40 Years of the Nelder-Mead Algorithm

David Byatt, Ian Coope, Chris Price

University of Canterbury
New Zealand

5 November, 2003

Abstract

The Nelder-Mead algorithm for unconstrained optimization has been used extensively to solve parameter estimation (and other) problems for almost 40 years. Despite its age it is still the method of choice for many practitioners in the fields of statistics, engineering and the physical and medical sciences because it is easy to code and very easy to use. It belongs to a class of methods which do not require derivatives and which are often claimed to be robust for problems with discontinuities or where function values are noisy. However, relatively recently (McKinnon, 1998) showed that the method can fail to converge or converge to non-solutions on certain classes of problems. Only very limited convergence results exist for a restricted class of problems in 1 or 2 dimensions (Lagarius et al, 1998). So why is the method still used? How can it be improved? Recent developments, are presented to help answer these questions.

Citation Classic

"The Nelder-Mead Simplex algorithm has enjoyed enduring popularity. Of all the direct search methods, the Nelder-Mead simplex algorithm is the one most found in numerical software packages. The original paper by Nelder and Mead is a Science Citatation Index classic with several thousand references across the scientific literature in journals ranging from Acta Anaestesiologica Scandinavica to Zhurnal Fizicheskio Khimii. In fact, there is an entire book from the chemical engineering community devoted to simplex search for optimization"

Lewis, Torczon, Trosset (2000)

Some Historical References

- Nelder, Mead (1965): Original Paper
- Parkinson, Hutchinson (1971): NAG Library implementation
- McKinnon(1998): Counter Examples
- Lagarius, Reeds, Wright, Wright(1998): MATLAB implementation (Fminsearch)
- Rykov(1983), Kelley(2000): Modifications (convergence not guaranteed)
- Tseng(2001), BCP(2002): Convergent Variants

How does Nelder-Mead work?

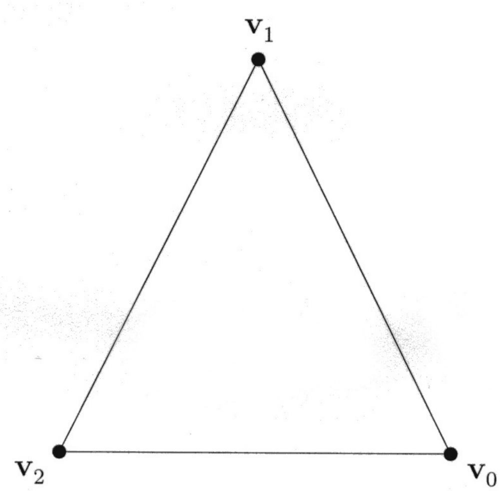
A simplex based direct search method

Direct search: Methods that use comparisons of the values of the objective function and do not require the use of any derivatives are called *direct search methods*.

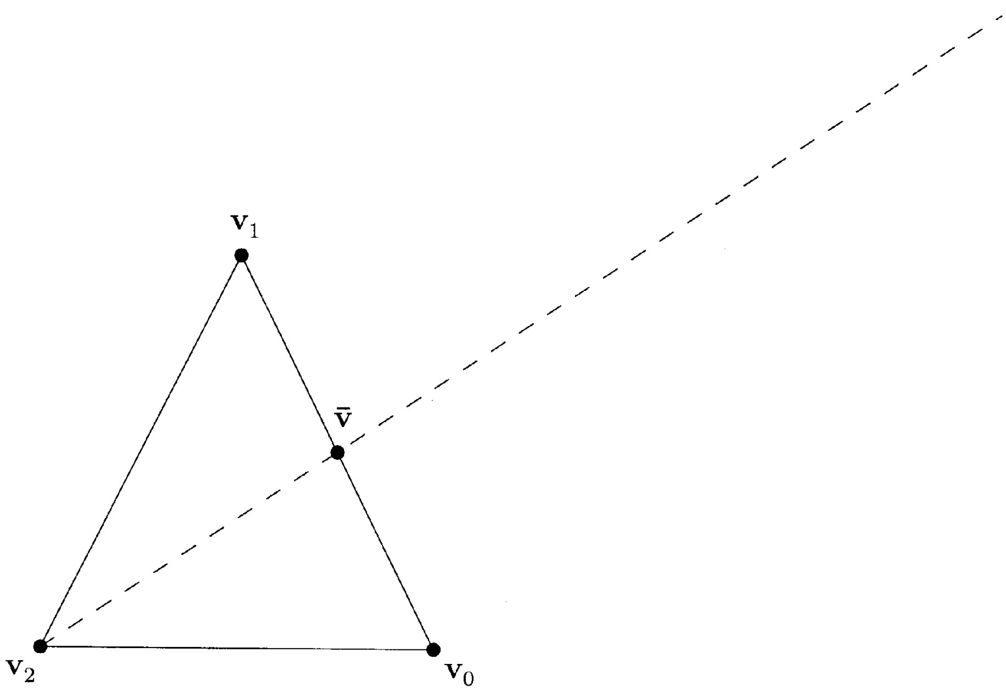
Simplex: A simplex in $\mathbf{R}^n$ is a set of $n + 1$ points that do not lie in a hyperplane.

Four basic operations:

