Analysis of BFGS

Giuseppe Giorgio Colabufo
giuseppe.colabufo@polytechnique.edu

2019

BFGS Algorithm

1. Obtain a direction d_k by $d_k = -B_k^{BFGS}\nabla f(x_k)$.
2. Perform a one-dimensional optimization (line search) to find an acceptable step-size α_k in the direction found in the first step, so $\alpha_k = \arg\min_{\alpha} f(x_k + \alpha d_k)$.
3. Set $p_k = \alpha_k d_k$ and update $x_{k+1} = x_k + p_k$.
4. $q_k = \nabla f(x_{k+1}) - \nabla f(x_k)$.
5. $B_{k+1}^{BFGS} = B_k^{BFGS} + \frac{(p_k^T q_k + q_k^T B_k^{BFGS} q_k)(p_k p_k^T)}{(p_k^T q_k)^2} - \frac{B_k^{BFGS} q_k p_k^T + p_k q_k^T B_k^{BFGS}}{p_k^T q_k}$.

We consider the update for the matrix B as follows:

$$B_{k+1} = \begin{cases} B_k + \frac{(p_k^T q_k + q_k^T B_k q_k)(p_k p_k^T)}{(p_k^T q_k)^2} - \frac{B_k q_k p_k^T + p_k q_k^T B_k}{p_k^T q_k} & p_k \neq 0 \\ B_k & p_k = 0 \end{cases} \tag{1}$$

Remark 0.0.1. In the previous update we only consider the case in which $p_k = 0$ because $p_k = \iff q_k = 0$.

Proof. If $p_k = 0$ then $x_{k+1} = x_k$ and $\nabla f(x_{k+1}) = \nabla f(x_k)$, i.e. $q_k = 0$.
On the other hand, if $q_k = 0$, i.e. $\nabla f(x_{k+1}) = \nabla f(x_k + \alpha_k d_k) = \nabla f(x_k)$, there are two cases:

- $\alpha_k = 0$ that means $p_k = 0$.
- $\alpha_k \neq 0$, but since $\alpha_k = \arg\min \phi(\alpha) = \arg\min f(x_k + \alpha d_k)$ from the first

order condition

$$\begin{aligned}
0 &= \phi'(\alpha_k) \\
&= \nabla f(x_k + \alpha_k d_k)^T d_k \\
&= \nabla f(x_k)^T d_k \\
&= -\nabla f(x_k)^T B_k \nabla f(x_k) \leq 0
\end{aligned}$$

we obtain $\nabla f(x_k) = 0$ since B_k is positive definite. But then $d_k = 0$ and $p_k = 0$.

In both cases we showed that $q_k = 0 \Rightarrow p_k = 0$. □

This justifies the form of the update. ◀

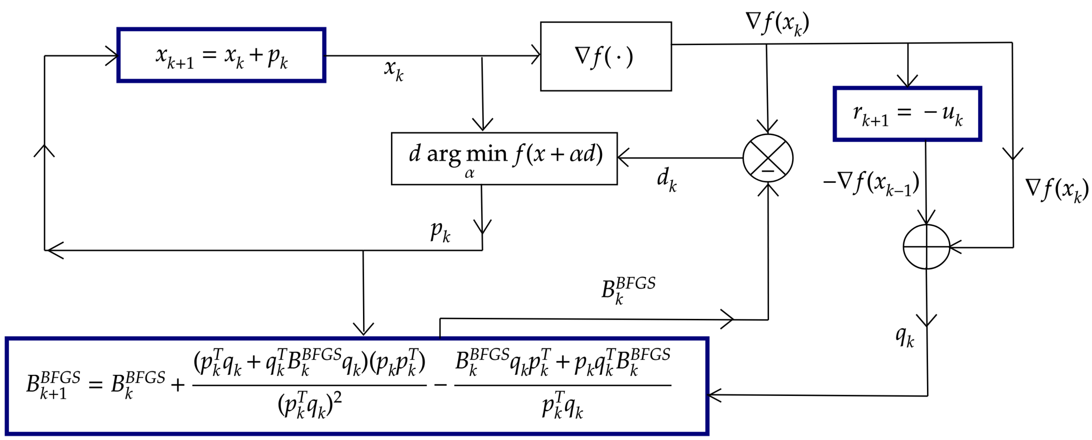

Figure 1: A diagram summarizing the BFGS algorithm.

