

10.4 Downhill Simplex Method in Multidimensions

With this section we begin consideration of multidimensional minimization, that is, finding the minimum of a function of more than one independent variable. This section stands apart from those which follow, however: All of the algorithms after this section will make explicit use of a one-dimensional minimization algorithm as a part of their computational strategy. This section implements an entirely self-contained strategy, in which one-dimensional minimization does not figure.

The *downhill simplex method* is due to Nelder and Mead [1]. The method requires only function evaluations, not derivatives. It is not very efficient in terms of the number of function evaluations that it requires. Powell's method (§10.5) is almost surely faster in all likely applications. However, the downhill simplex method may frequently be the *best* method to use if the figure of merit is “get something working quickly” for a problem whose computational burden is small.

The method has a geometrical naturalness about it which makes it delightful to describe or work through:

A *simplex* is the geometrical figure consisting, in N dimensions, of $N + 1$ points (or vertices) and all their interconnecting line segments, polygonal faces, etc. In two dimensions, a simplex is a triangle. In three dimensions it is a tetrahedron, not necessarily the regular tetrahedron. (The *simplex method* of linear programming, described in §10.8, also makes use of the geometrical concept of a simplex. Otherwise it is completely unrelated to the algorithm that we are describing in this section.) In general we are only interested in simplexes that are nondegenerate, i.e., that enclose a finite inner N -dimensional volume. If any point of a nondegenerate simplex is taken as the origin, then the N other points define vector directions that span the N -dimensional vector space.

In one-dimensional minimization, it was possible to bracket a minimum, so that the success of a subsequent isolation was guaranteed. Alas! There is no analogous procedure in multidimensional space. For multidimensional minimization, the best we can do is give our algorithm a starting guess, that is, an N -vector of independent variables as the first point to try. The algorithm is then supposed to make its own way downhill through the unimaginable complexity of an N -dimensional topography, until it encounters a (local, at least) minimum.

The downhill simplex method must be started not just with a single point, but with $N + 1$ points, defining an initial simplex. If you think of one of these points (it matters not which) as being your initial starting point \mathbf{P}_0 , then you can take the other N points to be

$$\mathbf{P}_i = \mathbf{P}_0 + \lambda \mathbf{e}_i \quad (10.4.1)$$

where the \mathbf{e}_i 's are N unit vectors, and where λ is a constant which is your guess of the problem's characteristic length scale. (Or, you could have different λ_i 's for each vector direction.)

The downhill simplex method now takes a series of steps, most steps just moving the point of the simplex where the function is largest (“highest point”) through the opposite face of the simplex to a lower point. These steps are called reflections,

Sample page from NUMERICAL RECIPES IN FORTRAN 77: THE ART OF SCIENTIFIC COMPUTING (ISBN 0-521-43064-X)
Copyright (C) 1986-1992 by Cambridge University Press. Programs Copyright (C) 1986-1992 by Numerical Recipes Software.
Permission is granted for internet users to make one paper copy for their own personal use. Further reproduction, or any copying of machine-readable files (including this one), to any server computer, is strictly prohibited. To order Numerical Recipes books or CDROMs, visit website <http://www.nr.com> or call 1-800-872-7423 (North America only), or send email to directcustserv@cambridge.org (outside North America).

