The method of steepest descent

The method of steepest descent that we saw in chapter 7 in the section on the conjugate gradient method was a method to find the minimum of a function

$$g(\mathbf{x}) = <\mathbf{x}, A\mathbf{x}> - 2<\mathbf{x},\mathbf{b}>$$

In any steepest descent method, we pick a starting point $\mathbf{x}_0$ that we think is close to the solution. We start our search for the solution by heading in the direction of steepest descent. This direction is the negative of the gradient of $g(\mathbf{x})$ at the starting point $\mathbf{x}_0$.

$$\mathbf{v} = - \nabla g(\mathbf{x}_0) = - \begin{bmatrix} \frac{\partial g}{\partial x_1}(\mathbf{x}_0) \\ \frac{\partial g}{\partial x_2}(\mathbf{x}_0) \\ \vdots \\ \frac{\partial g}{\partial x_n}(\mathbf{x}_0) \end{bmatrix}$$

In chapter 7, we had a further advantage. The function we were trying to find the minimum of had a particularly simple form. This made it easy to see what happened when we searched in the direction of steepest descent. All we had to do was define a function

$$h(t) = g(\mathbf{x}_0 + t\,\mathbf{v})$$

and solve for the t that minimized $h(t)$. Given the simple form of the function $g(\mathbf{x})$ in chapter 7, finding the minimum for $h(t)$ was fairly simple.

In the more general steepest descent method problem, we are trying to find the minimum of a more general function $g(\mathbf{x})$ from $\mathbb{R}^n$ to $\mathbb{R}$. Once again, we form

$$\mathbf{v} = - \nabla g(\mathbf{x}_0) = - \begin{bmatrix} \frac{\partial g}{\partial x_1}(\mathbf{x}_0) \\ \frac{\partial g}{\partial x_2}(\mathbf{x}_0) \\ \vdots \\ \frac{\partial g}{\partial x_n}(\mathbf{x}_0) \end{bmatrix}$$

and

$$h(t) = g(\mathbf{x}_0 + t\,\mathbf{v})$$

and seek the value of t that minimizes $h(t)$. The problem this time around is that $g(\mathbf{x})$ does not necessarily have a simple form, and consequently $h(t)$ will be a more complex function. In principle, we can find the minimum of $h(t)$ by differentiating with respect to t and locating critical points of

$h'(t)$. In practice, this is often too difficult to do.

As an alternative, we can try the following approach.

1. Compute $h(0)$.
2. Find a value of $x > 0$ such that $h(x) < h(0)$. (Start by trying $x = 1$, and keep halving x until $h(x) < h(0)$.)
3. Construct a quadratic polynomial in t that interpolates $(0,h(0))$, $(x/2,h(x/2))$, and $(x,h(x))$.
4. Find the value t that minimizes that quadratic polynomial.
5. Construct $\mathbf{x}_1 = \mathbf{x}_0 + t\,\mathbf{v}$.
6. Set $\mathbf{x}_0 = \mathbf{x}_1$, recompute $\mathbf{v}$, and repeat steps 1-5 until $g(\mathbf{x}_0)$ is small enough.

Finding roots and the method of steepest descent

In chapter 10 we are concerned with finding roots of vector valued functions

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(x_1,x_2,\ldots,x_n) \\ f_2(x_1,x_2,\ldots,x_n) \\ \vdots \\ f_n(x_1,x_2,\ldots,x_n) \end{bmatrix}$$

that map $\mathbb{R}^n$ to $\mathbb{R}^n$. If $\mathbf{f}(\mathbf{x})$ has a root at $\mathbf{x}$, the function

$$g(\mathbf{x}) = \sum_{k=1}^{n} \left(f_k(x_1,x_2,\ldots,x_n)\right)^2$$

has a minimum (of 0) at that same $\mathbf{x}$. Thus, to find roots, we construct the function $g(\mathbf{x})$ and use the method of steepest descent to find $\mathbf{x}$ that minimize $g(\mathbf{x})$.

The more practical approach is to use the method of steepest descent for a few iterations and then just use Newton's method from there. Newton's method only works well if we can get ourselves close to the root, so it makes sense to use the method of steepest descent to get closer to the root first.