Multi variables without constraint

The numerical optimization of general nonlinear multivariable objective functions requires efficient and robust techniques. Efficiency is important because these problems require an iterative solution procedure and trial and error becomes impractical for more than three or four variables. Robustness (the ability to achieve a solution) is desirable because a general non-linear function is unpredictable in its behavior, there may be relative maxima or minima, saddle points, region of convexity, concavity and so on.

We discuss the solution of unconstraint optimization Problem

Find $x^* = [x_1^*, x_2^* \dots x_n^*]^T$ that minimizes $f(x_1, x_2, x_3, \dots x_n) = f(x)$

Most effective iterative procedures alternate between two phases in the optimization

At iteration K, where the current x is x^k, they do the following

1. Choose a search direction s^k
2. Minimize along that direction (usually inexactly) to find a new point

$$x^{k+1} = x^k + \alpha^k s^k$$

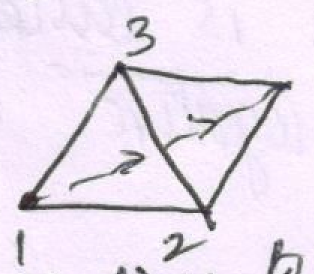
where α^k is a positive scalar called the step size. The step size is determined by an optimization process called a line search a

In addition to 1 and 2, an algorithm must specify

3. the initial starting vector $x^0 = [x_1^0 \; x_2^0 \dots x_n]^T$ and
4. the convergence criteria for termination.

