Quasi-Newton methods

Basic idea (§11.1)

- Newton's method:
  - Fast convergence if we start close enough to solution.
  - Requires Hessian inverse (which may be large).
- Quasi-Newton methods: approximate the Hessian inverse using only gradient information.
- Basic quasi-Newton algorithm:
$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k \boldsymbol{H}_k \boldsymbol{g}^{(k)},$$
where $\boldsymbol{H}_k$ takes the place of the true Hessian inverse in Newton's algorithm.
- The matrix $\boldsymbol{H}_{k+1}$ is computed using $\boldsymbol{x}^{(k)}$, $\boldsymbol{x}^{(k+1)}$, $\boldsymbol{g}^{(k)}$, $\boldsymbol{g}^{(k+1)}$, and $\boldsymbol{H}_k$.
- $\boldsymbol{H}_k$ is supposed to "mimic" $\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}$.
- What properties of $\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}$ should it mimic?
- At least $\boldsymbol{H}_k$ should be symmetric.
- Another property that $\boldsymbol{H}_k$ should mimic is the "secant" property.
- To explain this property, assume that f is quadratic, with Hessian $\boldsymbol{Q}$.
- Note that $\boldsymbol{Q}$ satisfies
$$\boldsymbol{g}^{(k+1)} - \boldsymbol{g}^{(k)} = \boldsymbol{Q}(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}),$$
or
$$\boldsymbol{Q}^{-1}(\boldsymbol{g}^{(k+1)} - \boldsymbol{g}^{(k)}) = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}.$$
- Let
$$\begin{aligned} \Delta\boldsymbol{g}^{(k)} &\triangleq \boldsymbol{g}^{(k+1)} - \boldsymbol{g}^{(k)}, \\ \Delta\boldsymbol{x}^{(k)} &\triangleq \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}. \end{aligned}$$
- At any k, $\boldsymbol{Q}^{-1}$ satisfies:
$$\boldsymbol{Q}^{-1}\Delta\boldsymbol{g}^{(i)} = \Delta\boldsymbol{x}^{(i)}, \qquad 0 \le i \le k.$$
- To mimic $\boldsymbol{Q}^{-1}$, we want $\boldsymbol{H}_{k+1}$ to also satisfy
$$\boldsymbol{H}_{k+1}\Delta\boldsymbol{g}^{(i)} = \Delta\boldsymbol{x}^{(i)}, \qquad 0 \le i \le k.$$
- The above is called the quasi-Newton (or secant) condition.

Summary of quasi-Newton algorithm

- Form of algorithm:

$$\begin{aligned}\boldsymbol{d}^{(k)} &= -\boldsymbol{H}_k \boldsymbol{g}^{(k)} \\ \alpha_k &= \arg\min_{\alpha \geq 0} f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}) \\ \boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)},\end{aligned}$$

  where the matrices $\boldsymbol{H}_0, \boldsymbol{H}_1, \ldots$ are symmetric.

- In the quadratic case, the above matrices are required to satisfy

$$\boldsymbol{H}_{k+1} \Delta \boldsymbol{g}^{(i)} = \Delta \boldsymbol{x}^{(i)}, \qquad 0 \leq i \leq k.$$

- Theorem (11.1): Any quasi-Newton algorithm is a conjugate direction algorithm.

- Specifically, suppose the quasi-Newton (secant) condition holds: for $0 \leq k < n-1$,

$$\boldsymbol{H}_{k+1} \Delta \boldsymbol{g}^{(i)} = \Delta \boldsymbol{x}^{(i)}, \qquad 0 \leq i \leq k.$$

  For $0 \leq k < n-1$, if $\alpha_i \neq 0$, $0 \leq i \leq k$, then $\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(k+1)}$ are $\boldsymbol{Q}$-conjugate.

Proof of theorem:

- We use induction.

