Nelder, Mead, and the Other Simplex Method

Margaret H. Wright

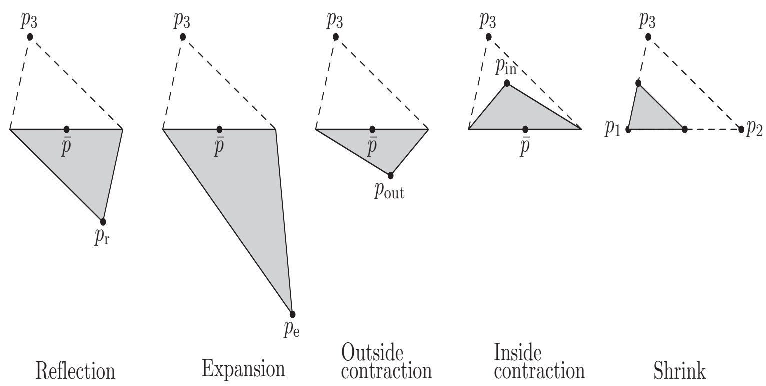
2010 Mathematics Subject Classification: 49M30, 65K10, 90C56
Keywords and Phrases: Nelder-Mead, direct search simplex method, derivative-free optimization, non-derivative optimization

In the mid-1960s, two English statisticians working at the National Vegetable Research Station invented the Nelder–Mead "simplex" direct search method. The method emerged at a propitious time, when there was great and growing interest in computer solution of complex nonlinear real-world optimization problems. Because obtaining first derivatives of the function f to be optimized was frequently impossible, the strong preference of most practitioners was for a "direct search" method that required only the values of f; the new Nelder–Mead method fit the bill perfectly. Since then, the Nelder–Mead method has consistently been one of the most used and cited methods for unconstrained optimization.

We are fortunate indeed that the late John Nelder[1] has left us a detailed picture of the method's inspiration and development [11, 14]. For Nelder, the starting point was a 1963 conference talk by William Spendley of Imperial Chemical Industries about a "simplex" method recently proposed by Spendley, Hext, and Himsworth for response surface exploration [15]. Despite its name, this method is not related to George Dantzig's simplex method for linear programming, which dates from 1947. Nonetheless, the name is entirely appropriate because the Spendley, Hext, and Himsworth method is defined by a simplex; the method constructs a pattern of $n+1$ points in dimension n, which moves across the surface to be explored, sometimes changing size, but always retaining the same shape.

Inspired by Spendley's talk, Nelder had what he describes as "one useful new idea": while defining each iteration via a simplex, add the crucial ingredient that the shape of the simplex should "adapt itself to the local landscape" [12]. During a sequence of lively discussions with his colleague Roger Mead, where "each of us [was] able to try out the ideas of the previous evening on the other the following morning", they developed a method in which the simplex could "elongate itself to move down long gentle slopes", or "contract itself on to the final minimum" [11]. And, as they say, the rest is history.

[1] 8 October 1924 – 7 August 2010.



The 1965 Nelder–Mead paper [12] appeared in the *Computer Journal*, a prestigious publication of the British Computer Society. Implementations and numerical testing followed almost immediately in which the Nelder–Mead method performed well compared to existing algorithms. In addition, one should not underestimate the degree to which the Nelder–Mead method appealed to practitioners because its moves are easy to describe. The Nelder–Mead simplex can change in five different ways during an iteration, as illustrated here in two dimensions. Except in the case of a shrink, the worst vertex of the simplex at iteration k (the point p_3 in the figure) is replaced at iteration $k + 1$ by one of the reflection, expansion, or contraction points. Based on this picture, users felt (and feel) that they understand what the method is doing. As Nelder said while trying to explain the method's popularity [11], "... the underlying ideas are extremely simple – you do not have to know what a Hessian matrix is to understand them".

