Newton’s method

Exact relation:

\[ f(x) = f(x_k) + (x - x_k)^T \nabla f(x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(\xi)(x - x_k), \]

where \( \xi = \alpha x_k + (1 - \alpha)x \) for some \( \alpha \in [0, 1] \).

Quadratic model for Newton:

\[ \phi_k(x) = f(x_k) + (x - x_k)^T \nabla f(x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k) \]

Recall: The solution to \( Hs = -g \) is the unique minimizer of \( \phi(s) = \alpha + g^T s + \frac{1}{2}s^T Hs \), for \( H \) symmetric, positive definite.

Thus, if \( \nabla^2 f(x_k) \) is positive definite, \( \phi_k \) has a unique minimum for

\[ x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k) \]

Algorithm (Basic Newton)

1. Choose initial guess \( x_0 \) & tolerance \( \epsilon \)
2. For \( k = 0, 1, \ldots \)
   2.1 If \( \| \nabla f(x_k) \| \leq \epsilon \) stop
   2.2 Solve \( \nabla^2 f(x_k)s_k = -\nabla f(x_k) \)
   2.3 Set \( x_{k+1} = x_k + s_k \)

Newton’s method converges quadratically to a local minimum \( x^* \)

- if the Hessian is positive definite at \( x^* \), and
- if started close enough to the solution

Problems with basic Newton

(i) Needs derivatives: may be impossible or expensive to compute
(ii) Needs solution of a linear system each iteration. Demanding for large problems
(iii) Converges only when starting close enough
(iv) The Hessian may be singular. Yields problems to solve the linear system

Guaranteeing descent (problem (iv))

- \( p \) is a descent direction for \( f \) at \( x_k \) if \( p^T \nabla f(x_k) < 0 \)
- If \( p \) is a descent direction for \( f \) at \( x_k \), then \( f(x_k + \alpha p) < f(x_k) \) for a sufficiently small \( \alpha > 0 \)
- The Newton direction \( s_k = - (\nabla^2 f(x_k))^{-1} \nabla f(x_k) \) is a descent direction if the Hessian \( \nabla^2 f(x_k) \) is positive definite
- The Cholesky decomposition: an alternative to \( LU \)-factorization for symmetric positive definite matrices
  - Needs no pivoting
  - Half the computational effort and half the storage compared to \( LU \)
- The Cholesky decomposition can be modified in order to produce a consistently positive definite matrix: \( LL^T = E + \nabla^2 f(x_k) \), where \( E \geq 0 \) diagonal so that \( LL^T \) positive definite
Globalization—line search (problem (iii))

- A general descent method: \( x_{k+1} = x_k + \alpha_k p_k \), where \( p_k \) is a descent direction
- \( p_k \) descent direction: \( f(x_k + \alpha p_k) < f(x_k) \) for \( \alpha > 0 \) small enough
- Too large \( \alpha \) may cause \( f(x_k + \alpha p_k) > f(x_k) \)!
- The line search subproblem: \( \min_{\alpha > 0} f(x_k + \alpha p_k) \):
- Not enough to demand \( f(x_k + \alpha p_k) < f(x_k) \) (to guarantee convergence). Convergence theory needs a stronger condition, such as the Armijo condition:

\[
\begin{align*}
  f(x_k + \alpha p_k) &\leq f(x_k) + \alpha \mu p_k^T \nabla f(x_k) \\
\end{align*}
\]

Very small \( \mu = 10^{-4} \) OK.

Damped Newton

Line search with Newton’s method is called **Damped Newton**

Backtracking line search:
- Start with \( \alpha = 1 \)
- Check the Armijo condition
- If needed, reduce \( \alpha \) until the Armijo condition is satisfied.
- \( \alpha < 1 \) only needed when far from the solution. When close enough, Newton will converge without step size reduction.

Damped Newton with modified Cholesky

Algorithm

1. Choose initial guess \( x_0 \), tolerance \( \epsilon \), and parameter \( \mu \)
2. For \( k = 0, 1, \ldots \)
   2.1 If \( \| \nabla f(x_k) \| \leq \epsilon \) stop
   2.2 Compute modified factorization \( LL^T \) of \( \nabla^2 f(x_k) \)
   \((LL^T = E + \nabla^2 f(x_k) \) for some diagonal \( E \geq 0 \))
   2.3 Solve \( LL^T s_k = -\nabla f(x_k) \)
   2.4 \( \alpha_k \leftarrow 1 \)
   2.5 While \( f(x_k + \alpha_k s_k) > f(x_k) + \mu \alpha_k s_k^T \nabla f(x_k), \)
      \( \alpha \leftarrow \alpha/2 \)
   2.6 Set \( x_{k+1} \leftarrow x_k + \alpha s_k \)