- For $k = 0$, we have

$$\begin{aligned}\boldsymbol{d}^{(1)T} \boldsymbol{Q} \boldsymbol{d}^{(0)} &= -\boldsymbol{g}^{(1)T} \boldsymbol{H}_1 \boldsymbol{Q} \boldsymbol{d}^{(0)} \\ &= -\boldsymbol{g}^{(1)T} \boldsymbol{H}_1 \frac{\boldsymbol{Q} \Delta \boldsymbol{x}^{(0)}}{\alpha_0} \\ &= -\boldsymbol{g}^{(1)T} \frac{\boldsymbol{H}_1 \Delta \boldsymbol{g}^{(0)}}{\alpha_0} \\ &= -\boldsymbol{g}^{(1)T} \frac{\Delta \boldsymbol{x}^{(0)}}{\alpha_0} \\ &= -\boldsymbol{g}^{(1)T} \boldsymbol{d}^{(0)} \\ &= 0\end{aligned}$$

  because of our choice of α_0.

- Suppose the result is true for $k-1$; i.e., $\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(k)}$ are $\boldsymbol{Q}$-conjugate.

- We now prove the result for k; i.e., that $\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(k+1)}$ are $\boldsymbol{Q}$-conjugate.

- It suffices to show that $\boldsymbol{d}^{(k+1)T}\boldsymbol{Q}\boldsymbol{d}^{(i)} = 0$, $0 \leq i \leq k$.

- Given i, $0 \leq i \leq k$, we have

$$
\begin{aligned}
\boldsymbol{d}^{(k+1)T}\boldsymbol{Q}\boldsymbol{d}^{(i)} &= -\boldsymbol{g}^{(k+1)T}\boldsymbol{H}_{k+1}\boldsymbol{Q}\boldsymbol{d}^{(i)} \\
&= -\boldsymbol{g}^{(k+1)T}\boldsymbol{H}_{k+1}\frac{\boldsymbol{Q}\Delta\boldsymbol{x}^{(i)}}{\alpha_i} \\
&= -\boldsymbol{g}^{(k+1)T}\frac{\boldsymbol{H}_{k+1}\Delta\boldsymbol{g}^{(i)}}{\alpha_i} \\
&= -\boldsymbol{g}^{(k+1)T}\frac{\Delta\boldsymbol{x}^{(i)}}{\alpha_i} \\
&= -\boldsymbol{g}^{(k+1)T}\boldsymbol{d}^{(i)}.
\end{aligned}
$$

- Since $\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(k)}$ are $\boldsymbol{Q}$-conjugate by assumption, by the "expanding subspace" lemma, we have $\boldsymbol{g}^{(k+1)T}\boldsymbol{d}^{(i)} = 0$.

- By the previous theorem, we conclude that if we apply a quasi-Newton algorithm to a quadratic, it terminates in at most n steps.

- How do we generate the matrices $\boldsymbol{H}_k$ in such a way that it satisfies the quasi-Newton condition?

- There are several update formulas available for computing $\boldsymbol{H}_{k+1}$ based on $\boldsymbol{H}_k$, $\Delta\boldsymbol{g}^{(k)}$, and $\Delta\boldsymbol{x}^{(k)}$.

- Methods for generating the $\boldsymbol{H}_k$:
  - Rank one formula
  - DFP formula
  - BFGS formula

- All have the form:

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \boldsymbol{U}_k$$

  where $\boldsymbol{U}_k$ is an update (correction) term that depends on $\boldsymbol{H}_k$, $\Delta\boldsymbol{g}^{(k)}$, and $\Delta\boldsymbol{x}^{(k)}$.

Descent Property

- We want the descent property to hold.

- Recall that to have the descent property, the search direction $\boldsymbol{d}^{(k)} = -\boldsymbol{H}_k\boldsymbol{g}^{(k)}$ must have positive inner product with $-\boldsymbol{g}^{(k)}$:

$$\boldsymbol{g}^{(k)T}\boldsymbol{H}_k\boldsymbol{g}^{(k)} > 0.$$

- Prop. (11.1): If $\boldsymbol{H}_k > 0$, then the algorithm has the descent property.

