ECE 580

Optimization Methods for Systems and Control

Chapter 11

Quasi-Newton Methods

Stan Żak

February 25, 2013

Quasi-Newton Algorithms

-

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

Quasi-Newton Algorithms

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

- The matrices $\boldsymbol{H}_k = \boldsymbol{H}_k^\top > 0$ satisfy

$$\boldsymbol{H}_{k+1} \Delta \boldsymbol{g}^{(i)} = \Delta \boldsymbol{x}^{(i)}, \qquad i = 0, 1, 2 \ldots, k$$

Quasi-Newton Algorithms

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

- The matrices $\boldsymbol{H}_k = \boldsymbol{H}_k^\top > 0$ satisfy

$$
\boldsymbol{H}_{k+1} \Delta \boldsymbol{g}^{(i)} = \Delta \boldsymbol{x}^{(i)}, \qquad i = 0, 1, 2 \ldots, k
$$

- Question,

$$
\boldsymbol{H}_k = ?
$$

The Rank-One Correction Formula

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

The Rank-One Correction Formula

- $\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \Delta \boldsymbol{H}_k$
- Want

$$\Delta \boldsymbol{H}_k = \Delta \boldsymbol{H}_k^\top$$

The Rank-One Correction Formula

- $\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \Delta \boldsymbol{H}_k$
- Want

$$\Delta \boldsymbol{H}_k = \Delta \boldsymbol{H}_k^\top$$

- Try the dyadic or outer product,

$$\Delta \boldsymbol{H}_k = a_k \begin{bmatrix} \\ \boldsymbol{z}^{(k)} \\ \\ \end{bmatrix} \begin{bmatrix} & \boldsymbol{z}^{(k)\top} & \end{bmatrix}$$

The Rank-One Correction Formula

- $\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \Delta \boldsymbol{H}_k$
- Want

$$\Delta \boldsymbol{H}_k = \Delta \boldsymbol{H}_k^\top$$

- Try the dyadic or outer product,

$$\Delta \boldsymbol{H}_k = a_k \begin{bmatrix} \\ \boldsymbol{z}^{(k)} \\ \\ \end{bmatrix} \begin{bmatrix} & \boldsymbol{z}^{(k)\top} & \end{bmatrix}$$

- Note that

$$\Delta \boldsymbol{H}_k^\top = a_k \left(\boldsymbol{z}^{(k)\top} \right)^\top \boldsymbol{z}^{(k)\top} = a_k \boldsymbol{z}^{(k)} \boldsymbol{z}^{(k)\top} = \Delta \boldsymbol{H}_k$$

Problem Statement for Finding the Rank-One Update

- We have

$$\Delta \boldsymbol{H}_k = a_k \boldsymbol{z}^{(k)} \boldsymbol{z}^{(k)\top}$$

Problem Statement for Finding the Rank-One Update

- We have

$$\Delta \boldsymbol{H}_k = a_k \boldsymbol{z}^{(k)} \boldsymbol{z}^{(k)\top}$$

- Given: $\boldsymbol{H}_k$, $\Delta \boldsymbol{g}^{(k)}$, $\Delta \boldsymbol{x}^{(k)}$

Problem Statement for Finding the Rank-One Update

- We have

$$\Delta \boldsymbol{H}_k = a_k \boldsymbol{z}^{(k)} \boldsymbol{z}^{(k)\top}$$

- Given: $\boldsymbol{H}_k$, $\Delta \boldsymbol{g}^{(k)}$, $\Delta \boldsymbol{x}^{(k)}$
- What must a_k and $\boldsymbol{z}^{(k)}$ be to force

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

Problem Statement for Finding the Rank-One Update

- We have
$$\Delta \boldsymbol{H}_k = a_k \boldsymbol{z}^{(k)} \boldsymbol{z}^{(k)\top}$$
- Given: $\boldsymbol{H}_k$, $\Delta \boldsymbol{g}^{(k)}$, $\Delta \boldsymbol{x}^{(k)}$
- What must a_k and $\boldsymbol{z}^{(k)}$ be to force
$$\boldsymbol{H}_{k+1} \Delta \boldsymbol{g}^{(k)} = \Delta \boldsymbol{x}^{(k)}?$$
- That is, what must a_k and $\boldsymbol{z}^{(k)}$ be to force
$$\left(\boldsymbol{H}_k + a_k \boldsymbol{z}^{(k)} \boldsymbol{z}^{(k)\top}\right) \Delta \boldsymbol{g}^{(k)} = \Delta \boldsymbol{x}^{(k)}?$$

