Nelder-Mead Toolbox Manual – Simplex Theory –

Version 0.2
September 2009

Michaël BAUDIN

Contents

Chapter 1

Simplex theory

In this section, we present the various definitions connected to simplex algorithms. We introduce several methods to measure the size of a simplex, including the oriented length. We present several methods to compute an initial simplex, that is, the regular simplex used by Spendley et al., the axis-by-axis simplex, Pfeffer's simplex and the randomized bounds simplex.

1.1 The simplex

A *simplex* S in $\mathbb{R}^n$ is the convex hull of $n+1$ points $S = \{\mathbf{x}_i\}_{i=1,n+1}$, where $\mathbf{x}_i \in \mathbb{R}^n$ for $i = 1, n+1$.

Box extended the Nelder-Mead algorithm to handle bound and non linear constraints [1]. To be able to manage difficult cases, he uses a *complex* made of $k \geq n+1$ vertices. In this section, we will state clearly when the definition and results can be applied to a complex. Indeed, some definitions such as the simplex gradient cannot be extended to a *complex* and are only applicable to a *simplex*.

The point $\mathbf{x}_i \in \mathbb{R}^n$ is the i-th vertex of S, where $x_i^j \in \mathbb{R}$ is the j-th coordinate of the i-th vertex of the simplex S. Given a function $f(\mathbf{x}) \in \mathbb{R}$, each vertex is associated with a function value $f_i = f(\mathbf{x}_i)$ for $i = 1, n+1$. In simplex algorithms, the vertex are sorted by increasing function values

$$f_1 \leq f_2 \leq \ldots \leq f_n \leq f_{n+1}. \tag{1.1}$$

The sorting order is not precisely defined neither in Spendley's et al paper [9] nor in Nelder and Mead's [6]. In [4], the sorting rules are defined precisely to be able to state a theoretical convergence result. In practical implementations, though, the ordering rules have no measurable influence.

Let V denote the $n \times n$ matrix of simplex directions

$$V(S) = (\mathbf{x}_2 - \mathbf{x}_1, \mathbf{x}_3 - \mathbf{x}_1, \ldots, \mathbf{x}_{n+1} - \mathbf{x}_1) = (\mathbf{v}_1, \ldots, \mathbf{v}_n). \tag{1.2}$$

We say that the simplex S is nonsingular if the matrix of simplex directions $V(S)$ is nonsingular.

1.2 The size of the simplex

Several methods are available to compute the size of a simplex.

The simplex diameter $diam(S)$ is defined by

$$diam(S) = \max_{i,j=1,n+1} \|\mathbf{x}_i - \mathbf{x}_j\|_2, \tag{1.3}$$

where $\|.\|_2$ is the euclidian norm defined by

$$\|\mathbf{x}\|_2 = \sum_{i=1,n} (x^j)^2. \tag{1.4}$$

In practical implementations, computing the diameter requires two nested loops over the vertices of the simplex, i.e. requires $(n+1)^2$ operations. This is why authors generally prefer to use lengths which are less expensive to compute.

The two oriented lengths $\sigma_-(S)$ and $\sigma_+(S)$ are using the first vertex as the reference point and are defined by

$$\sigma_+(S) = \max_{i=2,n+1} \|\mathbf{x}_i - \mathbf{x}_1\|_2 \qquad \text{and} \qquad \sigma_-(S) = \min_{i=2,n+1} \|\mathbf{x}_i - \mathbf{x}_1\|_2 \tag{1.5}$$

The following inequalities are satisfied between the diameter and the maximum oriented length

$$\sigma_+(S) \leq diam(S) \leq 2\sigma_+(S) \tag{1.6}$$

In Nash's book [5], the size of the simplex $s_N(S)$ is measured based on the 1-norm and is defined by

$$s_N(S) = \sum_{i=2,n+1} \|\mathbf{x}_i - \mathbf{x}_1\|_1 \tag{1.7}$$

where the 1-norm is defined by

$$\|\mathbf{x}\|_1 = \sum_{j=1,n} |x^j| \tag{1.8}$$

where $x_i^j \in \mathbb{R}$ is the j-th coordinate of the i-th vertex of the simplex S.