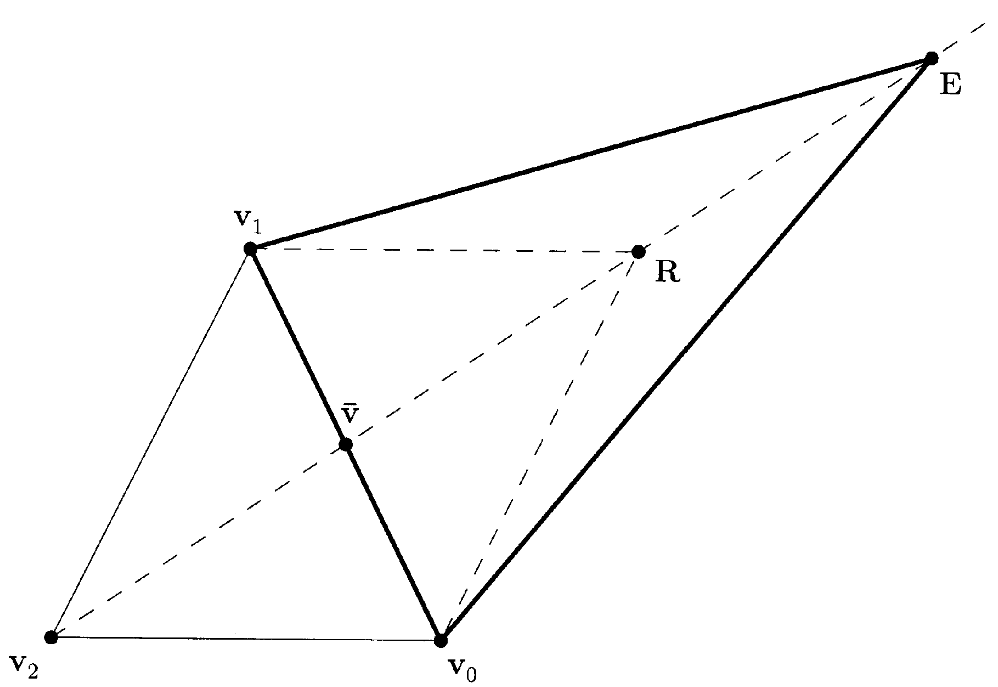
- Reflect
- Expand
- Contract
- Shrink



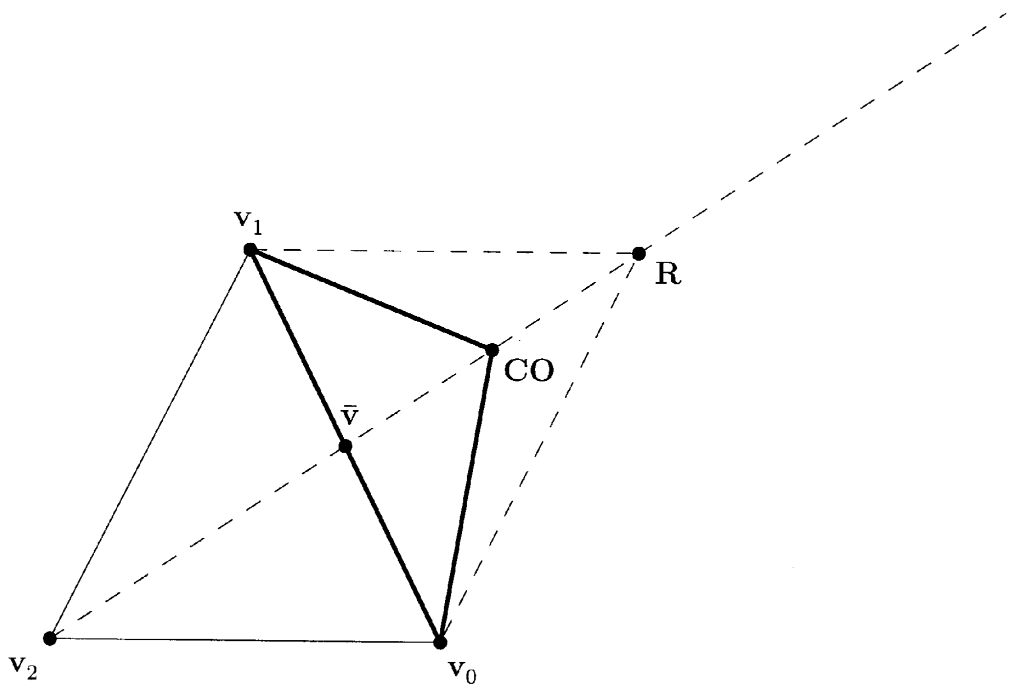
A simplex (2-d)



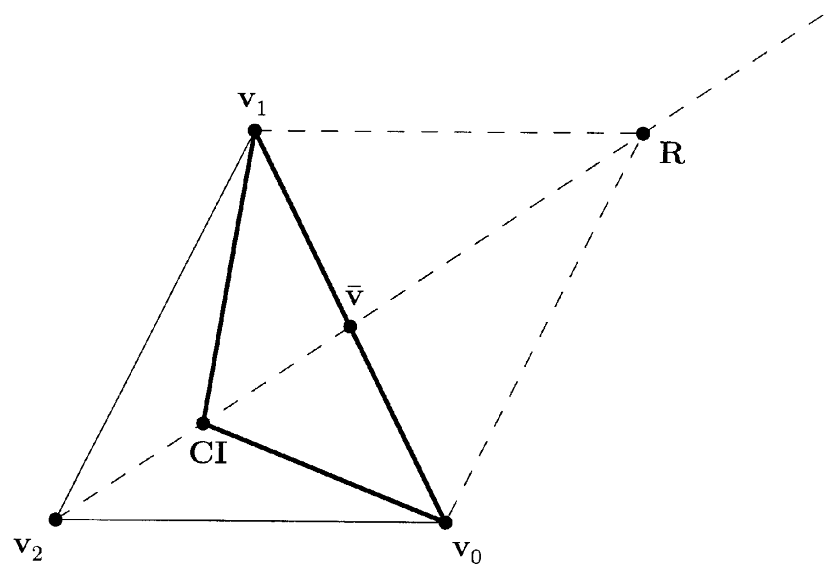
The search direction for the Nelder-Mead algorithm



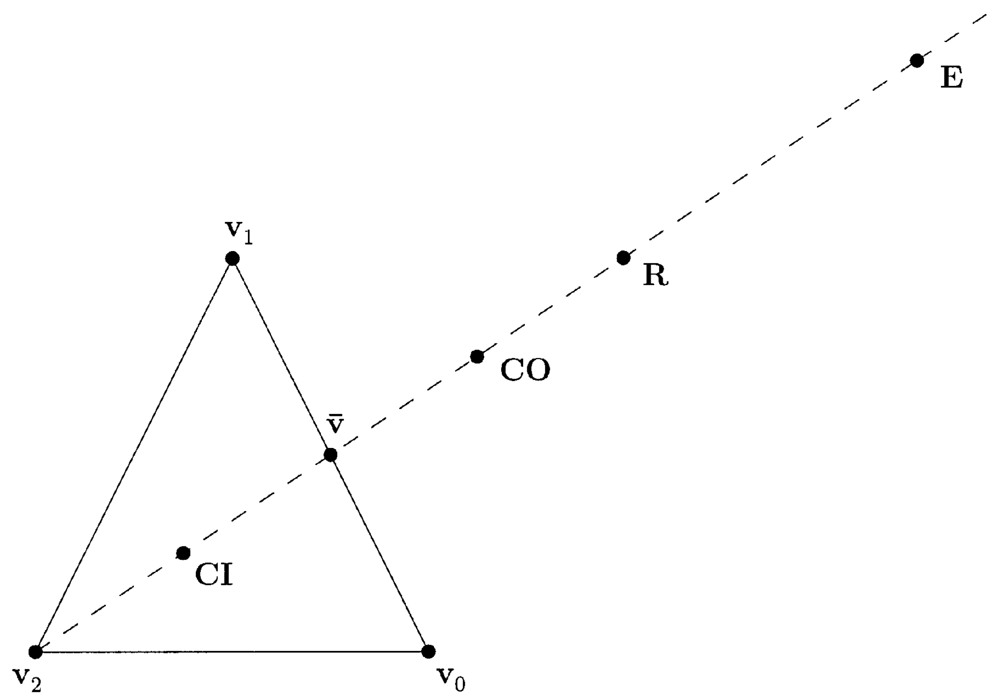
Expand step



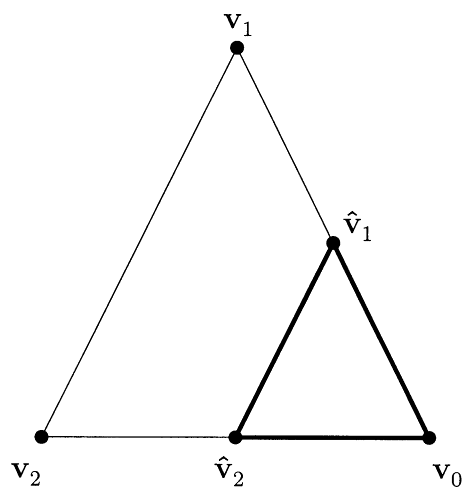
Contract outside step



Contract inside step



All of the trial points for the Nelder-Mead algorithm



Shrink step

What can go wrong?

