Chong, Edwin Kah Pin., and Stanislaw H. Żak. *An Introduction to Optimization*. Fourth edition., John Wiley & Sons, Inc., 2013.

6 QUASI-NEWTON METHODS

6.1 INTRODUCTION

- A computational drawback of **Newton's method** is the need to evaluate $\boldsymbol{F}(\boldsymbol{x}^{(k)})$ and solve the equation $\boldsymbol{F}(\boldsymbol{x}^{(k)})\boldsymbol{d}^{(k)} = -\boldsymbol{g}^{(k)}$ [i.e., compute $\boldsymbol{d}^{(k)} = -\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}\boldsymbol{g}^{(k)}$].

- To avoid the computation of $\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}$, the **quasi-Newton methods** use an **approximation** to $\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}$ in place of the true inverse. This approximation is updated at every stage so that it exhibits some properties of $\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}$.

- Consider the formula

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha \boldsymbol{H}_k \boldsymbol{g}^{(k)},$$

  where $\boldsymbol{H}_k$ is an $n \times n$ real matrix and α is a positive search parameter.

- Note that $\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} = -\alpha \boldsymbol{H}_k \boldsymbol{g}^{(k)}$. Expanding f about $\boldsymbol{x}^{(k)}$ yields

$$\begin{aligned} f(\boldsymbol{x}^{(k+1)}) &= f(\boldsymbol{x}^{(k)}) + \boldsymbol{g}^{(k)\top}(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) + o(\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\|) \\ &= f(\boldsymbol{x}^{(k)}) - \alpha \boldsymbol{g}^{(k)\top}\boldsymbol{H}_k \boldsymbol{g}^{(k)} + o(\alpha\|\boldsymbol{H}_k \boldsymbol{g}^{(k)}\|) \end{aligned}$$

  As α tends to zero, the second term on the right-hand side of this equation dominates the third. Thus, to guarantee a decrease in f for small α, we have to have

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

  A simple way to ensure this is to require that $\boldsymbol{H}_k$ be positive definite.

- **Proposition 6.1.1** Let $f \in C^1$, $\boldsymbol{x}^{(k)} \in \mathbb{R}^n$, $\boldsymbol{g}^{(k)} = \nabla f(\boldsymbol{x}^{(k)}) \neq \boldsymbol{0}$, and $\boldsymbol{H}_k$ an $n \times n$ real symmetric positive definite matrix. If we set $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha \boldsymbol{H}_k \boldsymbol{g}^{(k)}$, where $\alpha_k = \operatorname*{argmin}_{\alpha \geq 0} f(\boldsymbol{x}^{(k)} - \alpha \boldsymbol{H}_k \boldsymbol{g}^{(k)})$, then $\alpha_k > 0$ and $f(\boldsymbol{x}^{(k+1)}) < f(\boldsymbol{x}^{(k)})$.

6.2 APPROXIMATING THE INVERSE HESSIAN

- To begin, suppose that the objection function f is quadratic:

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}$$

- The Hessian matrix $\boldsymbol{F}(\boldsymbol{x})$ of the objective function f is constant and independent of $\boldsymbol{x}$.

- In other words, the Hessian $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{Q}$ for all $\boldsymbol{x}$, where $\boldsymbol{Q} = \boldsymbol{Q}^\top$.

- Note that $\boldsymbol{g}(\boldsymbol{x}) = \nabla \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}$. Then,

$$\boldsymbol{g}^{(0)} = \boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b}, \boldsymbol{g}^{(1)} = \boldsymbol{Q}\boldsymbol{x}^{(1)} - \boldsymbol{b}, \dots, \boldsymbol{g}^{(k)} = \boldsymbol{Q}\boldsymbol{x}^{(k)} - \boldsymbol{b}, \boldsymbol{g}^{(k+1)} = \boldsymbol{Q}\boldsymbol{x}^{(k+1)} - \boldsymbol{b},$$

