Ch.9 Newton's Method

9.1 Introduction

- Recall that the method of steepest descent uses only first derivatives (gradients) in selecting a suitable search direction. $\rightarrow$ not always effective.
- Newton's method uses first and second derivatives and performs better than the steepest descent method if the initial point is close to the minimizer.
- Quadratic approximation using the Taylor series expansion of f about the current point $x^{(k)}$

$$\begin{aligned} f(x) \quad &\approx \quad f(x^{(k)}) + (x - x^{(k)})^T g^{(k)} + \frac{1}{2}(x - x^{(k)})^T F(x^{(k)}) \\ & \qquad \times (x - x^{(k)}) \equiv q(x) \end{aligned}$$

- Applying the FONC to q yields

$$0 = \nabla q(x) = g^{(k)} + F(x^{(k)})(x - x^{(k)})$$

- If $F(x^{(k)}) > 0$, then q achieves a minimum at

$$x^{(k+1)} = x^{(k)} - F(x^{(k)})^{-1} g^{(k)}$$

- This recursive formula represents Newton's method.
- Example 9.1 Use Newton's method to minimize the

Powell function:

$$\begin{aligned} f(x_1,x_2,x_3,x_4) &= (x_1+10x_2)^2+5(x_3-x_4)^2 \\ &\quad +(x_2-2x_3)^4+10(x_1-x_4)^4 \end{aligned}$$

Use as the starting point $x^{(0)} = [3,-1,0,1]^T$.

$$\nabla f(x) = \begin{bmatrix} 2(x_1+10x_2)+40(x_1-x_4)^3 \\ 20(x_1+10x_2)+4(x_2-2x_3)^3 \\ 10(x_3-x_4)-8(x_2-2x_3)^3 \\ -10(x_3-x_4)-40(x_1-x_4)^3 \end{bmatrix}$$

$$F(x) = \begin{bmatrix} \cdots & \cdots \\ \cdots & \cdots \end{bmatrix}$$

Iteration 1.: find
$g^{(0)},\ F(x^{(0)}),\ F(x^{(0)})^{-1},\ F(x^{(0)})^{-1}g^{(0)}$
$x^{(1)} = x^{(0)} - F(x^{(0)})^{-1}g^{(0)} = [1.58,-0.15,0.25,0.25]^T.$
$f(x^{(1)}) = 31.8$
Iteration 2.: find
$g^{(1)},\ F(x^{(1)}),\ F(x^{(1)})^{-1},\ F(x^{(1)})^{-1}g^{(1)}$
$x^{(2)} = x^{(1)} - F(x^{(1)})^{-1}g^{(1)} = [1.05,-0.10,0.16,0.16]^T.$
$f(x^{(2)}) = 6.28.$
Iteration 3.: find
$g^{(2)},\ F(x^{(2)}),\ F(x^{(2)})^{-1},\ F(x^{(2)})^{-1}g^{(2)}$
$x^{(3)} = x^{(2)} - F(x^{(2)})^{-1}g^{(2)} = [0.70,-0.07,0.11,0.11]^T.$
$f(x^{(3)}) = 1.24.$

9.2 Analysis of Newton's Method

- Newton's method reaches the point x^* such that $\nabla f(x^*) = 0$ in just one step starting from any initial point $x^{(0)}$.
- To see this, suppose that $Q = Q^T$ is invertible, and

$$\begin{aligned} f(x) &= \frac{1}{2}x^T Q x - x^T b \\ g(x) &= \nabla f(x) = Qx - b \\ F(x) &= Q \\ x^{(1)} &= x^{(0)} - F(x^{(0)})^{-1} g^{(0)} \\ &= Q^{-1} b = x^* \end{aligned}$$

- **Theorem 9.1** *Suppose that $f \in C^3$, and $x^* \in \Re^n$ is a point such that $\nabla f(x^*) = 0$ and $F(x^*)$ is invertible. Then, for all $x^{(0)}$ sufficiently close to x^*, Newton's method is well defined for all k, and converges to x^* with order of convergence at least 2.*
- However, the method is not guaranteed to converge to the solution if we start far away from the solution.
- **Theorem 9.2** *Let $\{x^{(k)}\}$ be the sequence generated by Newton's method for minimizing a given objective function $f(x)$. If the Hessian $F(x^{(k)}) > 0$ and*

$g^{(k)} = \nabla f(x^{(k)}) \neq 0$, then the direction

$$d^{(k)} = -F(x^{(k)})^{-1} g^{(k)} = x^{(k+1)} - x^{(k)}$$

from $x^{(k)}$ to $x^{(k+1)}$ is a descent direction for f in the sense that there exists an $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha})$,

$$f(x^{(k)} + \alpha d^{(k)}) < f(x^{(k)})$$

- The above theorem motivates the following modification of Newton's method:

$$\begin{aligned} x^{(k+1)} &= x^{(k)} - \alpha_k F(x^{(k)})^{-1} g^{(k)} \\ \alpha_k &= \arg\min_{\alpha \geq 0} f(x^{(k)} - \alpha F(x^{(k)})^{-1} g^{(k)}) \end{aligned}$$

9.3 levenberg-Marquardt Modification

- Levenberg-Marquardt modification to Newton's algorithm:

$$x^{(k+1)} = x^{(k)} - (F(x^{(k)}) + \mu_k I)^{-1} g^{(k)}$$

  where $\mu_k \geq 0$.

- Eigenvalues of G are $\lambda_1 + \mu, \cdots, \lambda_n + \mu$.

$$\begin{aligned} G v_i &= (F + \mu I) v_i \\ &= (\lambda_i + \mu) v_i \end{aligned}$$

- If μ is sufficiently large, then all the eigenvalues of G are positive, and G is positive definite.

- Accordingly, if the parameter μ_k in the Levenberg-Marquardt modification of Newton's algorithm is sufficiently large, then the search direction $d^{(k)} = -(F(x^{(k)}) + \mu_k I)^{-1} g^{(k)}$ always points in a descent direction.