Manipulate

- Multiply and re-arrange

$$\left(\boldsymbol{H}_k + a_k \boldsymbol{z}^{(k)} \boldsymbol{z}^{(k)\top}\right) \Delta \boldsymbol{g}^{(k)} = \Delta \boldsymbol{x}^{(k)}$$

Manipulate

- Multiply and re-arrange

$$\left(\boldsymbol{H}_k + a_k \boldsymbol{z}^{(k)} \boldsymbol{z}^{(k)\top}\right) \Delta \boldsymbol{g}^{(k)} = \Delta \boldsymbol{x}^{(k)}$$

- We obtain

$$a_k \begin{bmatrix} \\ \boldsymbol{z}^{(k)} \\ \\ \end{bmatrix} \left(\begin{bmatrix} & \boldsymbol{z}^{(k)\top} & \end{bmatrix} \begin{bmatrix} \\ \Delta \boldsymbol{g}^{(k)} \\ \\ \end{bmatrix} \right) = \Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}$$

Manipulate

- Multiply and re-arrange

$$\left(\boldsymbol{H}_k + a_k \boldsymbol{z}^{(k)} \boldsymbol{z}^{(k)\top}\right) \Delta \boldsymbol{g}^{(k)} = \Delta \boldsymbol{x}^{(k)}$$

- We obtain

$$a_k \begin{bmatrix} \\ \boldsymbol{z}^{(k)} \\ \\ \end{bmatrix} \left(\begin{bmatrix} & \boldsymbol{z}^{(k)\top} & \end{bmatrix} \begin{bmatrix} \\ \Delta \boldsymbol{g}^{(k)} \\ \\ \end{bmatrix} \right) = \Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}$$

- Divide both sides by

$$a_k \boldsymbol{z}^{(k)\top} \Delta \boldsymbol{g}^{(k)}$$

Obtaining $z^{(k)}$

- The result of the above manipulations is

$$\left[\begin{array}{c} \\ \boldsymbol{z}^{(k)} \\ \\ \end{array}\right] = \frac{\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}}{a_k \boldsymbol{z}^{(k)\top} \Delta \boldsymbol{g}^{(k)}}$$

Obtaining $\boldsymbol{z}^{(k)}$

- The result of the above manipulations is

$$\begin{bmatrix} \\ \boldsymbol{z}^{(k)} \\ \\ \end{bmatrix} = \frac{\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}}{a_k \boldsymbol{z}^{(k)\top} \Delta \boldsymbol{g}^{(k)}}$$

- Hence,

$$\begin{aligned} \boldsymbol{H}_{k+1} &= \boldsymbol{H}_k + a_k \boldsymbol{z}^{(k)} \boldsymbol{z}^{(k)\top} \\ &= \boldsymbol{H}_k + \frac{\left(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}\right)\left(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}\right)^\top}{a_k \left(\boldsymbol{z}^{(k)\top} \Delta \boldsymbol{g}^{(k)}\right)^2} \end{aligned}$$

Obtaining $\boldsymbol{z}^{(k)}$

- The result of the above manipulations is

$$\left[\ \boldsymbol{z}^{(k)}\ \right] = \frac{\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}}{a_k \boldsymbol{z}^{(k)\top} \Delta \boldsymbol{g}^{(k)}}$$

- Hence,

$$\begin{aligned} \boldsymbol{H}_{k+1} &= \boldsymbol{H}_k + a_k \boldsymbol{z}^{(k)} \boldsymbol{z}^{(k)\top} \\ &= \boldsymbol{H}_k + \frac{\left(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}\right)\left(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}\right)^\top}{a_k \left(\boldsymbol{z}^{(k)\top} \Delta \boldsymbol{g}^{(k)}\right)^2} \end{aligned}$$