Rank one correction formula (§11.3)

- The rank one formula has the form

$$\boldsymbol{U}_k = a_k \boldsymbol{z}^{(k)} \boldsymbol{z}^{(k)T},$$

  where $a_k \in \mathbb{R}$ and $\boldsymbol{z}^{(k)} \in \mathbb{R}^n$.

- Note that

$$\operatorname{rank} \boldsymbol{z}^{(k)} \boldsymbol{z}^{(k)T} = \operatorname{rank} \left(\begin{bmatrix} z_1^{(k)} \\ \vdots \\ z_n^{(k)} \end{bmatrix} \begin{bmatrix} z_1^{(k)} & \cdots & z_n^{(k)} \end{bmatrix} \right) = 1.$$

  Hence the name *rank one* correction.

- Note: if we start with a symmetric matrix $\boldsymbol{H}_0$, then the $\boldsymbol{H}_k$ remain symmetric.

- What should a_k and $\boldsymbol{z}^{(k)}$ be? We need the quasi-Newton condition to hold.

- Answer: The quasi-Newton condition holds if and only if

$$\boldsymbol{U}_k = \frac{(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)})(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)})^T}{(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)})^T \Delta \boldsymbol{g}^{(k)}},$$

  which can be expressed as:

$$\begin{aligned} a_k &= \frac{1}{(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)})^T \Delta \boldsymbol{g}^{(k)}}, \\ \boldsymbol{z}^{(k)} &= \Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}. \end{aligned}$$

- Name: Rank one update formula.

- Derivation: tedious, but straightforward.

- Note that for each k,

$$\begin{aligned} \boldsymbol{H}_{k+1} \Delta \boldsymbol{g}^{(k)} &= \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)} + \boldsymbol{U}_k \Delta \boldsymbol{g}^{(k)} \\ &= \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)} + \Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)} \\ &= \Delta \boldsymbol{x}^{(k)}. \end{aligned}$$

- What about $\boldsymbol{H}_{k+1} \Delta \boldsymbol{g}^{(i)}$ for all $i = 0, \ldots, k$?

Theorem (11.2): The rank one formula satisfies the quasi-Newton condition.

Proof:

- Need to show that for each k,

$$\boldsymbol{H}_{k+1}\Delta\boldsymbol{g}^{(i)} = \Delta\boldsymbol{x}^{(i)}, \qquad 0 \le i \le k.$$

  Use induction.

- For $k = 0$, we know it is true (because we have already seen that $\boldsymbol{H}_{k+1}\Delta\boldsymbol{g}^{(k)} = \Delta\boldsymbol{x}^{(k)}$ for each k).

- Assume true for $k-1$; i.e., that $\boldsymbol{H}_k\Delta\boldsymbol{g}^{(i)} = \Delta\boldsymbol{x}^{(i)}$, $i < k$.

- We now show it is true for k.

- Since we already know that $\boldsymbol{H}_{k+1}\Delta\boldsymbol{g}^{(k)} = \Delta\boldsymbol{x}^{(k)}$, it remains to show that $\boldsymbol{H}_{k+1}\Delta\boldsymbol{g}^{(i)} = \Delta\boldsymbol{x}^{(i)}$ for $i < k$.

- Fix $i < k$. We have

$$\boldsymbol{H}_{k+1}\Delta\boldsymbol{g}^{(i)} = \boldsymbol{H}_k\Delta\boldsymbol{g}^{(i)} + \boldsymbol{U}_k\Delta\boldsymbol{g}^{(i)}.$$

- By the induction hypothesis, $\boldsymbol{H}_k\Delta\boldsymbol{g}^{(i)} = \Delta\boldsymbol{x}^{(i)}$.

