Termination Conditions for Line Search Methods

For Line Search Methods, we typically use the following options for termination criteria. The first three conditions are standard and can be used with any Line Search Method, the fourth is called the Newton Decrement, and is applicable only to the Newton method. $\epsilon_{grad}, \epsilon_{obj}$ and ϵ_{step} are numeric tolerances we use as constants in our algorithm runs (for example, $10^{-8}, 10^{-6}$, etc.), and the p_k in #4 denotes the Newton direction obtained in the k'th iteration.

1. $\|\nabla f(x_k)\| < \epsilon_{grad}$
2. $\|x_{k+1} - x_k\| < \epsilon_{step}$
3. $\|f(x_{k+1}) - f(x_k)\| < \epsilon_{obj}$
4. $\frac{1}{2} p_k^T \nabla^2 f(x_k) p_k < \epsilon_{obj}$

The Newton Decrement

We now introduce the underlying motivation and derivation of alternative expressions for the Newton decrement. Note that the first alternative we will arrive at involves the inverse of the Hessian, which we do not want to explicitly compute, at any stage of our algorithm. Hence, we will develop it further, arriving at the condition appearing above: the quadratic form of the Hessian (not the inverse) operating on the Newton step.

Definition

For Newton's method, assume $\nabla^2 f(x)$ is positive definite, and recall that the Newton step is given by:

$$p_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$$

First, define the Newton Decrement at x to be the quantity:

$$\lambda(x) = [\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)]^{\frac{1}{2}}$$

Make sure the dimensions are clear in this matrix multiplication: it is a quadratic expression Hessian inverse, operating on the gradient, hence a scalar, and then its square root.

Relation to decrease in objective function values

We now relate $\lambda(x)$ to the estimated decrease in function values, when taking the Newton step. At the constant location $x \in \mathbb{R}^n$, the quadratic approximation $\hat{f}$ of f near x (written as a function of $y \in \mathbb{R}^n$) is:

$$\hat{f}(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x)(y - x)$$

This above expression is the second order polynomial that we have used to motivate the Newton method and derive the step.

Now we consider $f(x) - \hat{f}(x+p)$ where p is the Newton step, namely, we give an estimate on the decrease in function value if we take the Newton step:

$$f(x) - \hat{f}(x+p) = f(x) - f(x) - \nabla f(x)^T p - \frac{1}{2} p^T \nabla^2 f(x) p$$

We substitute the step: $p = -\nabla^2 f(x)^{-1} \nabla f(x)$ and obtain:

$$f(x) - \hat{f}(x+p) = -\nabla f(x)^T [-\nabla^2 f(x)^{-1} \nabla f(x)] - \frac{1}{2} [-\nabla^2 f(x)^{-1} \nabla f(x)]^T \nabla^2 f(x) [-\nabla^2 f(x)^{-1} \nabla f(x)]$$

After the transpose and minus signs, rearranging gives:

$$f(x) - \hat{f}(x+p) = \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) - \frac{1}{2} \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla^2 f(x) [\nabla^2 f(x)^{-1} \nabla f(x)]$$

Which, after the Hessian and its inverse cancel out, yields:

$$f(x) - \hat{f}(x+p) = \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) - \frac{1}{2} \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) = \frac{1}{2} \lambda(x)^2$$

Hence we have shown that $\frac{1}{2}\lambda(x)^2$ is an estimate on the decrease in the objective function values, in the following sense: it the difference between the objective value at the current location and the value of its current quadratic approximation, upon taking the Newton step.

<u>Usage as a termination condition in the algorithm</u>

In the preceding section we motivated using $\frac{1}{2}\lambda^2 < \epsilon_{obj}$ as a termination condition for the algorithm, where ϵ_{obj} is a numeric tolerance on the objective value differences. We now find an alternative expression for λ, such that all involved quantities are available to us at each iteration. Specifically, we will not need the Hessian inverse: we already solved the system when obtaining the Newton step – we can use that.

To this end, consider the quadratic expression:

$$p^T \nabla^2 f(x) p,$$

where p is the Newton step. Again, substitute $p = -\nabla^2 f(x)^{-1} \nabla f(x)$, as in the definition of the Newton step, and we have, again after Hessian and inversed cancel out:

$$p^T \nabla^2 f(x) p = [-\nabla^2 f(x)^{-1} \nabla f(x)]^T \nabla^2 f(x) [-\nabla^2 f(x)^{-1} \nabla f(x)] = \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) = \lambda(x)^2$$

Thus, we can compute $\frac{1}{2}\lambda(x)^2$ for our termination condition, not using the original definition, which involves the Hessian inverse, but via $\frac{1}{2} p^T \nabla^2 f(x) p$, for which all involved vectors and matrix are already computed.