Quasi-Newton Methods

Saurav Samantaray

Department of Mathematics

Indian Institute of Technology Madras

March 2, 2024

Quasi-Newton Methods

- Quasi-Newton methods (like steepest descent), require only the gradient of the objective function to be supplied at each iterate.
- By measuring the changes in gradients, they construct a model of the objective function that is good enough to produce super-linear convergence.
- The improvement over steepest descent is dramatic, especially on difficult problems.
- Moreover, since second derivatives are not required (unlike Newton's), quasi-Newton methods are sometimes more efficient than Newton's method.
- Today, optimization software libraries contain a variety of quasi-Newton algorithms for solving unconstrained, constrained, and large-scale optimization problems.

THE BFGS METHOD

- BFGS method, named for its discoverers Broyden, Fletcher, Goldfarb, and Shanno.
- Consider the following quadratic model of the objective function at the current iterate x_k :

$$m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p. \tag{1}$$

B_k is a $n \times n$ symmetric positive definite matrix that will be revised or updated at every iteration.

- As the model is first-order accurate the function value f_k and the gradient ∇f_k, both match at $p = 0$.
- The minimiser of this model is

$$p_k = -B_k^{-1} \nabla f_k, \tag{2}$$

THE BFGS METHOD

- p_k is used as the search direction, and the new iterate is

$$x_{k+1} = x_k + \alpha_k p_k \tag{3}$$

  where the step length α_k is chosen to satisfy the Wolfe conditions.
- Instead of computing B_k afresh at every iteration, Davidon proposed to update it in a simple manner to account for the curvature measured during the (most recent) previous step.
- Suppose that we have generated a new iterate x_{k+1} and wish to construct a new quadratic model, of the form

$$m_{k+1}(p) = f_{k+1} + \nabla f_{k+1}^T p + \frac{1}{2} p^T B_{k+1} p. \tag{4}$$

- What requirements should we impose on B_{k+1}, based on the knowledge gained during the latest step?

THE BFGS METHOD

- One reasonable requirement is that the gradient of m_{k+1} should match the gradient of the objective function f at the latest two iterates x_k and x_{k+1}, i.e.

  $\nabla m_{k+1}(0) = \nabla f_{k+1}$ (condition at the second point is satisfied)

- The first condition can be written mathematically as

$$\nabla m_{k+1}(-\alpha_k p_k) = \nabla f_{k+1} - \alpha_k B_{k+1} p_k = \nabla f_k$$

- By rearranging we get

$$B_{k+1} \alpha_k p_k = \nabla f_{k+1} - \nabla f_k \tag{5}$$

THE BFGS METHOD

- Introduce the notations

$$s_k = x_{k+1} - x_k = \alpha_k p_k$$
$$y_k = \nabla f_{k+1} - \nabla f_k$$

- With the above notation (5) becomes

$$B_{k+1} s_k = y_k \tag{6}$$

  which is called the secant equation.

- Given the displacements s_k and the change of gradients y_k,the secant equation requires that the symmetric positive definite matrix B_{k+1} map s_k into y_k.