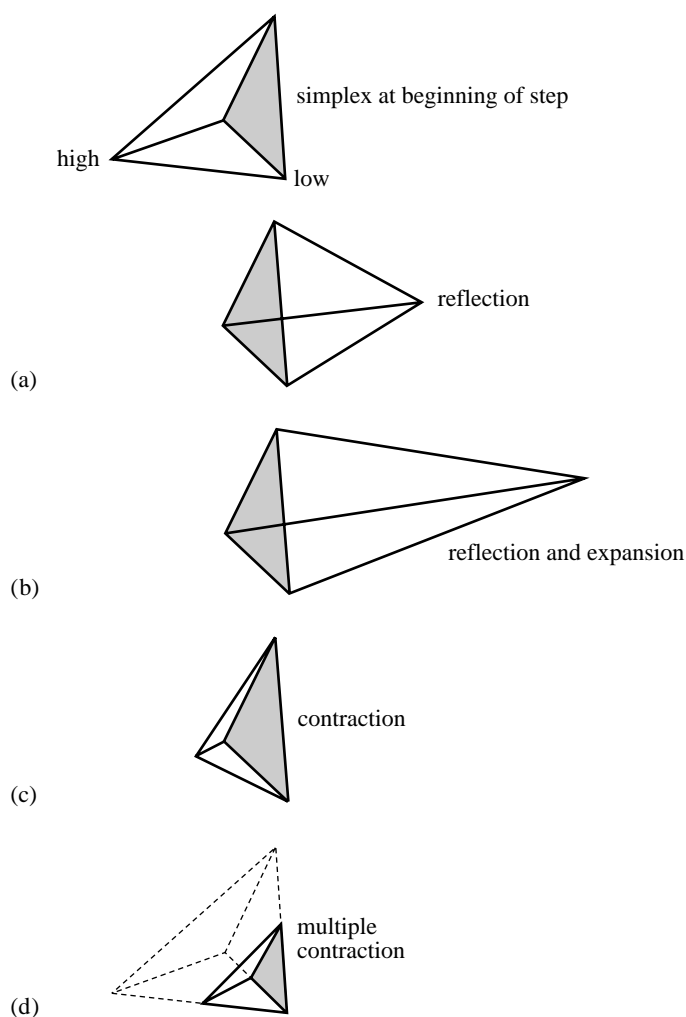


Figure 10.4.1. Possible outcomes for a step in the downhill simplex method. The simplex at the beginning of the step, here a tetrahedron, is shown, top. The simplex at the end of the step can be any one of (a) a reflection away from the high point, (b) a reflection and expansion away from the high point, (c) a contraction along one dimension from the high point, or (d) a contraction along all dimensions towards the low point. An appropriate sequence of such steps will always converge to a minimum of the function.

and they are constructed to conserve the volume of the simplex (hence maintain its nondegeneracy). When it can do so, the method expands the simplex in one or another direction to take larger steps. When it reaches a “valley floor,” the method contracts itself in the transverse direction and tries to ooze down the valley. If there is a situation where the simplex is trying to “pass through the eye of a needle,” it contracts itself in all directions, pulling itself in around its lowest (best) point. The routine name *amoeba* is intended to be descriptive of this kind of behavior; the basic moves are summarized in Figure 10.4.1.

Termination criteria can be delicate in any multidimensional minimization routine. Without bracketing, and with more than one independent variable, we no longer have the option of requiring a certain tolerance for a single independent

variable. We typically can identify one “cycle” or “step” of our multidimensional algorithm. It is then possible to terminate when the vector distance moved in that step is fractionally smaller in magnitude than some tolerance `tol`. Alternatively, we could require that the decrease in the function value in the terminating step be fractionally smaller than some tolerance `ftol`. Note that while `tol` should not usually be smaller than the square root of the machine precision, it is perfectly appropriate to let `ftol` be of order the machine precision (or perhaps slightly larger so as not to be diddled by roundoff).

Note well that either of the above criteria might be fooled by a single anomalous step that, for one reason or another, failed to get anywhere. Therefore, it is frequently a good idea to *restart* a multidimensional minimization routine at a point where it claims to have found a minimum. For this restart, you should reinitialize any ancillary input quantities. In the downhill simplex method, for example, you should reinitialize N of the $N + 1$ vertices of the simplex again by equation (10.4.1), with \mathbf{P}_0 being one of the vertices of the claimed minimum.

Restarts should never be very expensive; your algorithm did, after all, converge to the restart point once, and now you are starting the algorithm already there.