- Hence, enough to show that the $\boldsymbol{U}_k\Delta\boldsymbol{g}^{(i)} = 0$. For this, it is enough that

$$\begin{aligned}&(\Delta\boldsymbol{x}^{(k)} - \boldsymbol{H}_k\Delta\boldsymbol{g}^{(k)})^T\Delta\boldsymbol{g}^{(i)}\\ &\quad = \Delta\boldsymbol{x}^{(k)T}\Delta\boldsymbol{g}^{(i)} - \Delta\boldsymbol{g}^{(k)T}\boldsymbol{H}_k\Delta\boldsymbol{g}^{(i)} = 0.\end{aligned}$$

- We have

$$\begin{aligned}\Delta\boldsymbol{g}^{(k)T}\boldsymbol{H}_k\Delta\boldsymbol{g}^{(i)} &= \Delta\boldsymbol{g}^{(k)T}(\boldsymbol{H}_k\Delta\boldsymbol{g}^{(i)})\\ &= \Delta\boldsymbol{g}^{(k)T}\Delta\boldsymbol{x}^{(i)}\end{aligned}$$

  by the induction hypothesis.

- Since $\Delta\boldsymbol{g}^{(k)} = \boldsymbol{Q}\Delta\boldsymbol{x}^{(k)}$, we have

$$\Delta\boldsymbol{g}^{(k)T}\boldsymbol{H}_k\Delta\boldsymbol{g}^{(i)} = \Delta\boldsymbol{x}^{(k)T}\boldsymbol{Q}\Delta\boldsymbol{x}^{(i)} = \Delta\boldsymbol{x}^{(k)T}\Delta\boldsymbol{g}^{(i)}.$$

- Hence,

$$\begin{aligned}&(\Delta\boldsymbol{x}^{(k)} - \boldsymbol{H}_k\Delta\boldsymbol{g}^{(k)})^T\Delta\boldsymbol{g}^{(i)}\\ &\quad = \Delta\boldsymbol{x}^{(k)T}\Delta\boldsymbol{g}^{(i)} - \Delta\boldsymbol{x}^{(k)T}\Delta\boldsymbol{g}^{(i)} = 0,\end{aligned}$$

  which completes the proof.

Drawbacks of rank one formula

- The $\boldsymbol{H}_k$ may not be positive definite (because a_k may be negative), and hence $\boldsymbol{d}^{(k)} = -\boldsymbol{H}_k \boldsymbol{g}^{(k)}$ may not be a descent direction.
- There may be numerical problems if

$$\Delta \boldsymbol{g}^{(k)T}(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}) \approx 0.$$

- We seek more sophisticated update formulas that avoid the above problems.
- We study two other formulas: DFP and BFGS.

The DFP Algorithm (§11.4)

- DFP update formula:

$$\boldsymbol{U}_k = \frac{\Delta \boldsymbol{x}^{(k)} \Delta \boldsymbol{x}^{(k)T}}{\Delta \boldsymbol{x}^{(k)T} \Delta \boldsymbol{g}^{(k)}} - \frac{\boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)} \Delta \boldsymbol{g}^{(k)T} \boldsymbol{H}_k}{\Delta \boldsymbol{g}^{(k)T} \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}}.$$

- Davidon, 1959; Fletcher and Powell, 1963.
- Also called *variable metric algorithm*.
- Has two "rank one" terms.

Theorem (11.3): The DFP algorithm satisfies the quasi-Newton condition.

Proof:

- Need to show $\boldsymbol{H}_{k+1} \Delta \boldsymbol{g}^{(i)} = \Delta \boldsymbol{x}^{(i)}$, $0 \leq i \leq k$.
- For $i = k$:

$$\begin{aligned} \boldsymbol{H}_{k+1} \Delta \boldsymbol{g}^{(k)} &= \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)} + \boldsymbol{U}_k \Delta \boldsymbol{g}^{(k)} \\ &= \Delta \boldsymbol{x}^{(k)}. \end{aligned}$$

- For general case, use induction.
- For $k = 0$, already showed it is true.
- Assume true for $k - 1$: $\boldsymbol{H}_k \Delta \boldsymbol{g}^{(i)} = \Delta \boldsymbol{x}^{(i)}$, $0 \leq i \leq k - 1$.
- To show true for k, remains to consider the case $i < k$.

