not a Lyapunov function[3] for the iteration.

The two-dimensional results proved in Section 5 seem very weak, but conceivably represent the limits of what can be proved for arbitrary strictly convex functions. In particular, Theorem 5.2 leaves open the possibility that the ever-smaller simplices endlessly "circle" the contour line $f(x) = f^*$. Since no examples of this behavior are known, it may be possible to prove the stronger result that the simplices always converge to a *single* point $\mathbf{x}_*$.

An obvious question concerns *how* the Nelder-Mead method can fail to converge to a minimizer in the two-dimensional case. Further analysis suggests that, for suitable strictly convex functions (C^1 seems to suffice), failure can occur only if the simplices elongate indefinitely and their shape goes to "infinity" in the space of simplex shapes (as in the McKinnon counterexample).

An interesting open problem concerns whether there exists *any* function $f(\mathbf{x})$ in $\mathcal{R}^2$ for which the Nelder-Mead algorithm always converges to a minimizer. The natural candidate is $f(x, y) = x^2 + y^2$, which by affine invariance is equivalent to all strictly convex quadratic functions in two dimensions. A complete analysis of Nelder-Mead for $x^2 + y^2$ remains an open problem.

Our general conclusion about the Nelder-Mead algorithm is that the main mystery to be solved is not whether it ultimately converges to a minimizer—for general (nonconvex) functions, it does not—but rather why it tends to works so well in practice by producing a rapid initial decrease in function values.

Acknowledgments

The authors greatly appreciate the referee's detailed and constructive comments, which helped us to improve the content and presentation of the paper. We also thank Virginia Torczon for making

[3] See the discussion of Lyapunov functions in, for example, [11, pages 23–27] in the context of stability of nonlinear fixed points.

us aware of the connections described in Section 4.4. Margaret Wright is extremely grateful to Steve Fortune for many interesting discussions and his consistently enlightening geometric insights.

References

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

[2] C. T. Kelley (1997). Detection and remediation of stagnation in the Nelder-Mead algorithm using a sufficient decrease condition, Technical Report, Department of Mathematics, North Carolina State University, Raleigh, North Carolina, USA.

[3] J. C. Lagarias, B. Poonen, and M. H. Wright (1997). Convergence of the restricted Nelder-Mead algorithm in two dimensions, in preparation.

[4] Math Works (1994. Matlab Optimization Toolbox, The Math Works, Natick, Massachusetts, USA.

[5] K. I. M. McKinnon (1996). Convergence of the Nelder-Mead simplex method to a non-stationary point, Technical report, Department of Mathematics and Computer Science, University of Edinburgh, Edinburgh, UK. To appear, *SIAM J. Optimization*.

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

[7] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vettering (1988). *Numerical Recipes in C*, Cambridge University Press, Cambridge, UK.

[8] A. Rykov (1980a). Simplex direct search algorithms, *Automation and Robot Control* 41, 784–793.

[9] A. Rykov (1980b). Simplex methods of direct search, *Engineering Cybernetics* 18, 12–18.

[10] A. Rykov (1983). Simplex algorithms for unconstrained optimization, *Problems of Control and Information Theory* 12, 195–208.

[11] A. M. Stuart and A. R. Humphries (1996). *Dynamical Systems and Numerical Analysis*, Cambridge University Press, New York, USA.

[12] V. Torczon (1989). *Multi-directional Search: A Direct Search Algorithm for Parallel Machines.* Ph.D. thesis, Rice University, Houston, Texas, USA.

[13] V. Torczon (1997a). On the convergence of pattern search algorithms, *SIAM J. Optimization* 7, 1–25.

[14] V. Torczon (1997b). Private communication.

[15] P. Tseng (1995). Fortified-descent simplicial search method: a general approach, Technical report, Department of Mathematics, University of Washington, Seattle, Washington, USA.

[16] F. H. Walters, L. R. Parker, S. L. Morgan, and S. N. Deming (1991). *Sequential Simplex Optimization.* CRC Press, Boca Raton, USA.

[17] D. J. Woods (1985). *An Interactive Approach for Solving Multi-Objective Optimization Problems.* Ph.D. thesis, Rice University, Houston, Texas, USA.

[18] M. H. Wright (1996). Direct search methods: once scorned, now respectable. in *Numerical Analysis 1995: Proceedings of the 1995 Dundee Biennial Conference in Numerical Analysis*, D. F. Griffiths and G. A. Watson (eds.), 191–208, Addison Wesley Longman, Harlow, UK.