and

$$\begin{aligned} \boldsymbol{g}^{(i+1)} - \boldsymbol{g}^{(i)} &= \boldsymbol{Q}\left(\boldsymbol{x}^{(i+1)} - \boldsymbol{x}^{(i)}\right) \\ \Delta\boldsymbol{g}^{(i)} &= \boldsymbol{Q}\Delta\boldsymbol{x}^{(i)} \\ \boldsymbol{Q}^{-1}\Delta\boldsymbol{g}^{(i)} &= \Delta\boldsymbol{x}^{(i)}, \end{aligned}$$

where $\Delta\boldsymbol{g}^{(i)} \triangleq \boldsymbol{g}^{(i+1)} - \boldsymbol{g}^{(i)}$ and $\Delta\boldsymbol{x}^{(i)} = \boldsymbol{x}^{(i+1)} - \boldsymbol{x}^{(i)}$, $0 \leq i \leq k$.

- Therefore, we impose the requirement that the approximation $\boldsymbol{H}_{k+1}$ of the Hessian satisfy

  1. $\boldsymbol{H}_{k+1} = \boldsymbol{H}_{k+1}^\top$
  2. $\boldsymbol{H}_{k+1} \succ 0$,
  3. $\boldsymbol{H}_{k+1}\Delta\boldsymbol{g}^{(i)} = \Delta\boldsymbol{x}^{(i)}$ for $0 \leq i \leq k$.

6.3 THE RANK ONE CORRECTION FORMULA

- In the rank one correction formula, the correction term is symmetric and has the form $a_k\boldsymbol{z}^{(k)}\boldsymbol{z}^{(k)\top}$, where $a_k \in \mathbb{R}$ and $\boldsymbol{z}^{(k)} \in \mathbb{R}^n$. Therefore, the update equation is

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \alpha_k\boldsymbol{z}^{(k)}\boldsymbol{z}^{(k)\top}.$$

- Note that

$$\operatorname{rank} \boldsymbol{z}^{(k)}\boldsymbol{z}^{(k)\top} = \operatorname{rank}\left(\begin{bmatrix} z_1^{(k)} \\ \vdots \\ z_n^{(k)} \end{bmatrix}\left[z_1^{(k)}, \dots, z_n^{(k)}\right]\right) = 1$$

- Recall that we require that $\boldsymbol{H}_{k+1}\Delta\boldsymbol{g}^{(i)} = \Delta\boldsymbol{x}^{(i)}$, $i = 1, \dots, k$. In other words, given $\boldsymbol{H}_k$, $\Delta\boldsymbol{g}^{(k)}$, and $\Delta\boldsymbol{x}^{(k)}$, we wish to find a_k and $\boldsymbol{z}^{(k)}$ to ensure that

$$\boldsymbol{H}_{k+1}\Delta\boldsymbol{g}^{(k)} = \left(\boldsymbol{H}_k + \alpha_k\boldsymbol{z}^{(k)}\boldsymbol{z}^{(k)\top}\right)\Delta\boldsymbol{g}^{(k)} = \Delta\boldsymbol{x}^{(k)}.$$

First note that $\boldsymbol{z}^{(k)\top}\,\Delta\boldsymbol{g}^{(k)}$ is a scalar. Thus,

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

We can now determine

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

Hence,

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \alpha_k \boldsymbol{z}^{(k)} \boldsymbol{z}^{(k)\top} = \boldsymbol{H}_k + \frac{\left(\Delta\boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta\boldsymbol{g}^{(k)}\right)\left(\Delta\boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta\boldsymbol{g}^{(k)}\right)^\top}{\alpha_k (\boldsymbol{z}^{(k)\top} \Delta\boldsymbol{g}^{(k)})^2}.$$

- The next step is to express the denominator of the second term on the righthand side of the equation above as a function of the given quantities $\boldsymbol{H}_k$, $\Delta\boldsymbol{g}^{(k)}$, and $\Delta\boldsymbol{x}^{(k)}$ only:

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

- Hence,

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

- **Rank One Algorithm**

1. Set $k \coloneqq 0$; select $\boldsymbol{x}^{(0)}$ and a real symmetric positive definite $\boldsymbol{H}_0$.
2. If $\boldsymbol{g}^{(k)} = \boldsymbol{0}$, stop; else, $\boldsymbol{d}^{(k)} = -\boldsymbol{H}_k \boldsymbol{g}^{(k)}$.
3. Compute

$$\alpha_k = \underset{\alpha \ge 0}{\operatorname{argmin}} f\left(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}\right),$$