- Collapse of the simplex

    - McKinnon's functions (2-d)

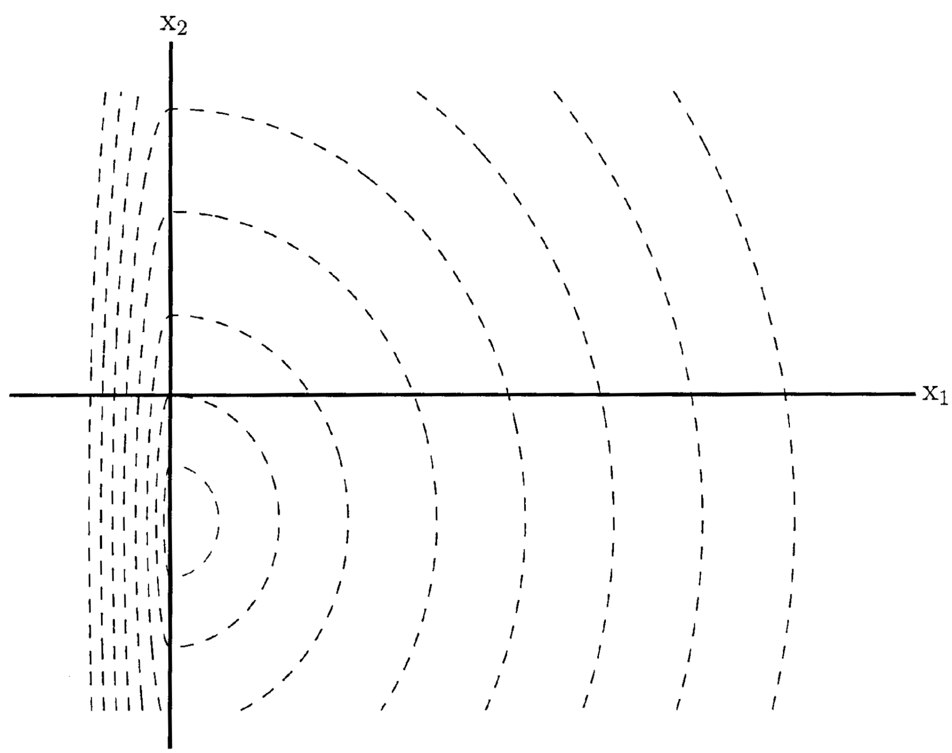
$$f(x_1, x_2) = \begin{cases} \theta\phi|x_1|^{\tau} + x_2 + x_2^2 & x \le 0 \\ \theta x_1^{\tau} + x_2 + x_2^2 & x \ge 0 \end{cases}$$