1.3 The initial simplex

While most of the theory can be developed without being very specific about the initial simplex, it plays a very important role in practice. All approaches are based on the initial guess $\overline{\mathbf{x}}_0 \in \mathbb{R}^n$ and create a geometric shape based on this point. (We denote the initial guess by $\overline{\mathbf{x}}_0$ instead of the usual $\mathbf{x}_0$ in order to distinguish the initial guess from the vertices $\{\mathbf{x}_i\}_{i=1,n+1}$.)

In this section, we present the various approach to design the initial simplex. In the first part, we emphasize the importance of the initial simplex in optimization algorithms. Then we present the regular simplex by Spendley et al., the axis-by-axis simplex, the randomized bounds approach by Box and Pfeffer's simplex.

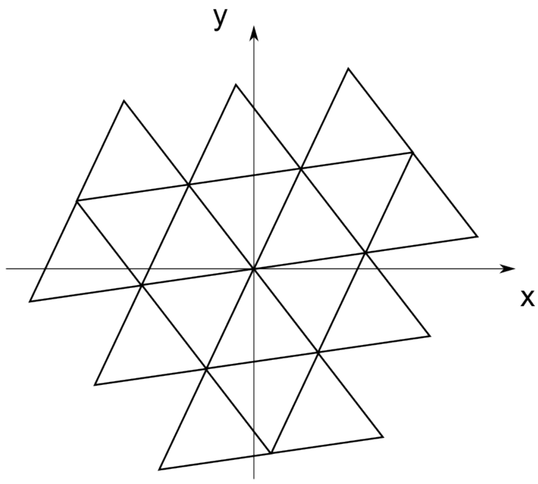

Fig. 1.1 : Typical pattern with fixed-shape Spendley's et al algorithm

1.3.1 Importance of the initial simplex

The initial simplex is particularily important in the case of Spendley's et al method, where the shape of the simplex is fixed during the iterations. Therefore, the algorithm can only go through points which are on the pattern defined by the initial simplex. The pattern presented in figure 1.1 is typical a fixed-shape simplex algorithm (see [10], chapter 3, for other patterns of a direct search method). If, by chance, the pattern is so that the optimum is close to one point defined by the pattern, the number of iteration may be small. On the contrary, the number of iterations may be large if the pattern does not come close to the optimum.

The variable-shape simplex algorithm designed by Nelder and Mead is also very sensitive to the initial simplex. One of the problems is that the initial simplex should be consistently scaled with respect to the unknown x. In "An investigation into the efficiency of variants on the simplex method" [7], Parkinson and Hutchinson explored several improvements of Nelder and Mead's algorithm. First, they investigate the sensitivity of the algorithm to the initial simplex. Two parameters were investigated, that is, the initial length and the orientation of the simplex. The conclusion of their study with respect to the initial simplex is the following. "The orientation of the initial simplex has a significant effect on efficiency, but the relationship can be too sensitive for an automatic predictor to provide sufficient accuracy at this time."

Since no initial simplex clearly improves on the others, in practice, it may be convenient to try different approaches.

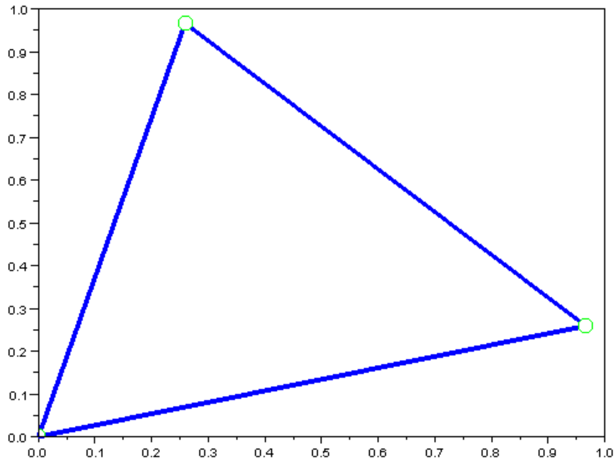

Fig. 1.2 : Regular simplex in 2 dimensions

1.3.2 Spendley's et al regular simplex

In their paper [9], Spendley et al. use a regular simplex with given size $\ell > 0$. We define the parameters $p, q > 0$ as