In the following we note $r_k = \nabla f(x_{k-1})$ for each $k \geq 1$.

Lemma 0.1. *The three following statement are equivalent:*

(i) $\exists K \in \mathbb{N} \cup \{+\infty\}\ \forall m \geq K\ x_m = x^*$;

(ii) $\exists K \in \mathbb{N} \cup \{+\infty\}\ \forall m \geq K\ B_m = B^*$;

(iii) $\exists K \in \mathbb{N} \cup \{+\infty\}\ \forall m \geq K\ r_m = r^*$;

Proof. First suppose $K < +\infty$.

$(i) \Rightarrow (ii)$ If (i) holds then $\forall m \geq K\ p_m = 0$ and that means $B_m = B_K = B^*$ because of the update formula (1).

$(i) \Rightarrow (iii)$ If (i) holds, since $\forall m \geq K$ $p_m = 0$ then $q_m = 0$, i.e. $\nabla f(x_m) = \nabla f(x^*)$ and $r_m = r^*$.

$(ii+iii) \Rightarrow (i)$ It follows from the algorithm that $d_m = -B_m \nabla f(x_m) = -B^* r^* =: d^*$. Then α_m satisfies

$$0 = \nabla f(x_m + \alpha_m d^*)^T d^* = {r^*}^T d^*$$

which implies $r^* = 0 \Rightarrow d^* = 0$ and $p_m = 0$ as before. This latter equation is equivalent to (i).

If $K = +\infty$ then the sequence x_k does not converge to x^* in finite time: for all $m \in \mathbb{N}$ the stepsize $p_m \neq 0$ which in particularly implies $q_m \neq 0$ and this is equivalent to the fact that r_k does not converge in finite time. The sequence B_k does not converge either, otherwise the secant equation $B_+ q = p$ would give a contraddiction. This proves the equivalence of the three statements in the case.

□

Remark 0.1.1. We can derive a lower bound on $\|B\|$ using a Taylor expansion on r:

$$\begin{aligned} r_{k+1} - r_k &= \nabla f(x_k) - \nabla f(x_{k-1}) \\ &= H(\tau)(x_k - x_{k-1}) \\ &= \alpha_{k-1} H(\tau) p_{k-1} \\ &= \alpha_{k-1} H(\tau) B_k q_{k-1} \\ &= \alpha_{k-1} H(\tau) B_k (r_{k+1} - r_k) \end{aligned}$$

where τ is a point between x_{k-1} and x_k. Taking the norms

$$\|r_{k+1} - r_k\| \leq |\alpha_{k-1}| \, \|H(\tau)\| \, \|B_k\| \, \|r_{k+1} - r_k\| \, .$$

Now, if $r_{k+1} \neq r_k$ we derive the lower bound on $\|B_k\|$:

$$\|B_k\| \geq \frac{1}{|\alpha_{k-1}| \, \|H(\tau)\|}$$

and if $r_{k+1} = r_k$ then the algorithm has converged (reasoning as in the proof of (Lemma 0.1)). ◀

Remark 0.1.2. Without loss of generality we can impose $\alpha \geq 0$ in the algorithm, to require p to be a descent direction. This means that $p_k^T \nabla f(x_k) < 0$ for every $k \geq 0$. In particular yields $f(x_{k+1}) \leq f(x_k)$. ◀

Remark 0.1.3. With the extra assumption $\alpha \geq 0$ and the update formula (1), we can conclude that the (BFGS Algorithm) converges to a stationary point x^*. If this happens in a finite number of steps $K < +\infty$, then, by (Lemma 0.1), the sequence of matrices B_k converges as well and their dynamics is clearly stable: the sequence is modified for K steps and after it becomes constant. ◀