Consider, then, our N -dimensional amoeba:

```

SUBROUTINE amoeba(p,y,mp,np,ndim,ftol,funk,iter)
  INTEGER iter,mp,ndim,np,NMAX,ITMAX
  REAL ftol,p(mp,np),y(mp),funk,TINY
  PARAMETER (NMAX=20,ITMAX=5000,TINY=1.e-10)      Maximum allowed dimensions and func-
  EXTERNAL funk                                     tion evaluations, and a small num-
C  USES amotry,funk                                  ber.
  Multidimensional minimization of the function funk(x) where x(1:ndim) is a vector
  in ndim dimensions, by the downhill simplex method of Nelder and Mead. The matrix
  p(1:ndim+1,1:ndim) is input. Its ndim+1 rows are ndim-dimensional vectors which are
  the vertices of the starting simplex. Also input is the vector y(1:ndim+1), whose com-
  ponents must be pre-initialized to the values of funk evaluated at the ndim+1 vertices (rows)
  of p; and ftol the fractional convergence tolerance to be achieved in the function value
  (n.b.!). On output, p and y will have been reset to ndim+1 new points all within ftol of
  a minimum function value, and iter gives the number of function evaluations taken.
  INTEGER i,ih,i,ilo,inhi,j,m,n
  REAL rtol,sum,swap,ysave,ytry,psum(NMAX),amotry
  iter=0
1  do 12 n=1,ndim                                Enter here when starting or have just overall contracted.
      sum=0.                                       Recompute psum.
      do 11 m=1,ndim+1
          sum=sum+p(m,n)
      enddo 11
      psum(n)=sum
  enddo 12
2  ilo=1                                           Enter here when have just changed a single point.
  if (y(1).gt.y(2)) then                          Determine which point is the highest (worst), next-highest,
      ih=1                                         and lowest (best),
      inhi=2
  else
      ih=2
      inhi=1
  endif
  do 13 i=1,ndim+1                                by looping over the points in the simplex.
      if(y(i).le.y(ilo)) ilo=i
      if(y(i).gt.y(ihi)) then
          inhi=ihi
          ihi=i
      else if(y(i).gt.y(inhi)) then
          if(i.ne.ihi) inhi=i
      endif
  enddo 13

```

Sample page from NUMERICAL RECIPES IN FORTRAN 77: THE ART OF SCIENTIFIC COMPUTING (ISBN 0-521-43064-X)
 Copyright (C) 1986-1992 by Cambridge University Press. Programs Copyright (C) 1986-1992 by Numerical Recipes Software.
 Permission is granted for internet users to make one paper copy for their own personal use. Further reproduction, or any copying of machine-readable files (including this one), to any server computer, is strictly prohibited. To order Numerical Recipes books or CDROMs, visit website <http://www.nr.com> or call 1-800-872-7423 (North America only), or send email to directcustserv@cambridge.org (outside North America).

```

        endif
    enddo 13
    rtol=2.*abs(y(ihi)-y(ilo))/(abs(y(ihi))+abs(y(ilo))+TINY)
    Compute the fractional range from highest to lowest and return if satisfactory.
    if (rtol.lt.ftol) then      If returning, put best point and value in slot 1.
        swap=y(1)
        y(1)=y(ilo)
        y(ilo)=swap
        do 14 n=1,ndim
            swap=p(1,n)
            p(1,n)=p(ilo,n)
            p(ilo,n)=swap
        enddo 14
        return
    endif
    if (iter.ge.ITMAX) pause 'ITMAX exceeded in amoeba'
    iter=iter+2
    Begin a new iteration. First extrapolate by a factor -1 through the face of the simplex across
    from the high point, i.e., reflect the simplex from the high point.
    ytry=amotry(p,y,psum,mp,np,ndim,funk,ihi,-1.0)
    if (ytry.le.y(ilo)) then
        Gives a result better than the best point, so try an additional extrapolation by a factor 2.
        ytry=amotry(p,y,psum,mp,np,ndim,funk,ihi,2.0)
    else if (ytry.ge.y(ihi)) then
        The reflected point is worse than the second-highest, so look for an intermediate lower point,
        i.e., do a one-dimensional contraction.
        ysave=y(ihi)
        ytry=amotry(p,y,psum,mp,np,ndim,funk,ihi,0.5)
        if (ytry.ge.ysave) then      Can't seem to get rid of that high point. Better contract
            do 16 i=1,ndim+1          around the lowest (best) point.
                if(i.ne.ilo)then
                    do 15 j=1,ndim
                        psum(j)=0.5*(p(i,j)+p(ilo,j))
                        p(i,j)=psum(j)
                    enddo 15
                    y(i)=funk(psum)
                endif
            enddo 16
            iter=iter+ndim          Keep track of function evaluations.
            goto 1                  Go back for the test of doneness and the next iteration.
        endif
    else
        iter=iter-1                Correct the evaluation count.
    endif
    goto 2
END