- We have

$$\begin{aligned}\boldsymbol{H}_{k+1}\Delta\boldsymbol{g}^{(i)} &= \Delta\boldsymbol{x}^{(i)} \\ &\quad + \frac{\Delta\boldsymbol{x}^{(k)}}{\Delta\boldsymbol{x}^{(k)T}\Delta\boldsymbol{g}^{(k)}}(\Delta\boldsymbol{x}^{(k)T}\Delta\boldsymbol{g}^{(i)}) \\ &\quad - \frac{\boldsymbol{H}_k\Delta\boldsymbol{g}^{(k)}}{\Delta\boldsymbol{g}^{(k)T}\boldsymbol{H}_k\Delta\boldsymbol{g}^{(k)}}(\Delta\boldsymbol{g}^{(k)T}\Delta\boldsymbol{x}^{(i)}).\end{aligned}$$

- Now,

$$\begin{aligned}\Delta\boldsymbol{x}^{(k)T}\Delta\boldsymbol{g}^{(i)} &= \Delta\boldsymbol{x}^{(k)T}\boldsymbol{Q}\Delta\boldsymbol{x}^{(i)} \\ &= \alpha_k\alpha_i\boldsymbol{d}^{(k)T}\boldsymbol{Q}\boldsymbol{d}^{(i)} \\ &= 0\end{aligned}$$

  by the induction hypothesis and the conjugate direction property.

- Similarly, $\Delta\boldsymbol{g}^{(k)T}\Delta\boldsymbol{x}^{(i)} = 0$.

- Hence,

$$\boldsymbol{H}_{k+1}\Delta\boldsymbol{g}^{(i)} = \Delta\boldsymbol{x}^{(i)},$$

  and the proof is completed.

- Theorem (11.4): Suppose $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$. In the DFP algorithm, if $\boldsymbol{H}_k$ is positive definite, then so is $\boldsymbol{H}_{k+1}$.

- Proof: Tedious but straightforward.

- DFP algorithm better than rank one algorithm.

- DFP algorithm may have problems in some cases (getting stuck).

The BFGS Algorithm (§11.5)

- BFGS update algorithm:

$$\begin{aligned}\boldsymbol{U}_k &= \left(1 + \frac{\Delta\boldsymbol{g}^{(k)T}\boldsymbol{H}_k\Delta\boldsymbol{g}^{(k)}}{\Delta\boldsymbol{g}^{(k)T}\Delta\boldsymbol{x}^{(k)}}\right)\frac{\Delta\boldsymbol{x}^{(k)}\Delta\boldsymbol{x}^{(k)T}}{\Delta\boldsymbol{x}^{(k)T}\Delta\boldsymbol{g}^{(k)}} \\ &\quad - \frac{\boldsymbol{H}_k\Delta\boldsymbol{g}^{(k)}\Delta\boldsymbol{x}^{(k)T} + (\boldsymbol{H}_k\Delta\boldsymbol{g}^{(k)}\Delta\boldsymbol{x}^{(k)T})^T}{\Delta\boldsymbol{g}^{(k)T}\Delta\boldsymbol{x}^{(k)}}.\end{aligned}$$

- Broyden, Fletcher, Goldfarb, and Shanno, 1970.

- The BFGS formula is derived from the DFP formula using a technique called *complementarity*.

- Consider the quasi-Newton condition:

$$\boldsymbol{H}_{k+1}\Delta \boldsymbol{g}^{(i)} = \Delta \boldsymbol{x}^{(i)}, \qquad 0 \le i \le k.$$

- Consider a modified condition in which the roles of $\Delta \boldsymbol{g}^{(i)}$ and $\Delta \boldsymbol{x}^{(i)}$ are interchanged:

$$\boldsymbol{B}_{k+1}\Delta \boldsymbol{x}^{(i)} = \Delta \boldsymbol{g}^{(i)}, \qquad 0 \le i \le k.$$

  Call the above the "complementary quasi-Newton" condition.