e.g. $\theta = 6, \quad \phi = 60, \quad \tau = 2$

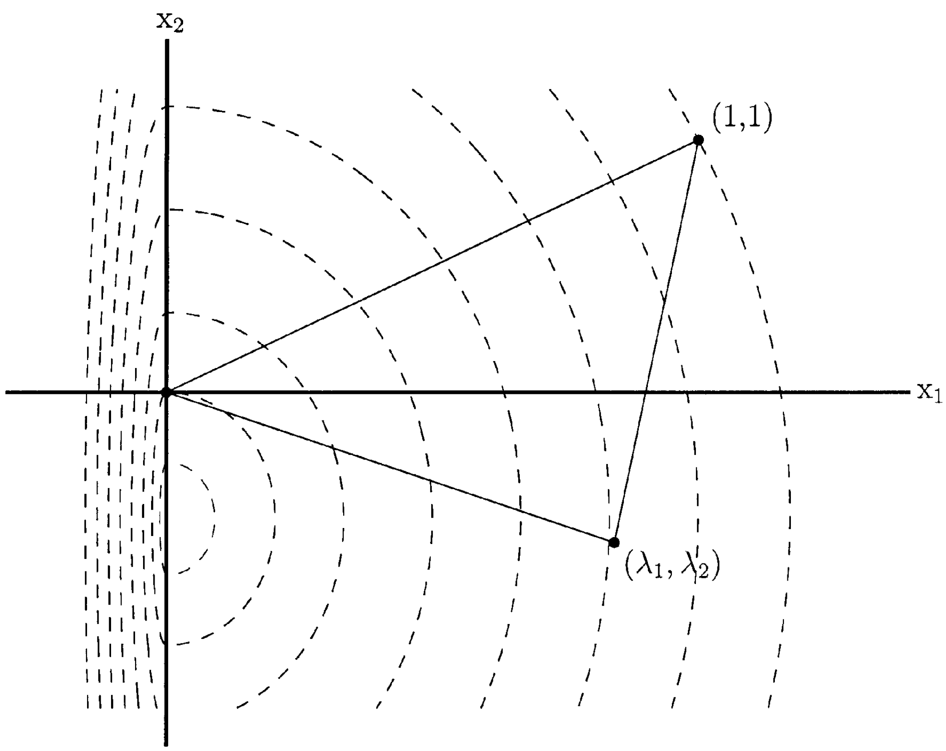
- Failure to make sufficient descent

    - Standard quadratics

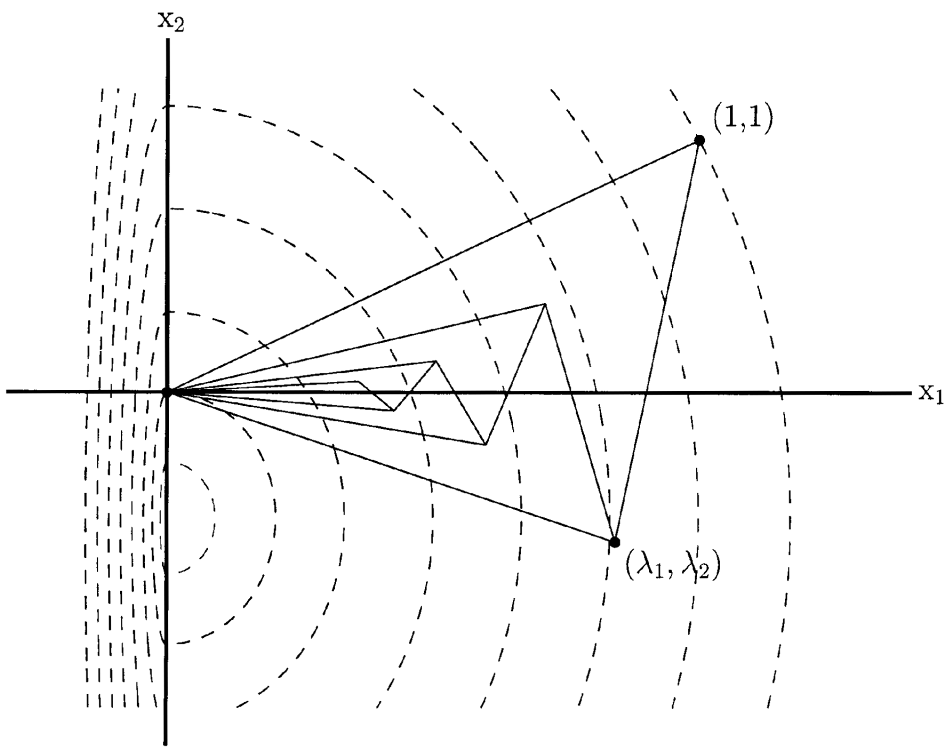
- No convergence proof



Contours for M^cKinnon's function

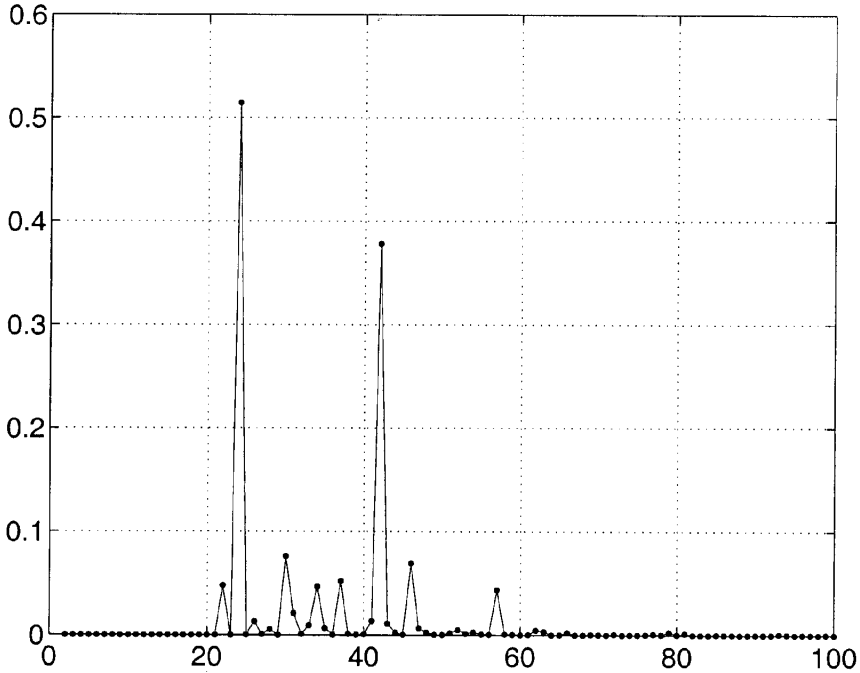


Contours for McKinnon's function and the initial simplex



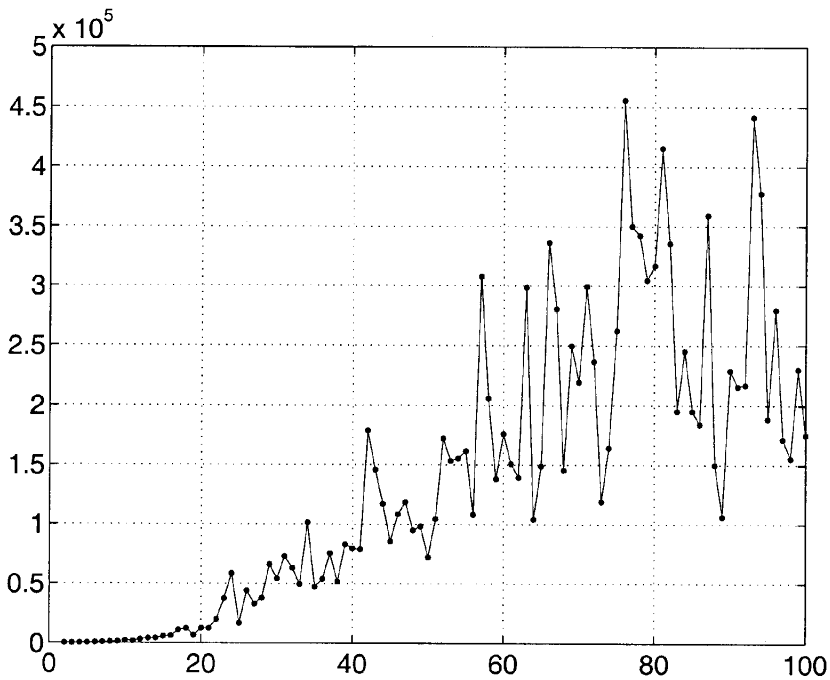
Contours for $M^cKinnon$'s function showing the collapse of successive simplices

The minima found by FMINSEARCH for the standard quadratic for dimensions from two through to 100



Dimension

The number of function evaluations required by FMINSEARCH before terminating



Dimension

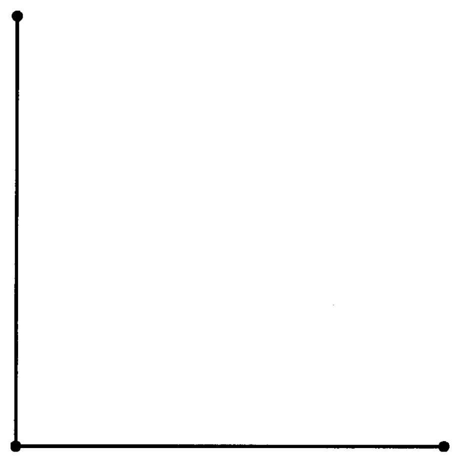
Positive bases

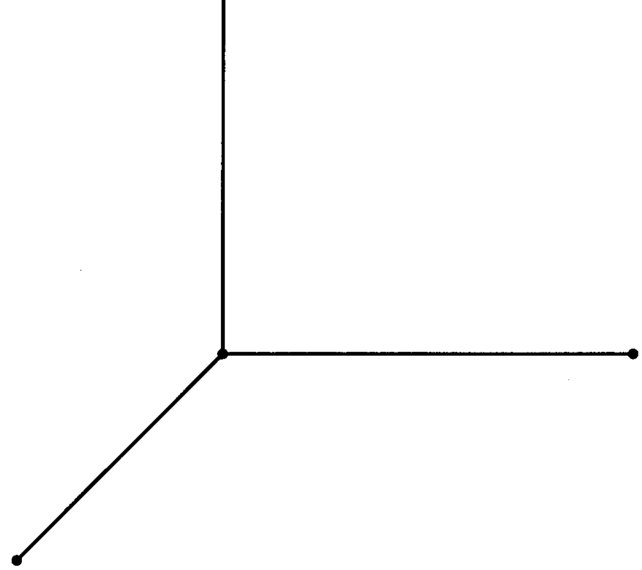
Positive basis: A positive basis P_+ is a set of vectors such that the following conditions hold:

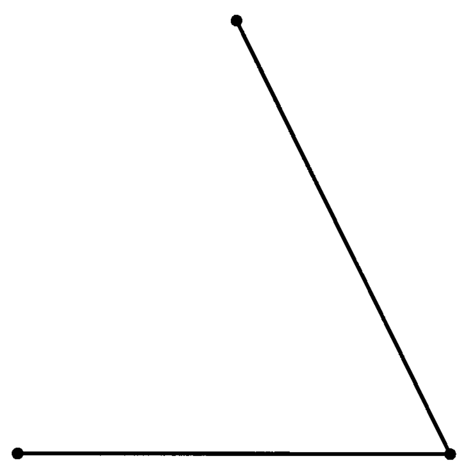
- Every vector in $\mathbf{R}^n$ can be written as a non-negative combination of the vectors in the positive basis.

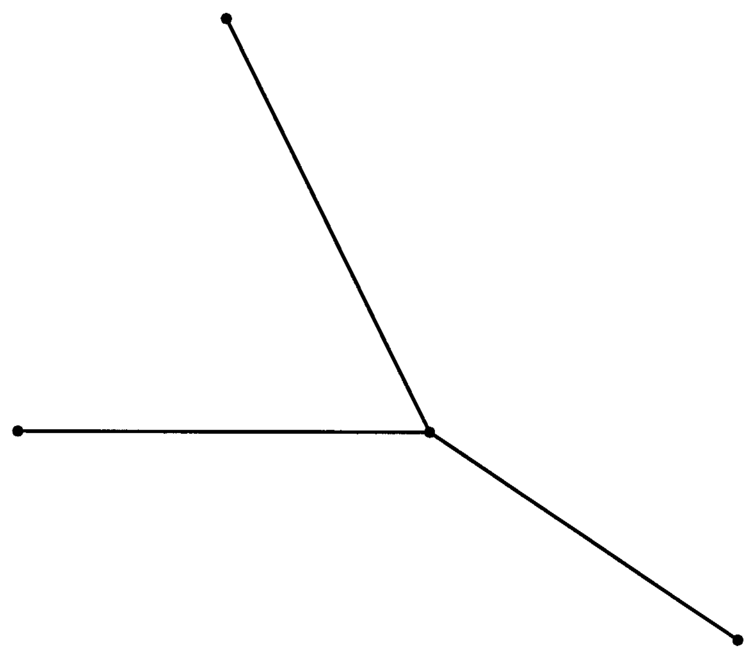
- No elements of P_+ is expressible as a non-negative combination of the remaining elements of P_+

$$P_+ = \left\{ \mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n, \ -\frac{1}{n}\sum_{i=1}^{n} \mathbf{p}_i \right\}$$ where $\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n$ form a basis for $\mathbf{R}^n$





