Quasi-Newton methods

Exact relation:

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

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

Newton:

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

Steepest-descent:

\[ \phi^{SD}_k(x) = f(x_k) + (x - x_k)^T \nabla f(x_k) + \frac{1}{2} (x - x_k)^T I(x - x_k) \]

Quasi-Newton:

\[ \phi^{QN}_k(x) = f(x_k) + (x - x_k)^T \nabla f(x_k) + \frac{1}{2} (x - x_k)^T B_k(x - x_k) \]
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**:

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}$$

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-definiteness.

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 $B_{k+1} = B_k - \frac{(B_k s_k)(B_k s_k)^T}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}$
Quadratic forms

\[ f(x) = \frac{1}{2} x^T A x = \frac{1}{2} x^T A_S x \]

where

\[ A_S = \frac{1}{2} (A + A^T) \]

Example: \( A_S = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \). Then

\[
\begin{align*}
(x_1 & \ 0) A_S \begin{pmatrix} x_1 \\ 0 \end{pmatrix} = x_1^2 \\
(0 & \ x_2) A_S \begin{pmatrix} 0 \\ x_2 \end{pmatrix} = -x_2^2
\end{align*}
\]
Quadratic forms

The spectral theorem: any real symmetric matrix $A$ can be diagonalized with an orthogonal matrix of eigenvectors to $A$

$$Q^T AQ = \Lambda, \quad A = Q\Lambda Q^T$$

where $Q^T Q = I$, $\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots \\ 0 & \lambda_2 & \cdots \\ \vdots & \vdots & \ddots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$

$$f(x) = x^T Ax = x^T Q\Lambda Q^T x = \tilde{x}^T \Lambda \tilde{x}, \text{ with } \tilde{x} = Q^T x$$

- $\tilde{x} = Q^T x$ is an orthogonal coordinate change (rotations and reflections)
- In the new coordinates $\tilde{x}$, back to situation on slide 4
  - $\lambda_n > 0$ indicates convex direction $\tilde{x} = e_n$
  - $\lambda_m < 0$ indicates concave direction $\tilde{x} = e_m$