- We now wish to express a_k in terms of the given data

Obtaining a_k

- To find a_k, consider

$$\Delta \boldsymbol{g}^{(k)\top} \boldsymbol{z}^{(k)} = \frac{\Delta \boldsymbol{g}^{(k)\top} \left(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)} \right)}{a_k \left(\boldsymbol{z}^{(k)\top} \Delta \boldsymbol{g}^{(k)} \right)}$$

Obtaining a_k

- To find a_k, consider

$$\Delta \boldsymbol{g}^{(k)\top} \boldsymbol{z}^{(k)} = \frac{\Delta \boldsymbol{g}^{(k)\top} \left(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)} \right)}{a_k \left(\boldsymbol{z}^{(k)\top} \Delta \boldsymbol{g}^{(k)} \right)}$$

- Note that

$$\boldsymbol{z}^{(k)\top} \Delta \boldsymbol{g}^{(k)} = \Delta \boldsymbol{g}^{(k)\top} \boldsymbol{z}^{(k)}$$

Obtaining a_k

- To find a_k, consider

$$\Delta \boldsymbol{g}^{(k)\top} \boldsymbol{z}^{(k)} = \frac{\Delta \boldsymbol{g}^{(k)\top} \left(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)} \right)}{a_k \left(\boldsymbol{z}^{(k)\top} \Delta \boldsymbol{g}^{(k)} \right)}$$

- Note that

$$\boldsymbol{z}^{(k)\top} \Delta \boldsymbol{g}^{(k)} = \Delta \boldsymbol{g}^{(k)\top} \boldsymbol{z}^{(k)}$$

- Hence,

$$a_k \left(\boldsymbol{z}^{(k)\top} \Delta \boldsymbol{g}^{(k)} \right)^2 = \Delta \boldsymbol{g}^{(k)\top} \left(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)} \right)$$

Finally, the Rank-One Update Formula!

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

Finally, the Rank-One Update Formula!

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

- One can show (Thm 11.2) that the above update formula applied to the quadratic gives

$$\boldsymbol{H}_{k+1} \Delta \boldsymbol{g}^{(i)} = \Delta \boldsymbol{x}^{(i)}, \qquad i = 0, 1, 2 \ldots, k$$

Rank-One Algorithm

- Step 1. Select $\boldsymbol{x}^{(0)}$ and $\boldsymbol{H}_0 = \boldsymbol{H}_0^\top > 0$. Calculate $\boldsymbol{g}^{(0)}$ and set $k = 0$

Rank-One Algorithm

- Step 1. Select $\boldsymbol{x}^{(0)}$ and $\boldsymbol{H}_0 = \boldsymbol{H}_0^\top > 0$. Calculate $\boldsymbol{g}^{(0)}$ and set $k = 0$
- Step 2. Calculate (Newton-like direction)

$$\boldsymbol{d}^{(k)} = -\boldsymbol{H}_k \boldsymbol{g}^{(k)}$$

Rank-One Algorithm

- Step 1. Select $\boldsymbol{x}^{(0)}$ and $\boldsymbol{H}_0 = \boldsymbol{H}_0^\top > 0$. Calculate $\boldsymbol{g}^{(0)}$ and set $k = 0$
- Step 2. Calculate (Newton-like direction)

$$\boldsymbol{d}^{(k)} = -\boldsymbol{H}_k \boldsymbol{g}^{(k)}$$

- Step 3. Perform line search

$$\alpha_k = \arg\min_{\alpha} f\left(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}\right)$$

and compute

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}$$

Rank-One Algorithm—Contd.

- Step 4. Calculate

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

Rank-One Algorithm—Contd.

- Step 4. Calculate

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

- Step 5. Set

$$k := k + 1$$

and go to Step 2

Rank-One Algorithm—Contd.

- Step 4. Calculate

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

- Step 5. Set

$$k := k + 1$$

  and go to Step 2

- In the Rank-One Correction algo, $\boldsymbol{H}_{k+1}$ may not be p.d.! Thus $\boldsymbol{d}^{(k+1)}$ may not be a descent direction.

