Notes on conjugate gradient methods

Written by Peter Sarvari

Based on the Notes by Professor Alessandro Astolfi, Imperial College London

During a line search, we try to minimize the following function

Where is the original function to be minimized, is the point that the minimization algorithm gave as a result after k steps and is the direction of research at the kth step that is set by the minimization algorithm. For example, in case of the gradient method, .

**T1** Now we will prove that in case of exact line search (when is chosen such that is minimized) we have the following relation:

Proof. Let’s suppose at the kth step we perform exact line search. By the necessary condition, we have . Using the definition of the derivative, this means that

Using that

This –by definition of the directional derivative – equals

During minimization, it is important to explore the space extensively. Hence, if n following direction of research (n being the dimension of the minimization problem) are linearly independent, then the space is explored thoroughly.   
**T2** Note that if k vectors are Q-conjugate, then they are linearly independent.

Proof. By definition two vectors are Q-conjugate if the following is true, where Q is a positive definite matrix ().

K vectors are linearly independent if is only true for . Suppose without loss of generality that the jth alpha is non-zero. Now multiply by from the left. Then we have:

All terms will become zero (since ) except for , which has to be zero by (1). Hence since because Q is positive definite. This is in contradiction with our assumption, so all alphas must be zero, i.e. the k vectors are linearly independent.

**T3** Now we will show that choosing the following way will result in exact line search.

Proof. Since , we know that

For a quadratic function , , hence

Left multiplying with yields

By T1, when , we have exact line search.

In conjugate gradient direction methods, the following rule is used to calculate the direction of research at step k:

**T4** We will now show that with this selection, and are Q-conjugate.

Proof.

Since is a scalar, the second term can be rewritten as:

Hence (3) becomes

**T5** We will show that this selection also results in the following equality:

Proof. Using the formula how we select and plugging it into the equation above yields:

Where we used that is a scalar and hence we can move it within the term. Note that by our selection of alpha, as shown by T3. Hence the whole second term disappears.

Since we would like to use this method also when the function to minimize is not quadratic, we need to replace Q with something that can be calculated for any general (smooth) function. One way we can do this is via equation (2).

This condition gives us as a function of Hence, we cannot use it to calculate , because depends on what we choose in the first place. So for non-quadratic function, we still need to use line search. However, if it is close to exact, we retain all the desirable properties. For the calculation of the direction of research, we can plug (4) into the equation for :

The last equality follows from T3.

Furthermore, using T5, we have

Which is the Polak-Ribiere formula.