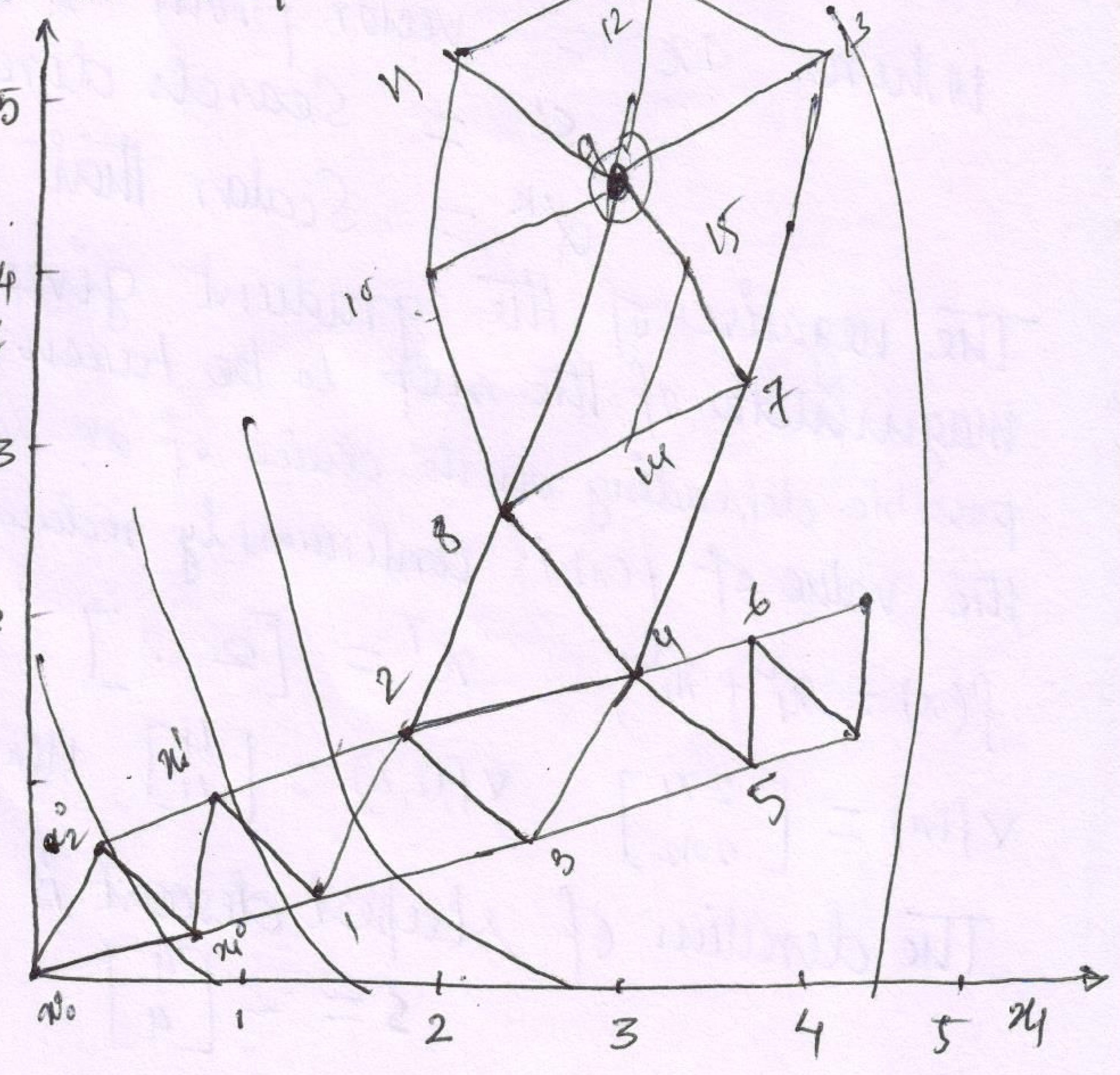
-Simplex search method-

The method of the sequential simplex formulated by Spendly, Hext Himsworth (1962). selects points at the vertices of the simplex at which to evaluate $f(x)$.



In two dimension the figure is an equilateral triangle. In three dimensions this figure becomes a regular tetrahedron. Reflection to a new point in the simplex method. at pt 1 $f(x)$ is greater than f at points 2 or 3. and so on. Each search direction points away from the vertex having the highest value of $f(x)$ to the other vertices in the simplex. thus the direction of search changes, but the step size is fixed for a given size simplex.

Progression to the vicinity of the optimum and oscillation around the optimum using the simplex method of search. The original vertices are x_0^0, x_1^0, x_2^0. The next point (vertex) is x_1^1. succeeding new vertices are numbered starting with 1 and continuing to 13 at which point a cycle starts to repeat. The size of the simplex is reduced, to the triangle determined by points 7, 14, 15 then the procedure is continued.



— Gradient Method —

Methods that use first derivatives

A good search direction should reduce (for minimization) the objective function so that if x^0 is the original point and x^1 is the new point $f(x^1) < f(x^0)$. Such a direction s is called a descent direction and satisfies the following requirement at any point $\nabla^T f(x) s < 0$

To see why, examine the two vectors $\nabla f(x^k)$ and s^k. The angle between them is θ, hence $\nabla^T f(x) s^k = |\nabla f(x^k)|\,|s^k| \cos\theta$

If $\theta = 90^\circ$ as in figure then step along s^k do not reduce (improve) the value of $f(x)$. If $0 \le \theta < 90^\circ$, no improvement is possible and $f(x)$ increases. Only if $\theta > 90^\circ$ does the search direction yield smaller values of $f(x)$, hence $\nabla^T f(x^k) s^k < 0$.

We first examine the classic (gradient) steepest descent method of using the gradient and then examine a conjugate gradient method.

- Steepest Descent -

The gradient is the vector at a point x that gives the (local) direction of the greastest rate of increase in $f(x)$. It is orthogonal to the contour of $f(x)$ at x. For maximization, the search direction is simply the gradient (when used the algorithm is called 'steepest ascent') for minimization, the search direction is the negative of the gradient ("steepest descent")

$$s^k = -\nabla f(x^k)$$

In steepest descent at the k^{th} stage, the transition from the current point x^k to the new point x^{k+1} is given by following expression

$$x^{k+1} = x^k + \Delta x^k = x^k + \alpha^k s^k = x^k - \alpha^k \nabla f(x^k)$$