$$p = \frac{1}{n\sqrt{2}}\left(n - 1 + \sqrt{n+1}\right), \qquad (1.9)$$

$$q = \frac{1}{n\sqrt{2}}\left(\sqrt{n+1} - 1\right). \qquad (1.10)$$

We can now define the vertices of the simplex $S = \{\mathbf{x}_i\}_{i=1,n+1}$. The first vertex of the simplex is the initial guess

$$\mathbf{x}_1 = \overline{\mathbf{x}}_0. \qquad (1.11)$$

The other vertices are defined by $\mathbf{x}_i = (x_i^1, \ldots x_i^n) \in \mathbb{R}^n$ where the coordinates x_i^j are

$$x_i^j = \begin{cases} \overline{x}_0^j + \ell p, & \text{if } j = i - 1, \\ \overline{x}_0^j + \ell q, & \text{if } j \neq i - 1, \end{cases} \qquad (1.12)$$

for vertices $i = 2, n+1$ and components $j = 1, n$, where $\ell \in \mathbb{R}$ is the length of the simplex and satisfies $\ell > 0$. Notice that this length is the same for all the edges which keeps the simplex regular.

The regular simplex is presented in figure 1.2.

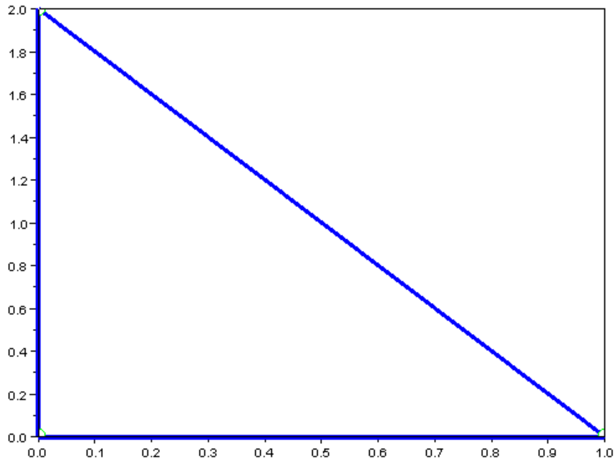

Fig. 1.3 : Axis-based simplex in 2 dimensions – Notice that the length along the x axis is 1 while the length along the y axis is 2.

1.3.3 Axis-by-axis simplex

A very efficient and simple approach leads to an axis-by-axis simplex. This simplex depends on a vector of positive lengths $\mathbf{l} \in \mathbb{R}^n$. The first vertex of the simplex is the initial guess

$$\mathbf{x}_1 \quad = \quad \overline{\mathbf{x}}_0. \tag{1.13}$$

The other vertices are defined by

$$x_i^j \quad = \quad \begin{cases} \overline{x}_0^j + \ell_j, \text{ if } j = i-1, \\ \overline{x}_0^j, \text{ if } j \neq i-1, \end{cases} \tag{1.14}$$

for vertices $i = 2, n+1$ and components $j = 1, n$.

This type of simplex is presented in figure 1.3, where $\ell_1 = 1$ and $\ell_2 = 2$. The axis-by-axis simplex is used in the Nelder-Mead algorithm provided in Numerical Recipes in C [8]. As stated in [8], the length vector $\mathbf{l}$ can be used as a guess for the characteristic length scale of the problem.

