The steepest-descent method

Computing the Newton step with the crude approximation $\nabla^2 f(x_k) \approx I$ yields $s_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k) \approx -\nabla f(x_k)$

The negative gradient:
- Is perpendicular to the level curves
- Points in the steepest downhill direction

The steepest-descent direction is the best direction locally

Quasi-Newton methods

Exact relation:
$$f(x) = f(x_k) + (x - x_k)^T \nabla f(x_k) + \frac{1}{2} (x - x_k)^T \nabla^2 f(x_k) (y - x),$$
where $\xi = \alpha x_k + (1 - \alpha)x$ for some $\alpha \in [0, 1]$.

Newton:
$$\phi_k^N(x) = f(x_k) + (x - x_k)^T \nabla f(x_k) + \frac{1}{2} (x - x_k)^T \nabla^2 f(x_k) (y - x)$$

Steepest-descent:
$$\phi_k^{SD}(x) = f(x_k) + (x - x_k)^T \nabla f(x_k) + \frac{1}{2} (x - x_k)^T I (y - x)$$

Quasi-Newton:
$$\phi_k^{QN}(x) = f(x_k) + (x - x_k)^T \nabla f(x_k) + \frac{1}{2} (x - x_k)^T B_k (y - x)$$

The secant equation yields $n$ equations for the $n^2$ coefficients of $B_k$:
$$B_{k+1} s_k = y_k,$$
where
$$s_k = x_{k+1} - x_k,$$
$$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

The symmetric rank-one update:
$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{y_k - B_k s_k} s_k$$

the only rank-one update satisfying the secant equation and preserving symmetry.

The BFGS update is a good, rank-two update, satisfying the secant equation, and preserving symmetry and positive-defiteness.
The BFGS quasi-Newton method with line search

1. Specify initial guess $x_0$, $B_0 (= I$ say)
2. For $k = 0, 1, \ldots$
   2.1 Check convergence (say, the size of $\| \nabla f(x_k) \|$)
   2.2 Solve $B_k p_k = -\nabla f(x_k)$
   2.3 $\alpha_k \leftarrow 1$
   2.4 While $f(x_k + \alpha_k p_k) > f(x_k) + \mu \alpha_k p_k^T \nabla f(x_k)$
      reduce $\alpha_k$ ($\alpha_k \leftarrow \alpha_k/2$ or something smarter)
   2.5 $x_{k+1} = x_k + \alpha_k p_k$
   2.6 $s_k = x_{k+1} - x_k$
   2.7 $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$
   2.8 $B_{k+1} = B_k - \frac{(B_k s_k)(B_k s_k)^T}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}$