1 General

line search conditions:

$$
\begin{aligned}
f_{k+1}^T &\leq f_k + c_1 \alpha_k \nabla f_k^T p_k, c_1, \alpha_k \in (0,1) && (1)\\
\nabla f_{k+1}^T p_k &\geq c_2 \nabla f_k^T p_k, 0 < c_1 < c_2 < 1 && (2)\\
&\text{where :} && (3)\\
f_k &= f(x_k) && (4)\\
f_{k+1} &= f(x_k + \alpha_k p_k) && (5)
\end{aligned}
$$

2 Quasi Newton

2.1 BFGS

properties: $O(n^2)$, self correcting, slightly more iterations than Newton Method, linear convergence order and superlinear rate of convergence

secant equation:

$$
\begin{aligned}
B_{k+1}(x_{k+1} - x_k) &= \nabla f_{k+1} - \nabla f_k\\
B_{k+1} s_k &= y_k\\
s_k &= \alpha_k p_k\\
y_k &= \nabla f_{k+1} - \nabla f_k
\end{aligned}
$$

$$
\begin{aligned}
B_{k+1} &\succ 0\\
s_k^T B_{k+1} s_k &= s^T y_k > 0
\end{aligned}
$$

Proof.

$$
\begin{aligned}
y_k^T s_k &= (\nabla f_{k+1} - \nabla f_k)^T s_k\\
\nabla f_{k+1}^T s_k &\geq c_2 \nabla f_k^T s_k\\
(\nabla f_{k+1} - \nabla f_k)^T s_k &\geq c_2 \nabla f_k^T s_k - \nabla f_k^T s_k\\
y_k^T s_k &\geq (c_2 - 1) \nabla f_k^T s_k
\end{aligned}
$$

$$c_2 < 1, s_k \text{ is a descent dir} \implies s_k^T y_k > 0$$

Curvature condition holds. □

constrain B by solving:

$$
\begin{aligned}
&\min_B ||B - B_k||\\
&s.t.\ B = B^T, B s_k = y_k
\end{aligned}
$$

similarly, constrain B's inverse, H where it satisfy secant equation:

$$H_{k+1} y_k = s_k$$

$$
\begin{aligned}
&\min_H ||H - H_k||\\
&s.t.\ H = H^T, H y_k = s_k
\end{aligned}
$$

using weighted Frobenius norm:

$$
\begin{aligned}
||A||_W &:= ||W^{1/2} A W^{1/2}||_F\\
||X||_F &:= (\sum_i \sum_j (X_{ij})^2)^{1/2}
\end{aligned}
$$

solved weight matrix W satisfy $W s_k = y_k$ solution given by:

$$
\begin{aligned}
H_{k+1} &= (I - \rho_k s_k y_k^T) H_k (I - \rho_k y_k s_k^T)\\
&\quad + \rho_k s_k^T s_k\\
\rho_k &= \frac{1}{y_k^T s_k}
\end{aligned}
$$

W is the average Hessian $\bar{G}$:

$$\bar{G} = \int_0^1 \nabla^2 f(x_k + \tau \alpha_k p_k) d\tau$$

initial H_0 can be chosen approximately (eg: finite differences, I)

Algorithm 1: BFGS Algorithm

$H_0, x_0, \epsilon > 0$**:** inverse Hessian approx., initial point, convergence tolerance
x **:** solution

$k \leftarrow 0$
while $||\nabla f_k|| > \epsilon$ **do**
  $\alpha_k \leftarrow$ LineSearch(..)
  $x_{k+1} \leftarrow x_k + \alpha_k p_k$
  $s_k \leftarrow x_{x+1} - x_k$
  $y_k \leftarrow \nabla f_{k+1} - \nabla f_k$
  $\rho_k \leftarrow \frac{1}{y_k^T s_k}$
  $H_{k+1} \leftarrow (I - \rho_k s_k y_k^T) H_k (I - \rho_k y_k s_k^T)$
  $+ \rho_k s_k^T s_k$
  $k \leftarrow k + 1$

using Sherman-Morrison-Woodbury formula to obtain Hessian update equation:

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

proper line search is required so that BFGS algo captures curvature information

inaccurate line search can be used to reduce computation cost

3 Trust Region Methods