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

4. Compute

$$\Delta\boldsymbol{x}^{(k)} = \alpha_k \boldsymbol{d}^{(k)},$$

$$\Delta\boldsymbol{g}^{(k)} = \boldsymbol{g}^{(k+1)} - \boldsymbol{g}^{(k)},$$

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

5. Set $k \coloneqq k + 1$; go to step 2.

Example 6.3.1. Let

$$f(x_1, x_2) = x_1^2 + \frac{1}{2} x_2^2 + 3.$$

Apply the rank one correction algorithm to minimize f. Use $\boldsymbol{x}^{(0)} = [1, 2]^\top$ and $\boldsymbol{H}_0 = \boldsymbol{I}_2$ (i.e., 2×2 identity matrix).

We can represent f as

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\top}\begin{bmatrix}2 & 0\\0 & 1\end{bmatrix}\boldsymbol{x} + 3$$

Thus,

$$\boldsymbol{g}^{(k)} = \begin{bmatrix}2 & 0\\0 & 1\end{bmatrix}\boldsymbol{x}^{(k)}.$$

Because $\boldsymbol{H}_0 = \boldsymbol{I}_2$,

$$\boldsymbol{d}^{(0)} = -\boldsymbol{H}_0\boldsymbol{g}^{(0)} = -\boldsymbol{g}^{(0)} = [-2,-2]^{\top}.$$

The objective function is quadratic, and hence

$$\alpha_0 = \operatorname*{argmin}_{\alpha\ge 0} f\left(\boldsymbol{x}^{(0)} + \alpha\boldsymbol{d}^{(0)}\right) = -\frac{\boldsymbol{g}^{(0)\top}\boldsymbol{d}^{(0)}}{\boldsymbol{d}^{(0)\top}\boldsymbol{Q}\boldsymbol{d}^{(0)}} = -\frac{[-2,-2]\begin{bmatrix}-2\\-2\end{bmatrix}}{[-2,-2]\begin{bmatrix}2 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}-2\\-2\end{bmatrix}} = \frac{2}{3},$$

and thus

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + \alpha_0\boldsymbol{d}^{(0)} = \left[-\frac{1}{3},\frac{2}{3}\right]^{\top}.$$

We then compute

$$\Delta\boldsymbol{x}^{(0)} = \alpha_0\boldsymbol{d}^{(0)} = \left[-\frac{4}{3},-\frac{4}{3}\right]^{\top},$$

$$\boldsymbol{g}^{(1)} = \boldsymbol{Q}\boldsymbol{x}^{(1)} = \left[-\frac{2}{3},\frac{2}{3}\right]^{\top},$$

$$\Delta\boldsymbol{g}^{(0)} = \boldsymbol{g}^{(1)} - \boldsymbol{g}^{(0)} = \left[-\frac{8}{3},-\frac{4}{3}\right]^{\top}.$$

Because

$$\Delta\boldsymbol{g}^{(0)\top}\left(\Delta\boldsymbol{x}^{(0)} - \boldsymbol{H}_0\Delta\boldsymbol{g}^{(0)}\right) = \left[-\frac{8}{3},-\frac{4}{3}\right]\left(\left[-\frac{4}{3},-\frac{4}{3}\right]^{\top} - \left[-\frac{8}{3},-\frac{4}{3}\right]^{\top}\right) = \left[-\frac{8}{3},-\frac{4}{3}\right]\begin{bmatrix}4/3\\0\end{bmatrix} = -\frac{32}{9}$$

we obtain

$$\boldsymbol{H}_1 = \boldsymbol{H}_0 + \frac{\left(\Delta\boldsymbol{x}^{(0)} - \boldsymbol{H}_0\Delta\boldsymbol{g}^{(0)}\right)\left(\Delta\boldsymbol{x}^{(0)} - \boldsymbol{H}_0\Delta\boldsymbol{g}^{(0)}\right)^{\top}}{\Delta\boldsymbol{g}^{(0)\top}\left(\Delta\boldsymbol{x}^{(0)} - \boldsymbol{H}_0\Delta\boldsymbol{g}^{(0)}\right)} = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} + \frac{[4/3,0]^{\top}[4/3,0]}{-32/9} = \begin{bmatrix}1/2 & 0\\0 & 1\end{bmatrix}.$$