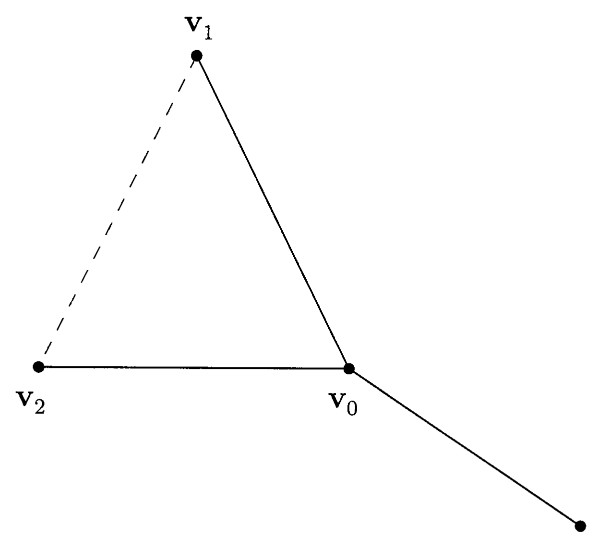




Frames

Frame: A frame F in $\mathbf{R}^n$ is a set of n + 2 points specified by a central frame point $\mathbf{c}$, a positive basis P_+ and a frame size h, such that:

$$F(\mathbf{c}, P_+, h) = \{\mathbf{c}\} \cup \{\mathbf{c} + h\mathbf{p}_i : \mathbf{p}_i \in P_+\}$$



Creating the frame about the best vertex of the current simplex.

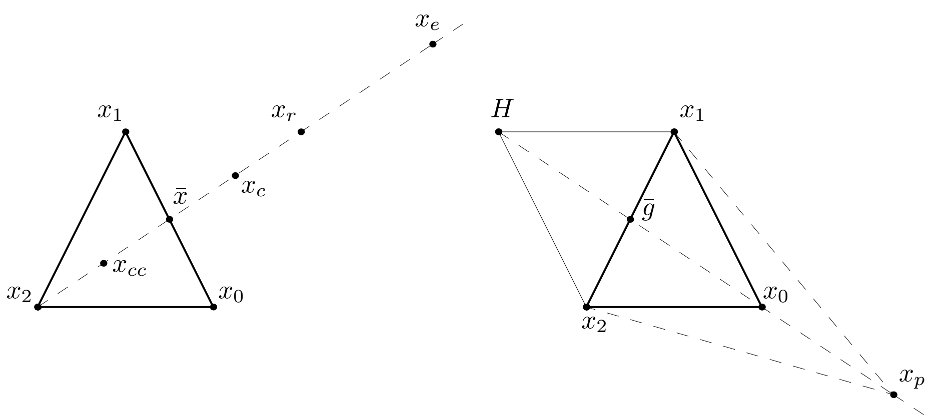


Figure 1: Trial points for the Nelder-Mead algorithm (left image) and the pseudo expand point from the ghost simplex for the modified Nelder-Mead algorithm (right image).