Nelder's recollection of events [11] following publication of the Nelder–Mead paper is that some "professional optimizers" were "surprised" because they "had convinced themselves that direct search methods ... were basically unpromising". Nelder notes with relish that "our address (National Vegetable Research Station) also caused surprise in one famous US laboratory,[2] whose staff clearly doubted if turnipbashers could be numerate".

The Nelder–Mead paper has been cited thousands of times, and qualified by the late 1970s as a "Science Citation Classic". The Nelder–Mead method soon became so much more popular than other simplex-based methods that it began to be called "the" simplex method, in the context of unconstrained optimization.[3]

The story of the subsequent position of the Nelder–Mead method in mainstream optimization clearly illustrates a sea change, sometimes called "math-

[2] To the present author's knowledge, this laboratory has never been identified.

[3] Because the LP simplex method is much better known, the Nelder–Mead method is sometimes lightheartedly called "the other simplex method".

ematization", that has taken place since the 1960s and early 1970s. A 1972 survey paper by Swann [16, page 28] concludes by saying

> Although the methods described above have been developed heuristically and no proofs of convergence have been derived for them, in practice they have generally proved to be robust and reliable ...

The lack of theoretical foundations and motivation would almost certainly be regarded as unacceptable in an optimization journal today.

As optimization became more mathematical, by the late 1970s textbooks tended to dismiss the Nelder–Mead method (and other direct search methods) as "ad hoc" or "heuristic". Of course there were a small number of scholarly works about the Nelder–Mead method (see the references in [20, 6]). Among these, the analysis of [4] is of particular interest.

Of equal or (to some) greater concern, the Nelder–Mead method was well known to experience practical difficulties ranging from stagnation to failure. As a result, even in its early years papers were published that described how the Nelder–Mead method could be modified so that it would work well on a particular problem.

Although not center stage in mainstream optimization, direct search methods other than Nelder–Mead were being studied and implemented, especially in China and the Soviet Union, but the associated work was not well known in the West. (Several references to these papers are given in [20, 6].) This situation changed significantly in 1989, when Virginia Torczon, a PhD student at Rice University advised by John Dennis, published a thesis [17] that not only proposed a direct search method ("multidirectional search"), but also provided a proof that, under various conditions, $\liminf \|\nabla f\| \to 0$, where f is the function to be optimized.

Once rigorous convergence results had been established for one method, the floodgates opened, and since 1989 there has been a subsequent (and still ongoing) renaissance of interest in derivative-free methods. The level of intensity has been especially high for research on model-based derivative-free methods, which (unlike Nelder–Mead and other direct search methods) create evolving simple models of f. A nice discussion of the different classes of derivative-free methods can be found in [2].

How does the Nelder–Mead method fit into today's landscape of derivative-free methods? It is fair to describe Nelder–Mead as a far outlier, even a singularity, in the emerging families of mathematically grounded direct search methods such as generalized pattern search and generating set search [2]. Hence the position of the Nelder–Mead method in mainstream nonlinear optimization is anomalous at best, and is subject to a wide range of attitudes.

From the positive end, several researchers have created modified Nelder–Mead methods with the goal of retaining the favorable properties of the original while avoiding its known deficiencies. See, for example, [19, 5, 18, 10, 13, 1]. Strategies for remedying the defects of the original Nelder–Mead include using a "sufficient decrease" condition for acceptance of a new vertex (rather than simple decrease) and restarting when the current simplex becomes excessively ill-conditioned.

Taking a negative view, some researchers believe that Nelder–Mead is passé because modern derivative-free methods are consistently better:

> The Nelder-Mead algorithm, however, can work very well and it is expected to survive a very long time. Nevertheless, it is seriously defective: it is almost never the best method and indeed it has no general convergence results ... we believe that ultimately more sophisticated and successful methods will earn their rightful place in practical implementations ... [2, page 7].

Whichever view prevails in the long run, as of 2012 the Nelder–Mead method is not fading away. As in its early days, it remains remarkably popular with practitioners in a wide variety of applications. In late May 2012, Google Scholar displayed more than 2,000 papers *published in 2012* that referred to the Nelder–Mead method, sometimes when combining Nelder–Mead with other algorithms.