where Δx^k = vector from x^k to x^{k+1}

s^k = search direction, the direction of steepest descent

α^k = Scalar that determines the step length in direction s^k

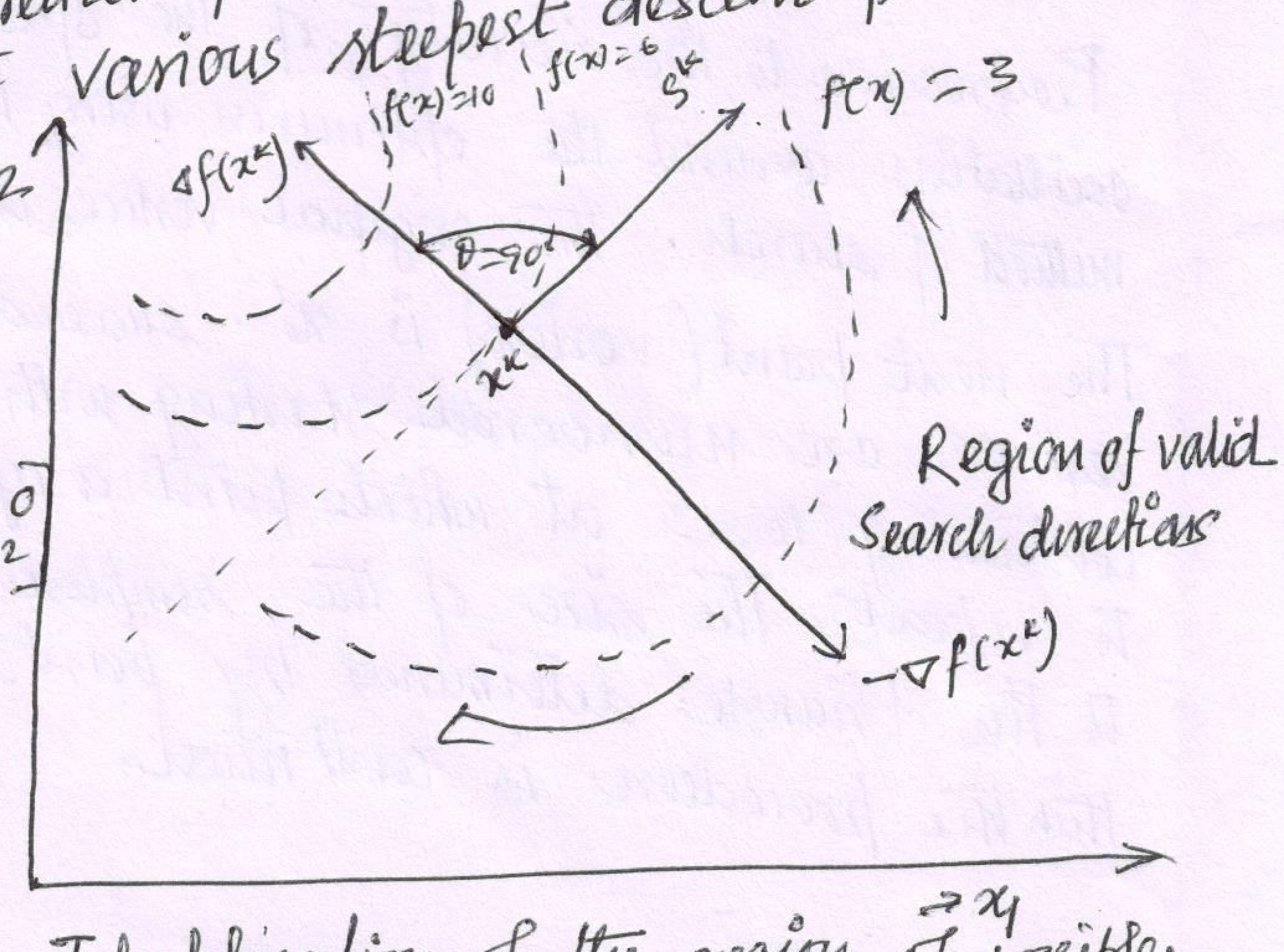
The negative of the gradient gives the direction for minimization but not the magnitude of the step to be taken, so that various steepest descent procedures are possible depending on the choice of α^k. We assume the value of $f(x)$ is continuously reduced.

$f(x) = x_1^2 + x_2^2$ $\quad x^T = [2 \;\; 2]$

$\nabla f(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$ $\quad \nabla f(2,2) = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$ $\quad H(x) = H = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

The direction of steepest descent is

$$s = -\begin{bmatrix} 4 \\ 4 \end{bmatrix}$$



Identification of the region of possible Search directions

A simpler procedure that may result in a suitable value of β is to apply a modified Cholesky factorization as follows

$$H(x^k) + D = LL^T$$

Where D is a diagonal matrix with non-negative elements [$d_{ii} = 0$, if $H(x^k)$ is positive definite] and L is a lower triangular matrix. Upper bounds on the elements in D are calculated using the Gershgorin circle theorem.

A simpler algorithm based on an arbitrary adjustment of β (a modified Marquardt's method) is listed.

The algorithm listed is to be applied to Rosenbrock's function $= 100(x_2 - x_1^2)^2 + (1-x_1)^2$ Starting at $x^0 = [-1.2, 1.0]^T$ with $H^0 = H(x^0)$.

