Operations Research Techniques and Algorithms (620-361)

Dr Yao-ban Chan

y.chan@ms.unimelb.edu.au
Telephone: 8344 9073
Office: Room 198, Richard Berry Building

Christina Burt
c.burt@ms.unimelb.edu.au
Telephone: 8344 1797
Office: 139 Barry St

Monday 17th March, 2008

Today's Lecture

Unimodal n-D unconstrained optimisation
- Descent methods

Descent methods

In our discussion of single-variable optimisation methods, we moved on from methods that assumed that we could solve $f'(x) = 0$ analytically to iterative methods which did not make this assumption.

The multivariable optimisation methods that we have discussed so far are analogous to first class of single variable methods - essentially they calculate the stationary points by solving $\nabla f(x) = 0$ and then working out which of the stationary points are local minima.

In this and the next few lectures we shall go on to discuss iterative methods for multivariable problems which do not assume that we can solve $\nabla f(x) = 0$ analytically (though we do assume that we can calculate $\nabla f(x)$ and in some cases, $\nabla^2 f(x)$ - what about non-derivative methods ?)

Remember that we are trying to find the minimum of a function. Recursive methods for doing this often involve evaluating the function f and its derivative ∇f at some point, choosing a direction in which f decreases and moving along that direction for some distance. The following definition is relevant.

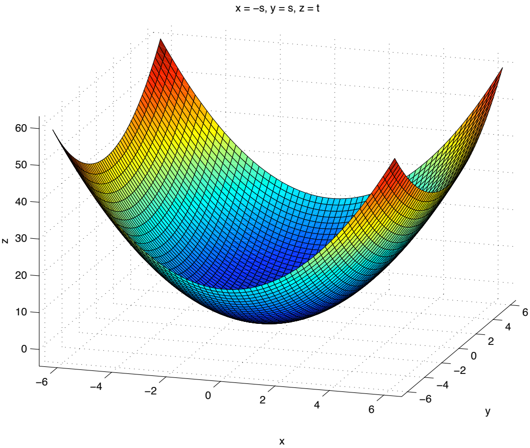
Definition:
Given $x \in \Re^n$, a vector $d \in \Re^n$ is a descent direction *for f at x if*

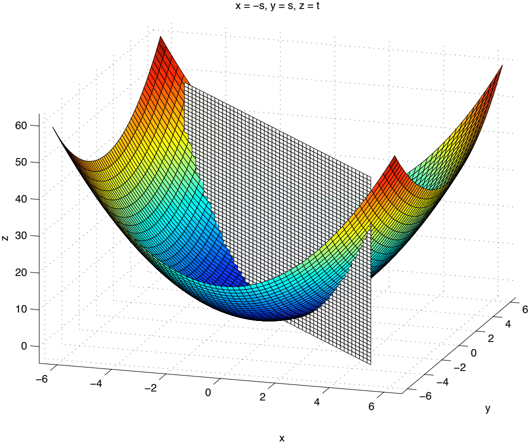
$$\nabla f(x)^T d \; < \; 0$$

Such a direction d is called a descent direction because the function value decreases as we move along it, starting from x, at least initially.

To help you understand why this is so, consider the one-dimensional slice of f in the direction d. Recall that this is a function of a single variable. Recall also that the derivative of this function is given by the directional derivative, $f'(x; d) = \langle \nabla f, d \rangle$. If this is negative, that is if $f'(x; d) < 0$, then the slope of this one-dimensional slice of f is negative as we move along d.

So f decreases, at least initially, as we move along d from x. In other words, d provides a direction of function *descent* from x.


x = −s, y = s, z = t
60
50
40
30
20
10
0
z
−6
−4
−2
0
2
4
6
x
−6
−4
−2
0
2
4
6
y


x = −s, y = s, z = t
60
50
40
30
20
10
0
z
−6
−4
−2
0
2
4
6
x
−6
−4
−2
0
2
4
6
y

Once we have chosen a descent direction d at x, to decrease the value of f we must choose a *step size* or *step length* $t > 0$ such that

$$f(x + td) < f(x).$$

Consider the unconstrained minimization problem

$$\min f(x) = x^2 - 2x$$

where $x \in \Re$.