In addition, certain theoretical questions remain open about the original Nelder–Mead method. Why is it sometimes so effective (compared to other direct search methods) in obtaining a rapid improvement in f? One failure mode is known because Ken McKinnon produced a fascinating family of strictly convex functions in two dimensions for which Nelder–Mead executes an infinite sequence of repeated inside contractions and thereby fails to converge to the minimizer from a specified starting configuration [9] – but are there other failure modes? An initial exploration of the effects of dimensionality [3] provides some insights, but there is more to be learned. Why, despite its apparent simplicity, should the Nelder–Mead method be difficult to analyze mathematically? (See [7, 8].) One can argue that, before the original method is retired, we should achieve the maximum possible mathematical understanding of how and why it works.

In an interview conducted in 2000, John Nelder said about the Nelder–Mead method:

> There are occasions where it has been spectacularly good ... Mathematicians hate it because you can't prove convergence; engineers seem to love it because it often works.

And he is still right.

We end with a picture of John Nelder and George Dantzig, fathers of two different simplex methods, together at the 1997 SIAM annual meeting at Stanford University:

John Nelder and George Dantzig, Stanford University, 1997, photographed by Margaret Wright

References

[1] Burmen, A., Puhan, J., and Tuma, T., Grid restrained Nelder–Mead algorithm, *Computational Optimization and Applications* 34 (2006), 359–375.

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

[3] Han, L. and Neumann, M., Effect of dimensionality on the Nelder–Mead simplex method, *Optimization Methods and Software* 21 (2006), 1–16.

[4] Hensley, D., Smith, P., and Woods, D., Simplex distortions in Nelder–Mead reflections, IMSL Technical Report Series No. 8801, IMSL, Inc., Houston, Texas (1988).

[5] Kelley, C. T., Detection and remediation of stagnation in the Nelder–Mead algorithm using a sufficient decrease condition, *SIAM Journal on Optimization* 10 (1999), 43–55.

[6] Kolda, T. G., Lewis, R. M., and Torczon, V., Optimization by direct search: new perspectives on some classical and modern methods, *SIAM Review* 45 (2003), 385–482.

[7] Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E., Convergence properties of the Nelder–Mead simplex algorithm in low dimensions, *SIAM Journal on Optimization* 9 (1998), 112–147.

[8] Lagarias, J. C., Poonen, B., and Wright, M. H., Convergence of the restricted Nelder–Mead method in two dimensions, *SIAM Journal on Optimization* 22 (2012), 501–532.

[9] McKinnon, K. I. M., Convergence of the Nelder–Mead simplex method to a non-stationary point, *SIAM Journal on Optimization* 9 (1998), 148–158.

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

[11] Nelder, J. A., This week's citation classic, *Citation Classics Commentaries* 15 (1979).

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

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

[14] Senn, S., A conversation with John Nelder, *Statistical Science* 18 (2003), 118–131.

[15] Spendley, W., Hext, G. R., and Himsworth, F. R., Sequential application of simplex designs in optimization and Evolutionary Operation, *Technometrics* 4 (1962), 441–461.

[16] Swann, W. H., "Direct search methods", in *Numerical Methods for Unconstrained Optimization* (P. E. Gill and W. Murray, eds.), Academic Press, London, 13–28 (1972).

[17] Torczon, V., *Multi-Directional Search: A Direct Search Algorithm for Parallel Machines*, PhD thesis, Department of Mathematical Sciences, Rice University, Houston, Texas, 1989.

[18] Tseng, P., Fortified-descent simplicial search method: A general approach, *SIAM Journal on Optimization*, 10 (1999), 269–288.

[19] Woods, D. J., *An Interactive Approach for Solving Multi-Objective Optimization Problems*, PhD thesis, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 1985.

[20] Wright, M. H., 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, 1996.

Margaret H. Wright
Courant Institute
of Mathematical Sciences
New York, New York 10012
USA
`mhw@cs.nyu.edu`