```

```

FUNCTION amotry(p,y,psum,mp,np,ndim,funk,ihi,fac)
INTEGER ihi,mp,ndim,np,NMAX
REAL amotry,fac,p(mp,np),psum(np),y(mp),funk
PARAMETER (NMAX=20)
EXTERNAL funk

```

C *USES funk*

```

    Extrapolates by a factor fac through the face of the simplex across from the high point,
    tries it, and replaces the high point if the new point is better.
    INTEGER j
    REAL fac1,fac2,ytry,ptry(NMAX)
    fac1=(1.-fac)/ndim
    fac2=fac1-fac
    do 11 j=1,ndim
        ptry(j)=psum(j)*fac1-p(ihi,j)*fac2
    enddo 11

```

Sample page from NUMERICAL RECIPES IN FORTRAN 77: THE ART OF SCIENTIFIC COMPUTING (ISBN 0-521-43064-X)
 Copyright (C) 1986-1992 by Cambridge University Press. Programs Copyright (C) 1986-1992 by Numerical Recipes Software.
 Permission is granted for internet users to make one paper copy for their own personal use. Further reproduction, or any copying of machine-readable files (including this one), to any server computer, is strictly prohibited. To order Numerical Recipes books or CDROMs, visit website <http://www.nr.com> or call 1-800-872-7423 (North America only), or send email to directcustserv@cambridge.org (outside North America).

```

ytry=funk(ptry)           Evaluate the function at the trial point.
if (ytry.lt.y(ihi)) then  If it's better than the highest, then replace the highest.
  y(ihi)=ytry
  do 12 j=1,ndim
    psum(j)=psum(j)-p(ihi,j)+ptry(j)
    p(ihi,j)=ptry(j)
  enddo 12
endif
amotry=ytry
return
END

```

CITED REFERENCES AND FURTHER READING:

Nelder, J.A., and Mead, R. 1965, *Computer Journal*, vol. 7, pp. 308–313. [1]
 Yarbro, L.A., and Deming, S.N. 1974, *Analytica Chimica Acta*, vol. 73, pp. 391–398.
 Jacoby, S.L.S., Kowalik, J.S., and Pizzo, J.T. 1972, *Iterative Methods for Nonlinear Optimization Problems* (Englewood Cliffs, NJ: Prentice-Hall).

10.5 Direction Set (Powell's) Methods in Multidimensions

We know (§10.1–§10.3) how to minimize a function of one variable. If we start at a point \mathbf{P} in N -dimensional space, and proceed from there in some vector direction \mathbf{n} , then any function of N variables $f(\mathbf{P})$ can be minimized along the line \mathbf{n} by our one-dimensional methods. One can dream up various multidimensional minimization methods that consist of sequences of such line minimizations. Different methods will differ only by how, at each stage, they choose the next direction \mathbf{n} to try. All such methods presume the existence of a “black-box” sub-algorithm, which we might call `linmin` (given as an explicit routine at the end of this section), whose definition can be taken for now as

`linmin`: Given as input the vectors \mathbf{P} and \mathbf{n} , and the function f , find the scalar λ that minimizes $f(\mathbf{P} + \lambda\mathbf{n})$. Replace \mathbf{P} by $\mathbf{P} + \lambda\mathbf{n}$. Replace \mathbf{n} by $\lambda\mathbf{n}$. Done.

All the minimization methods in this section and in the two sections following fall under this general schema of successive line minimizations. (The algorithm in §10.7 does not need very accurate line minimizations. Accordingly, it has its own approximate line minimization routine, `lnsrch`.) In this section we consider a class of methods whose choice of successive directions does not involve explicit computation of the function's gradient; the next two sections do require such gradient calculations. You will note that we need not specify whether `linmin` uses gradient information or not. That choice is up to you, and its optimization depends on your particular function. You would be crazy, however, to use gradients in `linmin` and *not* use them in the choice of directions, since in this latter role they can drastically reduce the total computational burden.