# Conjugate Gradient Method

The conjugate gradient method is based on the idea that the convergence to the solution could be accelerated if we minimize Q over the hyperplane that contains all previous search directions, instead of minimizing Q over just the line that points down gradient. To determine $x^{i+1}$ we minimize Q over

$$x^0 + span(p^0, p^1, p^2, ..., p^i)$$

where the $p^k$ represent previous search directions. An added advantage to this approach is that, if we can select the $p^k$ to be linearly independent, then the dimension of the hyperplane

$$x^0 + span(p^0, p^1, p^2, ..., p^i)$$

will grow one dimension with each iteration of the conjugate gradient method. This would imply that (assuming infinite precision arithmetic) the solution of the linear system Ax=b would be obtained in no more than N steps, where N is the number of unknowns in the system.

Let $x^0$ be an initial estimate to the solution of **Ax=b**. For our first search direction we proceed down a Q-gradient and choose

$$x^1 = x^0 + \alpha_0 p^0$$

where

$$p^0 = r^0 = -\nabla Q(x^0) = b - Ax$$

From the discussion of the method of steepest descent, we have

$$\alpha_0 = \frac{r^0 \cdot r^0}{r^0 \cdot Ar^0} = \frac{r^0 \cdot p^0}{p^0 \cdot Ap^0}$$

To understand what follows in the description of the conjugate gradient method, it is important to note that

$$r^1 \cdot r^0 = r^1 \cdot p^0 = 0$$

Rather than try to establish the above orthogonality relationships with a calculation, use the following calculus argument. By definition, $r^1$ is the gradient of Q at $x^1$, where $x^1$ is the conjugate gradient estimate to follow the initial guess $x^0$ . Also, $r^0 = p^0$ is the search direction along the line $x^0 + \alpha_0 p^0$. Calculus tells us that the gradient of Q at $x^1$ must be orthogonal to the search direction.

Not as a proof of the last statement about calculus and orthogonality, but to motivate the statement, consider the layers of an onion to be surfaces of Q = constant. Imagine piercing the onion with a skewer. In general, the skewer will pass through several outer layers of the onion, tangentially touch one of the inner layers, again pass through the other layers and then exit. The innermost layer that the skewer touches tangentially is given by $\nabla Q(x^1) = x^1$ and the direction of the skewer is $r^0 = p0$.

What does your mental image of this skewered onion tell you about the orientation of $r^1$ and $r^0 = p^0$?

The conjugate gradient method calls for defining successive approximates by

$$x^{i+1} = x^i + \alpha_i p^i$$

$$p^{i+1} = r^{i+1} + \beta_i p^i$$

with the scheme for choosing $\alpha_i$ and $\beta_i$ to be discussed next. The things to keep in mind when choosing $\alpha_i$ and $\beta_i$ are:

1.We want the span of the search directions to fill the space we are searching as the number of iterations increases;

2.Searching down Q-gradients was basically a good idea. But, to guarantee linearly independent successive search directions, we generally need to choose conjugate gradient search directions to be perturbations of steepest descent search directions.

We have already seen how to define $p^0$ and $\alpha_0$ , so $x^1$ and $r^1 = b - Ax^1$ can be considered known. To take the next step using the conjugate gradient method, we must determine values for $\alpha_0$ and $\beta_1$ so that we can calculate $p^1$ and $x^2$. Then we will see a pattern emerge.

In taking this next step in the conjugate gradient method we are seeking to minimize Q over the plane

$$x^0 + span(p^0, p^1)$$

This means that the residual $r^2$ will have to be orthogonal to both $p^0$ and $p^1$. (To support this last claim, imagine slicing through a lobe of an onion with a knife, forming a flat area spanned by vectors $p^0$ and $p^1$.)

The orthogonality condition $p^0 \cdot r^2 = 0$ will help us set the search direction $p^1$.

$$p^0 \cdot r^2 = p^0 \cdot \left[b - A\left(x^2\right)\right] = p^0 \cdot r^1 - \alpha_1 p^0 \cdot Ap^1$$

which is zero provided

$$\alpha_1 p^0 \cdot Ap^1 = 0$$

**Definition:** Two vectors u and v are said to be A-conjugate if $u \cdot Av = 0$.

From the requirement that $p^0 \cdot r^2 = 0$ , it follows that the search direction $p^1$ must be A-conjugate to the search direction $p^0$. We can now set $\beta_0$ as follows:

$$p^1 = r^1 + \beta_0 p^0$$

implies

$$Ap^1 = Ar^1 + \beta_0 Ap^0$$

Now

$$0 = p^0 \cdot Ap^1 = p^0 \cdot Ar^1 + \beta_0 p^0 \cdot Ap^0$$

implies

$$\beta_0 = -\frac{p^0 \cdot Ar^1}{p^0 \cdot Ap^0} = -\frac{r^1 \cdot Ar^1}{p^0 \cdot Ap^0}$$

Having decided to proceed from $x^1$ to $x^2$ along the search direction defined by $p^1 = r^1 + \beta_0 p^0$, the same calculus argument used to determine gives

$$\alpha_1 = \frac{r^1 \cdot p^1}{p^1 \cdot Ap^1}$$

so a step of the conjugate gradient method is complete. Having developed the CG method for one step it is easy to see that successive iterates are defined as follows.

**Conjugate Gradient Algorithm** (for A symmetric and positive definite)
**Step 1.** (initialize)
Choose
$x^0$
Set
i = 0 and imax = maximum number of iteration to be performed
$r^0 = b - Ax^0$
$p^0 = r^0$

**Step 2:** (Begin CG iteration)
If $i < imax$, Perform Steps 3 - 5
If $i = imax$, Exit and print message "i = imax"

**Step 3:** (Perform CG update)
Set
$\alpha_i = \frac{r^i \cdot p^i}{p^i \cdot Ap^i}$
$x^{i+1} = x^i + \alpha_i p^i$
$r^{i+1} = r^i - \alpha_i Ap^i$

**Step 4:** (check for convergence)
If $\|[r^{i+1}\| < tolerance$ go to Step 6.

**Step 5:** (Prepare for next CG update)
Set
$x^i = x^{i+1}$
$r^i = r^{i+1}$
$\beta_i = -\frac{r^i \cdot Ar^i}{p^i \cdot Ap^i}$

$p^i = r^i + \beta_i p^i$
i = i +1
Go to Step 3

**Step 6:**
print solution and residual norm