Therefore,

$$\boldsymbol{d}^{(1)} = -\boldsymbol{H}_1\boldsymbol{g}^{(1)} = -\begin{bmatrix}1/2 & 0\\0 & 1\end{bmatrix}\left[-\frac{2}{3},\frac{2}{3}\right]^{\top} = \left[\frac{1}{3},-\frac{2}{3}\right]^{\top}$$

and

$$\alpha_1 = \underset{\alpha \geq 0}{\operatorname{argmin}} f\left(\boldsymbol{x}^{(1)} + \alpha \boldsymbol{d}^{(1)}\right) = -\frac{\boldsymbol{g}^{(1)\top}\boldsymbol{d}^{(1)}}{\boldsymbol{d}^{(1)\top}\boldsymbol{Q}\boldsymbol{d}^{(1)}} = \frac{\left[-\frac{2}{3}, \frac{2}{3}\right]\left[\frac{1}{3}, -\frac{2}{3}\right]^\top}{\left[\frac{1}{3}, -\frac{2}{3}\right]\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\left[\frac{1}{3}, -\frac{2}{3}\right]^\top} = 1.$$

We now compute

$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} + \alpha_1 \boldsymbol{d}^{(1)} = [0, 0]^\top.$$

Note that $\boldsymbol{g}^{(2)} = 0$, and therefore $\boldsymbol{x}^{(2)} = \boldsymbol{x}^*$.

- Unfortunately, the rank one correction algorithm is not very satisfactory, for several reasons.
- First, the matrix $\boldsymbol{H}_{k+1}$ that the rank one algorithm generates may not be positive definite and thus $\boldsymbol{d}^{(k+1)}$ may not be a descent direction. This happens even in the quadratic case.
- Furthermore, if $\Delta \boldsymbol{g}^{(k)\top}\left(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}\right)$ is close to zero, there may be numerical problems in evaluating $\boldsymbol{H}_{k+1}$.

Example 6.3.2 Let

$$f(x_1, x_2) = \frac{x_1^4}{4} + \frac{x_2^2}{2} - x_1 x_2 + x_1 - x_2.$$

Apply the rank one correction algorithm to minimize f. Use $\boldsymbol{x}^{(0)} = [0.59607, 0.59607]^\top$ and $\boldsymbol{H}_0 = \begin{bmatrix} 0.94913 & 0.14318 \\ 0.14318 & 0.59702 \end{bmatrix}$.

Note that the eigenvalues of $\boldsymbol{H}_0$ are 1.000002 and 0.546148. $\boldsymbol{H}_0 > 0$ (i.e., $\boldsymbol{H}_0$ is positive definite).

We have

$$\Delta \boldsymbol{g}^{(0)\top}\left(\Delta \boldsymbol{x}^{(0)} - \boldsymbol{H}_0 \Delta \boldsymbol{g}^{(0)}\right) = -0.03276$$

and

$$\boldsymbol{H}_1 = \begin{bmatrix} 0.94481 & 0.23324 \\ 0.23324 & -1.2788 \end{bmatrix}$$

Note that the eigenvalues of $\boldsymbol{H}_1$ are 0.9690117 and -1.3030017. $\boldsymbol{H}_1$ is not positive definite.

6.4 THE DFP ALGORITHM

- The rank two update was originally developed by Davidon in 1959 and was subsequently modified by Fletcher and Powell in 1963; hence the name DFP algorithm.