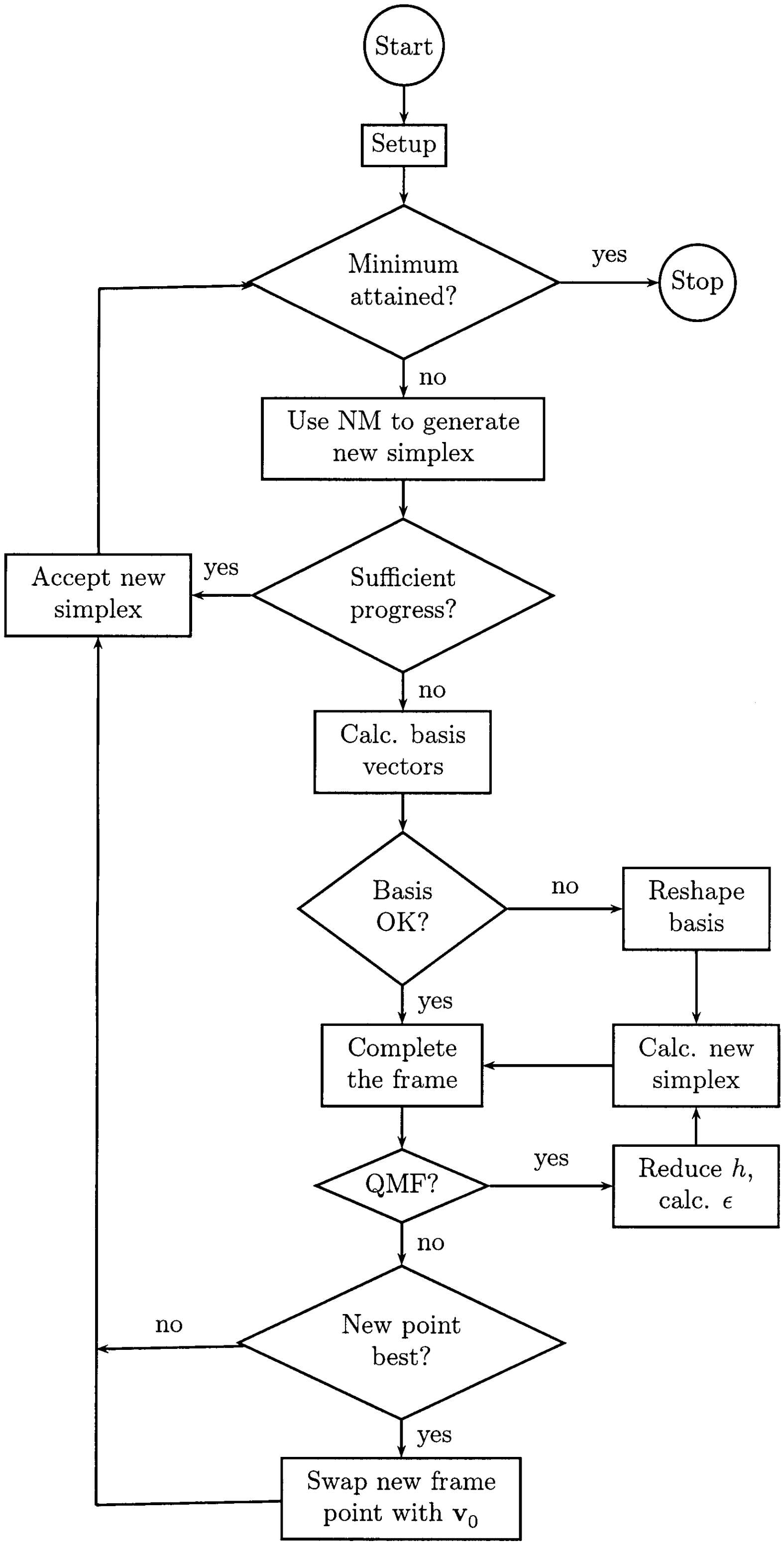
Sufficient descent

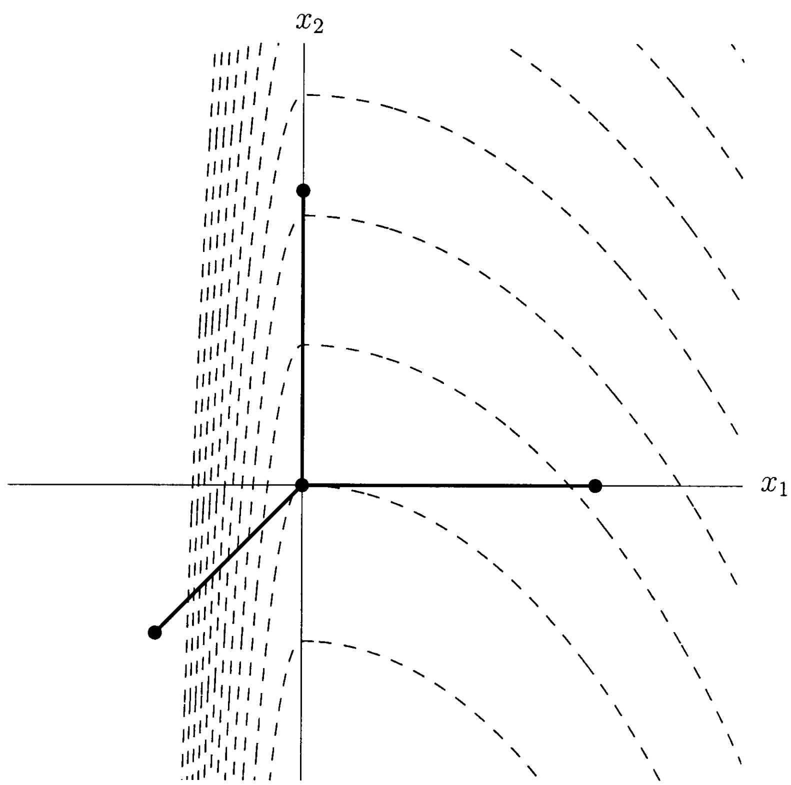
Sufficient descent: The algorithm is said to be making sufficient descent if the function value at the worst vertex of the simplex is reduced by a sufficient amount at each iteration. This can be determined by the use of a sufficient descent parameter ε, calculated using the equation:

$$\varepsilon = Nh^{\nu}$$

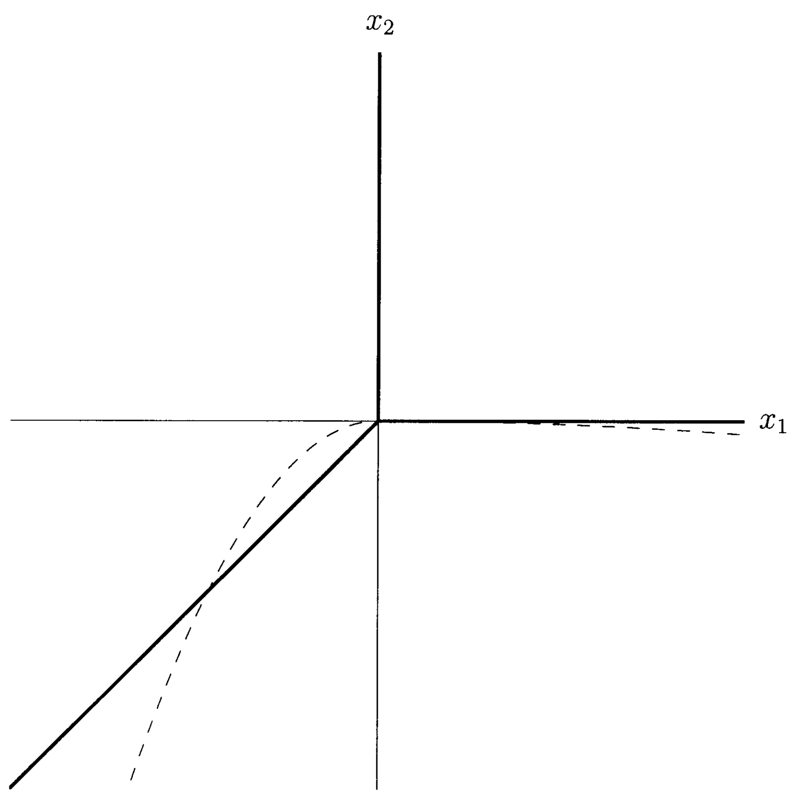
where N is a positive constant and $\nu > 1$ so that as $h \rightarrow 0$, so does ε

Schematic diagram for the variant





An orthogonal frame about the origin and M^cKinnon's function.

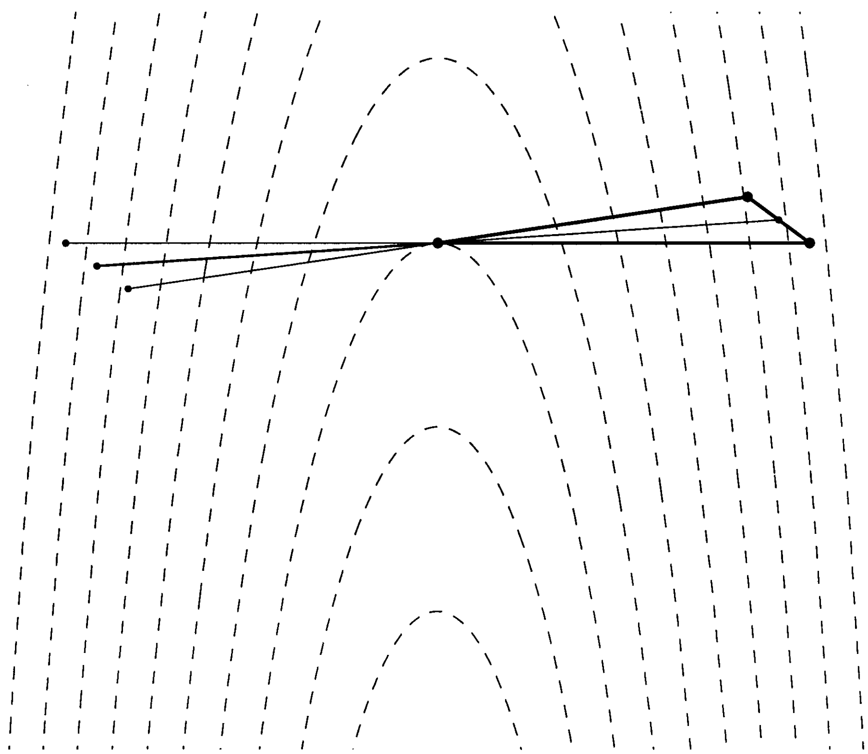


A close-up of the frame about the origin and M^{c}Kinnon's function.