- Think of $\boldsymbol{B}_k$ as an approximation to the Hessian (instead of inverse Hessian).

- Given: an update equation for $\boldsymbol{H}_k$ that satisfies the quasi-Newton condition.

- If we interchange $\Delta \boldsymbol{x}^{(k)}$ and $\Delta \boldsymbol{g}^{(k)}$ in the equation, and replace $\boldsymbol{H}_k$ by $\boldsymbol{B}_k$, then the resulting formula satisfies the complementary quasi-Newton condition.

- Based on the DFP formula,

$$\boldsymbol{B}_{k+1} = \boldsymbol{B}_k + \frac{\Delta \boldsymbol{g}^{(k)}\Delta \boldsymbol{g}^{(k)T}}{\Delta \boldsymbol{g}^{(k)T}\Delta \boldsymbol{x}^{(k)}} - \frac{\boldsymbol{B}_k\Delta \boldsymbol{x}^{(k)}\Delta \boldsymbol{x}^{(k)T}\boldsymbol{B}_k}{\Delta \boldsymbol{x}^{(k)T}\boldsymbol{B}_k\Delta \boldsymbol{x}^{(k)}}$$

- The above formula satisfies the complementary quasi-Newton condition.

- The previous formula for updating $\boldsymbol{B}_k$ is not immediately useful because what we need is the inverse Hessian.

- What we need is an update formula for $\boldsymbol{B}_k^{-1}$.

- The previous formula is of the form:

$$\boldsymbol{B}_{k+1} = \boldsymbol{B}_k + \boldsymbol{u}_1\boldsymbol{v}_1^T + \boldsymbol{u}_2\boldsymbol{v}_2^T.$$

- Hence,

$$\boldsymbol{B}_{k+1}^{-1} = \left(\boldsymbol{B}_k + \boldsymbol{u}_1\boldsymbol{v}_1^T + \boldsymbol{u}_2\boldsymbol{v}_2^T\right)^{-1}.$$

- Lemma (11.1): Let $\boldsymbol{A}$ be a nonsingular matrix. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be column vectors and assume that $1 + \boldsymbol{v}^T\boldsymbol{A}^{-1}\boldsymbol{u} \ne 0$. Then, $\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^T$ is nonsingular, and

$$(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^T)^{-1} = \boldsymbol{A}^{-1} - \frac{(\boldsymbol{A}^{-1}\boldsymbol{u})(\boldsymbol{v}^T\boldsymbol{A}^{-1})}{1 + \boldsymbol{v}^T\boldsymbol{A}^{-1}\boldsymbol{u}}.$$

- Proof: by verification.

- Name: *Sherman-Morrison formula*. Very useful!
- Note form:

$$(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^T)^{-1} = \boldsymbol{A}^{-1} + \boldsymbol{x}\boldsymbol{y}^T.$$

- Apply the Sherman-Morrison formula twice to

$$\boldsymbol{B}_{k+1}^{-1} = \left(\boldsymbol{B}_k + \boldsymbol{u}_1\boldsymbol{v}_1^T + \boldsymbol{u}_2\boldsymbol{v}_2^T\right)^{-1}.$$

- We obtain an update formula of the form:

$$\boldsymbol{B}_{k+1}^{-1} = \boldsymbol{B}_k^{-1} + \boldsymbol{u}_3\boldsymbol{v}_3^T + \boldsymbol{u}_4\boldsymbol{v}_4^T.$$

- If we now replace $\boldsymbol{B}_k^{-1}$ by the symbol $\boldsymbol{H}_k$, we obtain the BFGS formula!
- BFGS is the "complementary" formula to DFP.
- By the nature of complementarity, the BFGS formula inherits the properties of DFP.
- Theorem: The BFGS formula satisfies the quasi-Newton condition.
- Theorem: Suppose $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$. In the BFGS algorithm, if $\boldsymbol{H}_k$ is positive definite, then so is $\boldsymbol{H}_{k+1}$.