1.3.4 Randomized bounds

Assume that the variable $\mathbf{x} \in \mathbb{R}^n$ is bounded so that

$$m^j \leq x^j \leq M^j, \tag{1.15}$$

for $j = 1, n$, where $m_j, M_j \in \mathbb{R}$ are minimum and maximum bounds and $m_j \leq M_j$. A method suggested by Box in [1] is based on the use of pseudo-random numbers. Let $\{\theta_i^j\}_{i=1,n+1,j=1,n} \in [0, 1]$

be a sequence of random numbers uniform in the interval $[0,1]$. The first vertex of the simplex is the initial guess

$$\mathbf{x}_1 \quad = \quad \overline{\mathbf{x}}_0. \tag{1.16}$$

The other vertices are defined by

$$x_i^j \quad = \quad m^j + \theta_i^j (M^j - m^j), \tag{1.17}$$

for vertices $i = 2, n+1$ and components $j = 1, n$.

1.3.5 Pfeffer's method

This initial simplex is used in the function *fminsearch* and presented in [2]. According to [2], this simplex is due to L. Pfeffer at Stanford. The goal of this method is to scale the initial simplex with respect to the characteristic lengths of the problem. This allows, for example, to manage cases where $x_1 \approx 1$ and $x_2 \approx 10^5$. As we are going to see, the scaling is defined with respect to the initial guess $\overline{\mathbf{x}}_0$, with an axis-by-axis method.

The method proceeds by defining $\delta_u, \delta_z > 0$, where δ_u is used for usual components of $\overline{\mathbf{x}}_0$ and δ_z is used for the case where one component of $\overline{\mathbf{x}}_0$ is zero. The default values for δ_u and δ_z are

$$\delta_u = 0.05 \text{ and } \delta_z = 0.0075. \tag{1.18}$$

The first vertex of the simplex is the initial guess

$$\mathbf{x}_1 \quad = \quad \overline{\mathbf{x}}_0. \tag{1.19}$$

The other vertices are defined by

$$x_i^j \quad = \quad \begin{cases} \overline{x}_0^j + \delta_u \overline{x}_0^j, \text{ if } j = i-1 \text{ and } \overline{x}_0^{j-1} \neq 0, \\ \delta_z, \text{ if } j = i-1 \text{ and } \overline{x}_0^{j-1} = 0, \\ \overline{x}_0^j, \text{ if } j \neq i-1, \end{cases} \tag{1.20}$$

for vertices $i = 2, n+1$ and components $j = 1, n$.

1.4 References and notes

Some elements of the section 1.2 is taken from Kelley's book [3], "Iterative Methods for Optimization". While this document focus on Nelder-Mead algorithm, Kelley gives a broad view on optimization and present other algorithms for noisy functions, like implicit filtering, multidirectional search and the Hooke-Jeeves algorithm.

Bibliography

[1] M. J. Box. A new method of constrained optimization and a comparison with other methods. *The Computer Journal*, pages 42–52, 1965.

[2] Ellen Fan. Global optimization of lennard-jones atomic clusters. Technical report, McMaster University, February 2002.

[3] C. T. Kelley. *Iterative Methods for Optimization*, volume 19. SIAM Frontiers in Applied Mathematics, 1999. http://www.siam.org/books/textbooks/fr18_book.pdf.

[4] Jeffrey C. Lagarias, James A. Reeds, Margaret H. Wright, and Paul E. Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1):112–147, 1998. http://link.aip.org/link/?SJE/9/112/1.

[5] J. C. Nash. *Compact numerical methods for computers : linear algebra and function minimisation*. Hilger, Bristol, 1979.

[6] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, January 1965.

[7] Parkinson and Hutchinson. An investigation into the efficiency of variants on the simplex method. *F. A. Lootsma, editor, Numerical Methods for Non-linear Optimization*, pages 115–135, 1972.

[8] W. H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C, Second Edition*. 1992.

[9] W. Spendley, G. R. Hext, and F. R. Himsworth. Sequential application of simplex designs in optimisation and evolutionary operation. *Technometrics*, 4(4):441–461, 1962.

[10] Virginia Joanne Torczon. Multi-directional search: A direct search algorithm for parallel machines. Technical report, Rice University, 1989.