Improvements

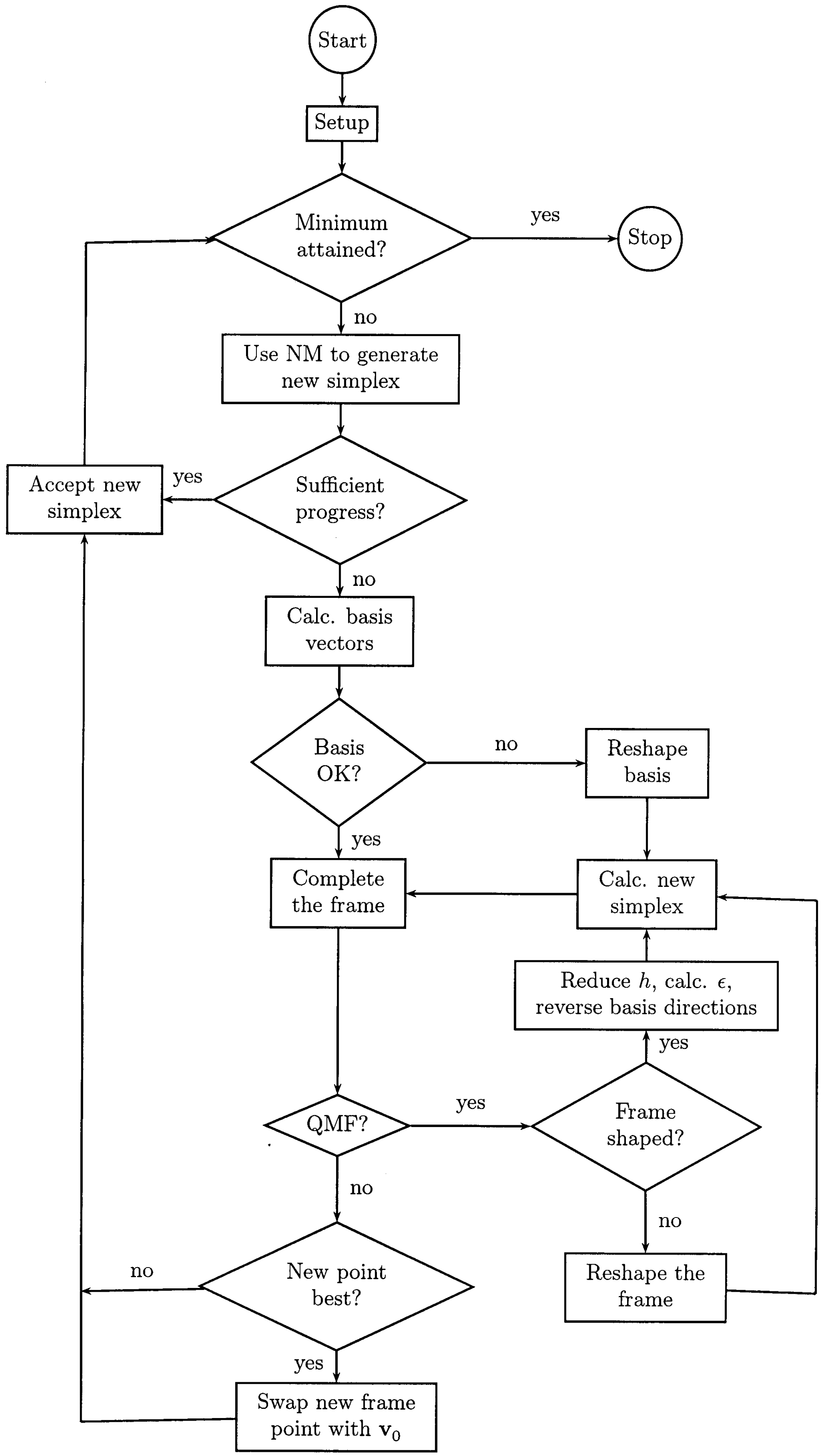
- Alternate the direction of the positive basis vectors every time a frame is reduced in size.

- If completing a frame, make it a good one.

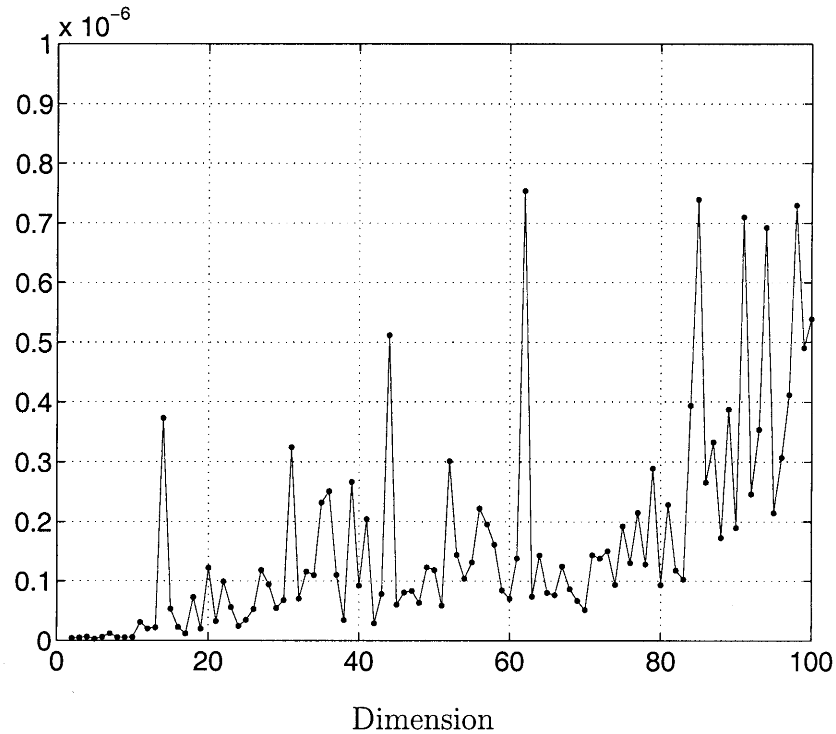


If a simplex is near collapse then even alternating the direction of the positive basis vectors may not produce a descent direction without a big reduction in the frame size.