Rank-Two Update

- There is a fix to the problem of $\boldsymbol{H}_{k+1}$ sometimes not being p.d. in the rank-one algo—a rank-two update!

Rank-Two Update

- There is a fix to the problem of $\boldsymbol{H}_{k+1}$ sometimes not being p.d. in the rank-one algo—a rank-two update!
- The Davidon-Fletcher-Powell (DFP) algo also called the variable metric algo

Rank-Two Update

- There is a fix to the problem of $\boldsymbol{H}_{k+1}$ sometimes not being p.d. in the rank-one algo—a rank-two update!
- The Davidon-Fletcher-Powell (DFP) algo also called the variable metric algo
- In the DFP algo, $\boldsymbol{H}_{k+1}$ inherits positive definiteness from $\boldsymbol{H}_k$—Thm 11.4

Rank-Two Update

- There is a fix to the problem of $\boldsymbol{H}_{k+1}$ sometimes not being p.d. in the rank-one algo—a rank-two update!
- The Davidon-Fletcher-Powell (DFP) algo also called the variable metric algo
- In the DFP algo, $\boldsymbol{H}_{k+1}$ inherits positive definiteness from $\boldsymbol{H}_k$—Thm 11.4
- Unfortunately, for “larger” non-quadratic problems, the DFP algo has the tendency of sometimes getting “stuck”

Rank-Two Update

- There is a fix to the problem of $\boldsymbol{H}_{k+1}$ sometimes not being p.d. in the rank-one algo—a rank-two update!
- The Davidon-Fletcher-Powell (DFP) algo also called the variable metric algo
- In the DFP algo, $\boldsymbol{H}_{k+1}$ inherits positive definiteness from $\boldsymbol{H}_k$—Thm 11.4
- Unfortunately, for "larger" non-quadratic problems, the DFP algo has the tendency of sometimes getting "stuck"
- There is a fix—the Broyden, Fletcher, Goldfarb, Shanno (BFGS) algo

Rank-Two Update

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

Deriving the BFGS Update

- Use the concept of duality or complementarity

Deriving the BFGS Update

- Use the concept of duality or complementarity
- Recall the updating formula for the approximation of the inverse of the Hessian,

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

Deriving the BFGS Update

- Use the concept of duality or complementarity
- Recall the updating formula for the approximation of the inverse of the Hessian,

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

- The above was inspired by

$$\boldsymbol{Q}^{-1}\Delta\boldsymbol{g}^{(i)} = \Delta\boldsymbol{x}^{(i)}$$

Deriving the BFGS Update

- Use the concept of duality or complementarity
- Recall the updating formula for the approximation of the inverse of the Hessian,

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

- The above was inspired by

$$\boldsymbol{Q}^{-1}\Delta\boldsymbol{g}^{(i)} = \Delta\boldsymbol{x}^{(i)}$$

- Suppose now we wish to approximate $\boldsymbol{Q}$ rather than $\boldsymbol{Q}^{-1}$

Deriving the BFGS Update

- Use the concept of duality or complementarity
- Recall the updating formula for the approximation of the inverse of the Hessian,

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

- The above was inspired by

$$\boldsymbol{Q}^{-1}\Delta\boldsymbol{g}^{(i)} = \Delta\boldsymbol{x}^{(i)}$$

- Suppose now we wish to approximate $\boldsymbol{Q}$ rather than $\boldsymbol{Q}^{-1}$
- Let $\boldsymbol{B}_k$ be our estimate of $\boldsymbol{Q}$ at the k-th step, where $\Delta\boldsymbol{g}^{(i)} = \boldsymbol{Q}\Delta\boldsymbol{x}^{(i)}$