So $\nabla f = 2x - 2$. If the graph of f is sketched, it is clear that $f(t) = f(0 + t \times 1)$ is less than $f(0)$ for small $t > 0$, i.e. it is likely that $d = 1$ is a descent direction at $x = 0$.

To check the latter, look at

$$\nabla f(0)^T d \ = \ (-2) \times 1 \ = \ -2 \ \leq \ 0,$$

so 1 is a descent direction at 0.

Let's try some different step sizes.

Step size $t = 1$ actually minimizes $f(t)$: $f(1) = -1$.

Step size $t = \frac{1}{2}$ yields a decrease in f:

$$f(\tfrac{1}{2}) = \tfrac{1}{4} - 1 = -\tfrac{3}{4} < 0 = f(0).$$

Step size $t = 3$ causes an increase in f (bad!):

$$f(3) = 9 - 6 = 3 > 0 = f(0).$$

Consider

$$f(x) = x_1^2 + \frac{1}{2}x_2^2$$

where $x = (x_1, x_2) \in \Re^2$; so f is a paraboloid surface.

Question: Given $x = (1, 0)$, what directions d are descent directions for f at x ?

Answer: Look at $\nabla f = (\partial f(x)/\partial x_1, \partial f(x)/\partial x_2) = (2x_1, \frac{1}{2}x_2)$. So $\nabla f(1,0) = (2,0)$ and $\nabla f(1,0)^T d = 2d_1$, where $d = (d_1, d_2)$; hence

$$\nabla f(1,0)^T d \; < \; 0 \quad \iff \quad 2d_1 \; < \; 0 \quad \iff \quad d_1 \; < \; 0.$$

For instance, each of $(-1,2)$, $(-2,0)$, and $(-0.00000001, 10000)$ is a descent direction at $(1,0)$.

Let us consider step sizes for the first two of these descent directions. To help us do so, we consider the one-dimensional slice of the function in each of these directions, positioned with the origin at $x = (1, 0)$. Note that the function value at this point is 1, i.e. $f(1, 0) = 1$.

We first examine direction $d = (-1, 2)$, and consider the function

$$\begin{aligned} \phi_1(t) &= f(x + td) = f((1, 0) + t(-1, 2)) = f(1 - t, 2t) \\ &= (1 - t)^2 + \frac{1}{2}(2t)^2 = 3t^2 - 2t + 1. \end{aligned}$$

From sketching ϕ_1, it is clear that it decreases as t increases from zero, and has a minimum value of $\frac{2}{3}$ at $t = \frac{1}{3}$.

So if we choose to go in this direction, we get the most function decrease by choosing a step size of $t = \frac{1}{3}$.

In this case, the new point would be $x := x + td = (1, 0) + \frac{1}{3}(-1, 2) = (\frac{2}{3}, \frac{2}{3})$, which has function value $f(x) = \frac{2}{3}$, as expected. While this new point has function value less than that of the old point (recall $f(1, 0) = 1$), it is not the minimum value of f, so we need to try again: we should seek a descent direction at this new point, determine a good step size, and repeat.

Now examine direction $d = (-2, 0)$, and consider the function

$$\begin{aligned}\phi_2(t) &= f(x+td) = f((1,0)+t(-2,0)) = f((1-2t,0)) \\ &= (1-2t)^2.\end{aligned}$$

From sketching ϕ_2, it is clear that it also decreases as t increases from zero, and has a minimum value of zero at $t = \frac{1}{2}$. So if we choose to go in this direction, we get the most function decrease by choosing a step size of $t = \frac{1}{2}$.

In this case, the new point would be $x := x + td = (1, 0) + \frac{1}{2}(-2, 0) = (0, 0)$, which has function value $f(x) = 0$. In fact, this is the minimum value for f, so this choice of descent direction was good: it got us to a minimizing point in one step!