- **DFP Algorithm**

  1. Set $k := 0$; select $\boldsymbol{x}^{(0)}$ and a real symmetric positive definite $\boldsymbol{H}_0$.
  2. If $\boldsymbol{g}^{(k)} = \mathbf{0}$, stop; else, $\boldsymbol{d}^{(k)} = -\boldsymbol{H}_k \boldsymbol{g}^{(k)}$.
  3. Compute
  $$\alpha_k = \operatorname*{argmin}_{\alpha \geq 0} f\left(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}\right),$$
  $$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}.$$
  4. Compute
  $$\Delta \boldsymbol{x}^{(k)} = \alpha_k \boldsymbol{d}^{(k)},$$
  $$\Delta \boldsymbol{g}^{(k)} = \boldsymbol{g}^{(k+1)} - \boldsymbol{g}^{(k)},$$
  $$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \frac{\Delta \boldsymbol{x}^{(k)} \Delta \boldsymbol{x}^{(k)\top}}{\Delta \boldsymbol{x}^{(k)\top} \Delta \boldsymbol{g}^{(k)}} - \frac{\left(\boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}\right)\left(\boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}\right)^\top}{\Delta \boldsymbol{g}^{(k)\top} \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}}.$$
  5. Set $k := k + 1$; go to step 2.

- **Theorem 6.4.2** Suppose that $\boldsymbol{g}^{(k)} \neq \mathbf{0}$. In the DFP algorithm, if $\boldsymbol{H}_k$ is positive definite, then so is $\boldsymbol{H}_{k+1}$.

Example 6.4.1 Locate the minimizer of

$$f(\boldsymbol{x}) = \frac{1}{2} \boldsymbol{x}^\top \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} \boldsymbol{x} - \boldsymbol{x}^\top \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \boldsymbol{x} \in \mathbb{R}^2.$$

Use the initial point $\boldsymbol{x}^{(0)} = [0, 0]^\top$ and $\boldsymbol{H}_0 = \boldsymbol{I}_2$.

Note that in this case

$$\boldsymbol{g}^{(k)} = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} \boldsymbol{x}^{(k)} - \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Hence,

$$\boldsymbol{g}^{(0)} = [1, -1]^\top$$

$$\boldsymbol{d}^{(0)} = -\boldsymbol{H}_0 \boldsymbol{g}^{(0)} = -\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Because f is a quadratic function,

$$\alpha_0 = \operatorname*{argmin}_{\alpha \geq 0} f\left(\boldsymbol{x}^{(0)} + \alpha \boldsymbol{d}^{(0)}\right) = -\frac{\boldsymbol{g}^{(0)\top}\boldsymbol{d}^{(0)}}{\boldsymbol{d}^{(0)\top}\boldsymbol{Q}\boldsymbol{d}^{(0)}} = -\frac{[1,-1]\begin{bmatrix}-1\\1\end{bmatrix}}{[-1,1]\begin{bmatrix}4 & 2\\2 & 2\end{bmatrix}\begin{bmatrix}-1\\1\end{bmatrix}} = 1,$$

and thus

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + \alpha_0 \boldsymbol{d}^{(0)} = [-1,1]^\top.$$

We then compute

$$\begin{aligned}\Delta\boldsymbol{x}^{(0)} &= \boldsymbol{x}^{(1)} - \boldsymbol{x}^{(0)} = \alpha_0 \boldsymbol{d}^{(0)} = [-1,1]^\top,\\ \boldsymbol{g}^{(1)} &= \begin{bmatrix}4 & 2\\2 & 2\end{bmatrix}\begin{bmatrix}-1\\1\end{bmatrix} - \begin{bmatrix}-1\\1\end{bmatrix} = \begin{bmatrix}-1\\-1\end{bmatrix},\\ \Delta\boldsymbol{g}^{(0)} &= \boldsymbol{g}^{(1)} - \boldsymbol{g}^{(0)} = [-2,0]^\top.\end{aligned}$$

Observe that

$$\Delta\boldsymbol{x}^{(0)}\Delta\boldsymbol{x}^{(0)\top} = \begin{bmatrix}-1\\1\end{bmatrix}[-1,1] = \begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix},$$

$$\Delta\boldsymbol{x}^{(0)\top}\Delta\boldsymbol{g}^{(0)} = [-1,1]\begin{bmatrix}-2\\0\end{bmatrix} = 2,$$

$$\boldsymbol{H}_0\Delta\boldsymbol{g}^{(0)} = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}-2\\0\end{bmatrix} = \begin{bmatrix}-2\\0\end{bmatrix}.$$