Approximating the Hessian

- We require

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

Approximating the Hessian

- We require

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

- We have for the quadratic case

$$\Delta \boldsymbol{g}^{(i)} = \boldsymbol{Q} \Delta \boldsymbol{x}^{(i)}$$

Approximating the Hessian

- We require

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

- We have for the quadratic case

$$\Delta \boldsymbol{g}^{(i)} = \boldsymbol{Q} \Delta \boldsymbol{x}^{(i)}$$

- Compare the above with the DFP update

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

Approximating the Hessian

- We require

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

- We have for the quadratic case

$$\Delta \boldsymbol{g}^{(i)} = \boldsymbol{Q} \Delta \boldsymbol{x}^{(i)}$$

- Compare the above with the DFP update

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

- We construct the formula for updating the Hessian by interchanging the roles of $\boldsymbol{B}_k$ and $\boldsymbol{H}_k$ and $\Delta \boldsymbol{g}^{(i)}$ and $\Delta \boldsymbol{x}^{(i)}$—the complementarity (duality) concept

Approximating the Hessian Using the Duality Concept

- The DFP formula for updating the inverse Hessian

$$\boldsymbol{H}_{k+1}^{DFP} = \boldsymbol{H}_k + \frac{\Delta \boldsymbol{x}^{(k)} \Delta \boldsymbol{x}^{(k)\top}}{\Delta \boldsymbol{x}^{(k)\top} \Delta \boldsymbol{g}^{(k)}} - \frac{\boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)} \Delta \boldsymbol{g}^{(k)\top} \boldsymbol{H}_k}{\Delta \boldsymbol{g}^{(k)\top} \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}}$$

Approximating the Hessian Using the Duality Concept

- The DFP formula for updating the inverse Hessian

$$\boldsymbol{H}_{k+1}^{DFP} = \boldsymbol{H}_k + \frac{\Delta\boldsymbol{x}^{(k)}\Delta\boldsymbol{x}^{(k)\top}}{\Delta\boldsymbol{x}^{(k)\top}\Delta\boldsymbol{g}^{(k)}} - \frac{\boldsymbol{H}_k\Delta\boldsymbol{g}^{(k)}\Delta\boldsymbol{g}^{(k)\top}\boldsymbol{H}_k}{\Delta\boldsymbol{g}^{(k)\top}\boldsymbol{H}_k\Delta\boldsymbol{g}^{(k)}}$$

- The formula for updating the Hessian— interchange the roles of $\boldsymbol{B}_k$ and $\boldsymbol{H}_k$ and $\Delta\boldsymbol{g}^{(i)}$ and $\Delta\boldsymbol{x}^{(i)}$

Approximating the Hessian Using the Duality Concept

- The DFP formula for updating the inverse Hessian

$$\boldsymbol{H}_{k+1}^{DFP} = \boldsymbol{H}_k + \frac{\Delta \boldsymbol{x}^{(k)} \Delta \boldsymbol{x}^{(k)\top}}{\Delta \boldsymbol{x}^{(k)\top} \Delta \boldsymbol{g}^{(k)}} - \frac{\boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)} \Delta \boldsymbol{g}^{(k)\top} \boldsymbol{H}_k}{\Delta \boldsymbol{g}^{(k)\top} \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}}$$

- The formula for updating the Hessian— interchange the roles of $\boldsymbol{B}_k$ and $\boldsymbol{H}_k$ and $\Delta \boldsymbol{g}^{(i)}$ and $\Delta \boldsymbol{x}^{(i)}$

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

The BFGS Formula for Approximating the Inverse Hessian

- Use the Sherman-Morrison formula for a matrix inverse

$$\left(\boldsymbol{A} + \begin{bmatrix} \boldsymbol{u} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}^\top \end{bmatrix}\right)^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^\top\boldsymbol{A}^{-1}}{1 + \boldsymbol{v}^\top\boldsymbol{A}^{-1}\boldsymbol{u}}$$

The BFGS Formula for Approximating the Inverse Hessian

- Use the Sherman-Morrison formula for a matrix inverse

$$\left(\boldsymbol{A} + \begin{bmatrix} \boldsymbol{u} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}^\top \end{bmatrix} \right)^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^\top\boldsymbol{A}^{-1}}{1+\boldsymbol{v}^\top\boldsymbol{A}^{-1}\boldsymbol{u}}$$

- Apply the above formula twice to compute $\boldsymbol{B}_{k+1}^{-1}$, that is,

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