					Elements of $[H(x^k) + \beta I]^{-1}$			
$f(x)$	x_1	x_2	$\frac{\partial f(x)}{\partial x_1}$	$\frac{\partial f(x)}{\partial x_2}$	$\tilde{h}^{-1}_{11}$	$\tilde{h}^{-1}_{12}$	$\tilde{h}^{-1}_{21}$	$\tilde{h}^{-1}_{22}$
24.2000	-1.2000	1.0000	-215.6000	-88.0000	0.0005	-0.0002	-0.0002	0.0009
4.1498	-1.0315	1.0791	2.1844	3.0284	0.0005	-0.0002	-0.0002	0.0009
4.1173	-1.0289	1.0557	-5.2448	-0.5768	0.0014	-0.0013	-0.0013	0.0034
3.9642	-0.9412	0.9301	12.7861	8.8552	0.0037	-0.0059	-0.0059	0.0130
3.4776	-0.8542	0.7098	-10.5031	-3.9772	0.0195	-0.0341	-0.0341	0.0641
2.7527	-0.6028	0.3206	-13.5391	-8.5706	0.0399	-0.0669	-0.0669	0.1170
1.9132	-0.3167	0.0580	-7.9993	-8.4706	0.0464	-0.0557	-0.0557	0.0718
1.1890	-0.0313	-0.0344	-2.5059	-7.0832	0.0579	-0.0328	-0.0328	0.0258
0.6885	0.2278	0.0215	1.2242	-6.0759	0.0616	-0.0039	-0.0039	0.0052
0.3266	0.4570	0.2031	3.2402	-4.5160	0.0706	0.0322	0.0322	0.0196
0.1275	0.6846	0.4520	3.9595	3.3523	0.0906	0.0861	0.0861	0.0818
0.0237	0.8705	0.7495	2.6299	-1.6593	0.1148	0.1573	0.1573	0.2203
0.0006	0.9870	0.9721	0.7700	-0.6033	0.1880	0.3273	0.3273	0.5748
0.0000	0.9974	0.9949	-0.0589	0.0269	0.3563	0.7033	0.7033	1.3922
0.0000	0.9999	0.9999	-0.0004	0.0002	0.5738	1.0249	1.0249	2.0489
0.0000	1.0000	1.0000	.0000	-0.0000	0.5001	1.0001	1.0001	2.0052
0.0000	1	1						

Rosenbrock's Method of Rotating Co-ordinates: This is a modified version of Hooke and Jeeves method in which the coordinate system is rotated in such a way that the first axis always orients to the locally estimated direction of best solution and all the axes are made mutually orthogonal and normal to the first one.

HOOKE-JEEVES

This program finds the minima of a multivariable, unconstrained, nonlinear function Minimize $F(x_1, x_2, \ldots x_n)$

The procedure is based on the direct search method proposed by Hooke and Jeeve. No derivatives are required. The procedure assumes a unimodal function, therefore, if more than one minimum exists or the shape of the surface is unknown, several sets of starting values are recommended. The algorithm proceeds as follows

1) A base point is picked and the objective function evaluated.

2) Local searches are made in each direction by stepping x_i a distance δ_i to each side and evaluating the objective function to see if a lower function is obtained.

3) If there is no function decrease the step size is reduced and searches are made from the previous best point.

4) If the value of the objective function has decreased, a "temporary head" x^{k+1}, is located using the two previous base point x_i^{k+1} and x_i^k

$$x_{i,0}^{k+1} = x_i^{k+1} + \alpha\left(x_i^{k+1} - x_i^k\right)$$

where i is the variable index $= 1, 2, 3 - - N$

0 denotes the temporary head

k is stage index (a stage is the end of N searches)

α is an acceleration factor, $\alpha \geq 1$

5) If the temporary head results in a lower function value, a new local search is performed about the temporary head, a new head is located and the value of F checked. This expansion continues as long as F decreases.

6) If the temporary head does not result in a lower function value, a search is made from the previous best point.

7) The procedure terminates when the convergence criterion is satisfied

Powell's Method

In word's Powell's method to minimize a function $f(x)$ in R^n can be described as follows.

- First initialize n search direction S_i, $i = 1, 2 - - n$ to the co-ordinate unit vector e_i; $i = 1, 2, n$
- Then starting at an initial guess x^0, perform and initial search in the S_n direction which gets you to the point X
- Store X in Y and then update X by performing n successive minimizations along the n search directions
- Create a new search direction $S_{n+1} = X - Y$ and minimize along this direction as well
- After this last search we check for convergence by comparing the relative change in function value at the most recent X with respect to the function value at Y.
- If we have not converged, then we discard the first search direction S_1 and let $S_i = S_{i+1}$, $i = 1, - - n$ and repeat.