Thus,

$$\left(\boldsymbol{H}_0\Delta\boldsymbol{g}^{(0)}\right)\left(\boldsymbol{H}_0\Delta\boldsymbol{g}^{(0)}\right)^\top = \begin{bmatrix}-2\\0\end{bmatrix}[-2,0] = \begin{bmatrix}4 & 0\\0 & 0\end{bmatrix},$$

and

$$\Delta\boldsymbol{g}^{(0)\top}\boldsymbol{H}_0\Delta\boldsymbol{g}^{(0)} = \begin{bmatrix}-2\\0\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}[-2,0] = 4.$$

Using the above, we now compute $\boldsymbol{H}_1$:

$$\begin{aligned}\boldsymbol{H}_1 &= \boldsymbol{H}_0 + \frac{\Delta\boldsymbol{x}^{(0)}\Delta\boldsymbol{x}^{(0)\top}}{\Delta\boldsymbol{x}^{(0)\top}\Delta\boldsymbol{g}^{(0)}} - \frac{\left(\boldsymbol{H}_0\Delta\boldsymbol{g}^{(0)}\right)\left(\boldsymbol{H}_0\Delta\boldsymbol{g}^{(0)}\right)^\top}{\Delta\boldsymbol{g}^{(0)\top}\boldsymbol{H}_0\Delta\boldsymbol{g}^{(0)}}\\ &= \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix} - \frac{1}{4}\begin{bmatrix}4 & 0\\0 & 0\end{bmatrix}\\ &= \begin{bmatrix}1/2 & -1/2\\-1/2 & 3/2\end{bmatrix}.\end{aligned}$$

We now compute $d^{(1)} = -H_1 g^{(1)} = [0,1]^\top$ and

$$\alpha_1 = \operatorname*{argmin}_{\alpha \geq 0} f\left(\boldsymbol{x}^{(1)} + \alpha \boldsymbol{d}^{(1)}\right) = -\frac{\boldsymbol{g}^{(1)\top}\boldsymbol{d}^{(1)}}{\boldsymbol{d}^{(1)\top}\boldsymbol{Q}\boldsymbol{d}^{(1)}} = \frac{1}{2}.$$

Hence,

$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} + \alpha_1 \boldsymbol{d}^{(1)} = [-1,3/2]^\top = \boldsymbol{x}^*.$$

6.5 THE BFGS ALGORITHM

- Recall that the approximation of the inverse of the Hessian matrix satisfies

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

which were derived from $\boldsymbol{Q}^{-1}\Delta\boldsymbol{g}^{(i)} = \Delta\boldsymbol{x}^{(i)}, 0 \le i \le k$.

- An alternative to approximating $\boldsymbol{Q}^{-1}$ is to approximate $\boldsymbol{Q}$ itself. To do this let $\boldsymbol{B}_k$ be our estimate of $\boldsymbol{Q}$ at the kth step. We require $\boldsymbol{B}_k$ to satisfy

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

- Notices that this set of equations is similar to the previous set of equations for $\boldsymbol{H}_{k+1}$, the only difference being that the roles of $\Delta\boldsymbol{x}^{(i)}$ and $\Delta\boldsymbol{g}^{(i)}$ are interchanged. Thus, given any update formula for $\boldsymbol{H}_k$, a corresponding update formula for $\boldsymbol{B}_k$ can be found by interchanging the roles of $\boldsymbol{B}_k$ and $\boldsymbol{H}_k$ and of $\Delta\boldsymbol{g}^{(i)}$ and $\Delta\boldsymbol{x}^{(i)}$:

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

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

- The approximation of the inverse Hessian is obtained by taking the inverse of $\boldsymbol{B}_{k+1}$:

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

- The result is based on the following formula for a matrix inverse, known as the **Sherman-Morrison formula**:

Lemma 6.1 Let $\boldsymbol{A}$ be a nonsingular matrix. Let u and v be column vectors such that $1 + v^\top A^{-1}u \neq 0$. Then, $A + uv^\top$ is nonsingular, and its inverse can be written in terms of A^{-1} using the following formula:

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

Chong, Edwin Kah Pin., and Stanislaw H. Żak. *An Introduction to Optimization*. Fourth edition., John Wiley & Sons, Inc., 2013.

