Converting a root finding problem to a minimization problem

Suppose we are given the problem of finding a root for a non-linear, vector-valued function $F(\mathbf{x})$.

$$F(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{bmatrix} = \mathbf{0}$$

We can turn this into a minimization problem by introducing the function

$$g(\mathbf{x}) = \sum_{i=1}^{n} \left(f_i(\mathbf{x})\right)^2$$

Since $g(\mathbf{x}) \geq 0$ for all $\mathbf{x}$, it is easy to see that $g(\mathbf{x})$ has a minimum where $F(\mathbf{x})$ has a root.

The primary advantage to this transformation is that it allows us to solve the root finding problem by using techniques normally used to locate minima.

The method of steepest descent

We know from having studied this problem earlier that we can make progress toward the minimum by searching in the direction given by the negative gradient, since the negative gradient gives you the local direction of steepest descent.

The only problem with this approach is that the local direction of steepest descent is not the global direction of steepest descent. That is, if you are currently in a location $\mathbf{x}_0$ and the global minimum is located at $\mathbf{x}_{min}$, there is no guarantee that $\mathbf{x}_{min}$ - $\mathbf{x}_0$ lies in the same direction as the negative gradient. Even so, by following the negative gradient we hope to at least make some progress toward the minimum.

Here now is a procedure that stands a chance of at least getting us to a better position.

1. Pick a starting guess $\mathbf{x}_0$.
2. Compute $-\nabla g(\mathbf{x}_0)$, the local direction of steepest descent.
3. Construct the function $h(\alpha) = g(\mathbf{x}_0 - \alpha \nabla g(\mathbf{x}_0))$.
4. Find the value α_0 that minimizes $h(\alpha)$.
5. Compute $\mathbf{x}_1 = \mathbf{x}_0 - \alpha_0 \nabla g(\mathbf{x}_0)$.

Computationally, the most difficult step in the procedure outlined here is step 4. If we want to solve for the miminizing α we would have to differentiate $h(\alpha)$ with respect to α, solve for a critical value of α, and confirm that we have a true minimum. Because these steps are difficult to carry out in practice, we most commonly replace step 4 with an approximation procedure that attemps to estimate where the minimizing α is. Here is that procedure.

1. Pick three distinct values of α: α_1, α_2, and α_3
2. Construct a second degree polynomial that interpolates the points $(\alpha_1 , h(\alpha_1))$, $(\alpha_2 , h(\alpha_2))$, and $(\alpha_3 , h(\alpha_3))$.
3. Find the value α_0 that minimizes that interpolating polynomial.

Mixing steepest descent and Newton's method

In section 10.2 we saw that Newton's method is a very effective root-finding technique for finding roots of nonlinear vector-valued functions. One weakness in Newton's method is that it does not work well if our initial guess is too far from the actual root. The method of steepest descent can be used as a preliminary step before using Newton, since steepest descent does at least guarantee that you will end up closer to the root each time you apply it.