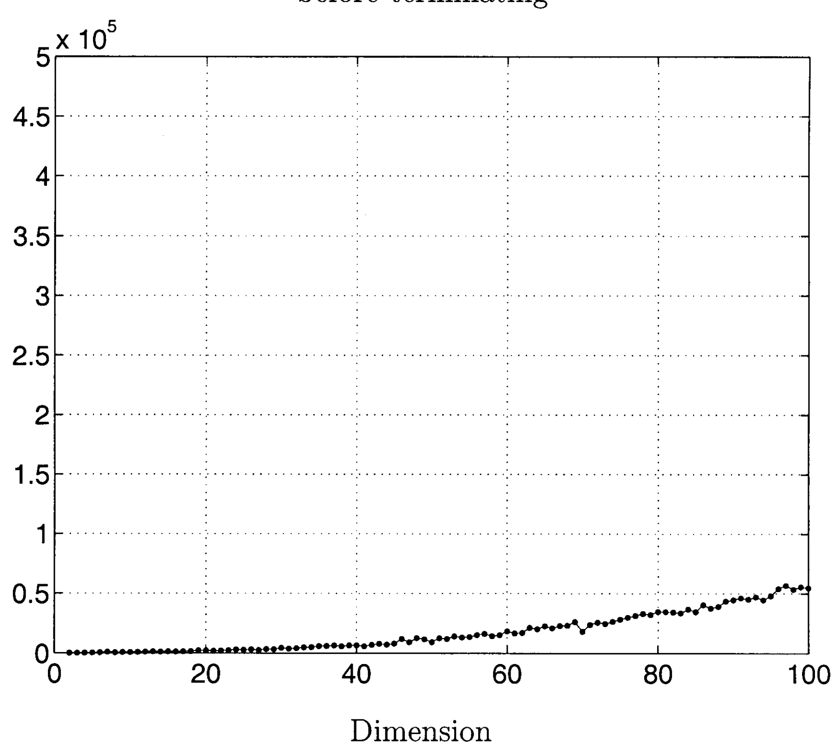
The improved variant



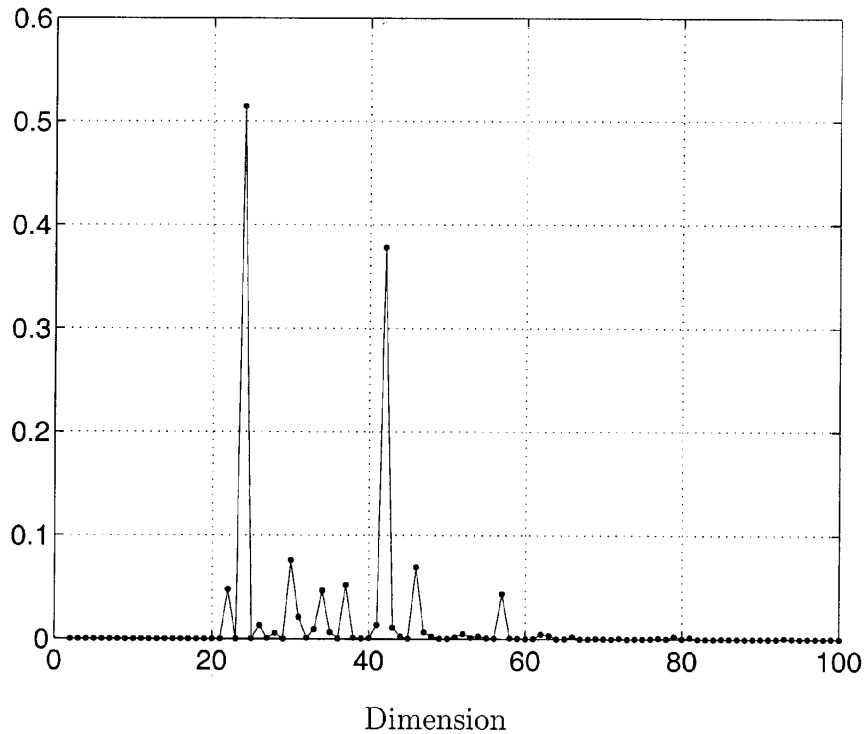
The minima found by the variant for the standard quadratic for dimensions from two through to 100



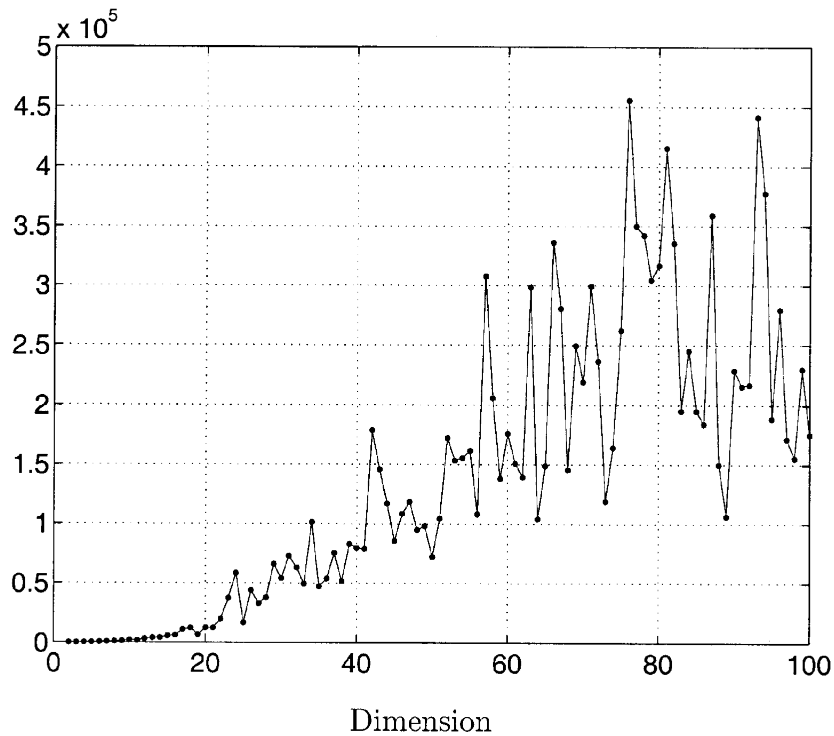
The number of function evaluations required by the variant before terminating



The minima found by FMINSEARCH for the standard quadratic for dimensions from two through to 100



The number of function evaluations required by FMINSEARCH before terminating



	FMINSEARCH		The variant	
Function	*FE*	*Minimum*	*FE*	*Minimum*
Rosenbrock 2-d	219	$1.09909e-18$	285	$1.39058e-17$
Freudenstein and Roth 2-d	172	$4.89843e+01$	217	$4.89843e+01$
Powell badly scaled 2-d	754	$1.11069e-25$	969	$4.23980e-25$
Brown badly scaled 2-d	335	$7.03868e-18$	498	$7.99797e-17$
Beale 2-d	162	$6.11428e-18$	191	$2.07825e-18$
Jennrich and Sampson 2-d	133	$1.24362e+02$	157	$1.24362e+02$
M^{c}Kinnon 2-d	290	$-2.50000e-01$	426	$-2.50000e-01$
Helical valley 3-d	428	$4.78479e-17$	342	$9.83210e-16$
Bard 3-d	*100004	$1.74287e+01$	1134	$1.74287e+01$
Gaussian 3-d	216	$1.12793e-08$	194	$1.12793e-08$
Meyer 3-d	*100004	$8.79459e+01$	2801	$8.79459e+01$
Gulf research 3-d	687	$1.13899e-22$	529	$5.44511e-19$
Box 3-d	701	$3.05741e-22$	478	$8.70459e-21$
Powell singular 4-d	956	$3.56353e-28$	1045	$6.73509e-26$
Wood 4-d	572	$1.56392e-17$	656	$2.57400e-16$
Kowalik and Osbourne 4-d	398	$3.07506e-04$	653	$3.07506e-04$
Brown and Dennis 4-d	*100001	$8.58222e+04$	603	$8.58222e+04$
Quadratic 4-d	326	$4.52859e-17$	440	$2.15350e-17$
Penalty (1) 4-d	1371	$2.24998e-05$	1848	$2.24998e-05$
Penalty (2) 4-d	3730	$9.37629e-06$	4689	$9.37629e-06$
Osbourne (1) 5-d	1098	$5.46489e-05$	1488	$5.46489e-05$
Brown almost linear 5-d	782	$1.45905e-18$	648	$1.08728e-18$
Biggs EXP6 6-d	1130	$5.65565e-03$	4390	$1.16131e-20$
Extended Rosenbrock 6-d	7015	$2.79071e-17$	3110	$1.35844e-14$
Brown almost-linear 7-d	1819	$9.72059e-18$	1539	$1.51163e-17$
Quadratic 8-d	1519	$2.93256e-16$	1002	$8.07477e-17$
Extended Rosenbrock 8-d	5958	$^{\dagger}6.66424e-01$	5314	$3.27909e-17$
Variably dimensional 8-d	3780	$2.08479e-16$	2563	$1.24784e-15$
Extended Powell 8-d	2513	$^{\dagger}5.13165e-07$	7200	$6.43822e-24$
Watson 9-d	3229	$^{\dagger}3.98475e-03$	5256	$1.39976e-06$
Extended Rosenbrock 10-d	6684	$^{\dagger}9.72338e+00$	7629	$2.22125e-16$
Penalty (1) 10-d	5479	$^{\dagger}7.56754e-05$	9200	$7.08765e-05$
Penalty (2) 10-d	6783	$^{\dagger}2.97789e-04$	32768	$2.93661e-04$
Trigonometric 10-d	3105	$2.79506e-05$	2466	$2.79506e-05$
Osbourne (2) 11-d	4926	$4.01377e-02$	6416	$4.01377e-02$
Extended Powell 12-d	6607	$^{\dagger}5.52519e-06$	20076	$1.11105e-20$
Quadratic 16-d	8543	$7.70363e-16$	2352	$1.41547e-16$
Quadratic 24-d	*100000	$^{\dagger}5.04216e-01$	4766	$1.21730e-15$

Comparison of the performance of FMINSEARCH and the variant.

Summary

- Have developed a provably convergent variant of the Nelder-Mead algorithm

- Appears to work well in practice

- Maintains the nice features of the Nelder-Mead algorithm and avoids some of its problems:
    - M^{c}Kinnon's functions
    - Standard quadratics