Example 6.5.1 Use the BFGS method to minimize

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix} \boldsymbol{x} - \boldsymbol{x}^\top \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \log(\pi), \quad \boldsymbol{x} \in \mathbb{R}^2.$$

Use the initial point $\boldsymbol{x}^{(0)} = [0, 0]^\top$ and $\boldsymbol{H}_0 = \boldsymbol{I}_2$.

We have

$$\boldsymbol{d}^{(0)} = -\boldsymbol{g}^{(0)} = -\left(\boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b}\right) = \boldsymbol{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Because f is a quadratic function,

$$\alpha_0 = \underset{\alpha \geq 0}{\operatorname{argmin}} f\left(\boldsymbol{x}^{(0)} + \alpha \boldsymbol{d}^{(0)}\right) = -\frac{\boldsymbol{g}^{(0)\top}\boldsymbol{d}^{(0)}}{\boldsymbol{d}^{(0)\top}\boldsymbol{Q}\boldsymbol{d}^{(0)}} = \frac{1}{2},$$

Therefore,

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + \alpha_0 \boldsymbol{d}^{(0)} = [0, 1/2]^\top.$$

To compute $\boldsymbol{H}_1 = \boldsymbol{H}_1^{BFGS}$, we need the following quantities:

$$\Delta\boldsymbol{x}^{(0)} = \boldsymbol{x}^{(1)} - \boldsymbol{x}^{(0)} = [0, 1/2]^\top,$$
$$\boldsymbol{g}^{(1)} = \boldsymbol{Q}\boldsymbol{x}^{(1)} - \boldsymbol{b} = [-3/2, 0]^\top,$$
$$\Delta\boldsymbol{g}^{(0)} = \boldsymbol{g}^{(1)} - \boldsymbol{g}^{(0)} = [-3/2, 1]^\top.$$

Therefore,

$$\begin{aligned} \boldsymbol{H}_1 &= \boldsymbol{H}_0 + \left(1 + \frac{\Delta\boldsymbol{g}^{(0)\top}\boldsymbol{H}_0\Delta\boldsymbol{g}^{(0)}}{\Delta\boldsymbol{x}^{(0)\top}\Delta\boldsymbol{g}^{(0)}}\right)\frac{\Delta\boldsymbol{x}^{(0)}\Delta\boldsymbol{x}^{(0)\top}}{\Delta\boldsymbol{x}^{(0)\top}\Delta\boldsymbol{g}^{(0)}} - \frac{\boldsymbol{H}_0\Delta\boldsymbol{g}^{(0)}\Delta\boldsymbol{x}^{(0)\top} + \left(\boldsymbol{H}_0\Delta\boldsymbol{g}^{(0)}\Delta\boldsymbol{x}^{(0)\top}\right)^\top}{\Delta\boldsymbol{x}^{(0)\top}\Delta\boldsymbol{g}^{(0)}} \\ &= \begin{bmatrix} 1 & 3/2 \\ 3/2 & 11/4 \end{bmatrix}. \end{aligned}$$

Hence, we have

$$d^{(1)} = -H_1 g^{(1)} = [3/2, 9/4]^\top,$$

$$\alpha_1 = \underset{\alpha \geq 0}{\operatorname{argmin}} f\left(\boldsymbol{x}^{(1)} + \alpha \boldsymbol{d}^{(1)}\right) = -\frac{\boldsymbol{g}^{(1)\top}\boldsymbol{d}^{(1)}}{\boldsymbol{d}^{(1)\top}\boldsymbol{Q}\boldsymbol{d}^{(1)}} = 2.$$

Therefore,

$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} + \alpha_1 \boldsymbol{d}^{(1)} = [3, 5]^\top = \boldsymbol{x}^*.$$