Exercises in Optimization
(ACM 40990 / ACM41030)

Dr Lennon Ó Náraigh

Exercises #3

Exercises #3 - BFGS again and Trust-Region Methods

1. A simple way to approximate the Hessian (i.e. simpler than BFGS) is to use the so-called **symmetric rank-1** formula, defined by:

$$B_{k+1} = B_k + \frac{(\boldsymbol{y}_k - B_k\boldsymbol{s}_k)(\boldsymbol{y}_k - B_k\boldsymbol{s}_k)^T}{\langle \boldsymbol{y}_k - B_k\boldsymbol{s}_k, \boldsymbol{s}_k\rangle}$$

Unfortunately, this formula does not guarantee that the approximate Hessian is positive-definite. However, you should:

(a) Check that the update satisfies the Secant equation:

$$B_{k+1}\boldsymbol{s}_k = \boldsymbol{y}_k,$$

where

$$\boldsymbol{s}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k \qquad \boldsymbol{y}_k = \nabla f_{k+1} - \nabla f_k.$$

(b) Check that B_k is a symmetric matrix, for all $k \in 0, 1, 2, \cdots$.

Furthermore,

(c) You should show that the inverted Hessians $H_k := B_k^{-1}$ satisfy:

$$H_{k+1} = H_k + \frac{(\boldsymbol{s}_k - H_k\boldsymbol{y}_k)(\boldsymbol{s}_k - H_k\boldsymbol{y}_k)^T}{(\boldsymbol{s}_k - H_k\boldsymbol{y}_k)^T\,\boldsymbol{y}_k}$$

Hint: Use the Sherman–Morrison formula. Suppose $A \in \mathbb{R}^{n\times n}$ is an invertible square matrix and $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ are column vectors. Then $A+\boldsymbol{u}\boldsymbol{v}^T$ is invertible if and only if $1 + \langle \boldsymbol{v}, A^{-1}\boldsymbol{u}\rangle \neq 0$. In this case,

$$\left(A + \boldsymbol{u}\boldsymbol{v}^T\right)^{-1} = A^{-1} - \frac{A^{-1}\boldsymbol{u}\boldsymbol{v}^T A^{-1}}{1 + \langle \boldsymbol{v}, A^{-1}\boldsymbol{u}\rangle}.$$

We start with Part (b) first, because it will make the other calculations easier if we can assume that B_k and B_{k+1} are symmetric. This is an induction argument. We assume that B_k is symmetric. Then, B_{k+1} is of the form:

$$B_{k+1} = B_k + \lambda \boldsymbol{u}\boldsymbol{u}^T,$$

where $\lambda = \langle \boldsymbol{y}_k - B_k \boldsymbol{s}_k, s_k \rangle^{-1}$ and $\boldsymbol{u} = \boldsymbol{y}_k - B_k \boldsymbol{s}_k$. The components of this matrix are:

$$(B_{k+1})_{ij} = (B_k)_{ij} + \lambda u_i u_j.$$

By the induction hypothesis, $(B_k)_{ij} = (B_k)_{ji}$ so the matrix on the RHS is symmetric, hence B_{k+1} is a symmetric matrix. Therefore, provided the starting matrix B_0 is symmetric, by mathematical induction, all of the matrices $B_1, B_2, \cdots$ will be symmetric.

We look at Part (a) next. We will look at the components of the vector $B_{k+1}\boldsymbol{s}_k$ with respect to the usual basis $\{\boldsymbol{e}_i\}_{i=1}^n$. We have:

$$\langle B_{k+1}\boldsymbol{s}_k, \boldsymbol{e}_i \rangle = \langle B\boldsymbol{s}, \boldsymbol{e}_i \rangle + \frac{\langle (\boldsymbol{y} - B\boldsymbol{s})(\boldsymbol{y} - B\boldsymbol{s})^T \boldsymbol{s}, \boldsymbol{e}_i \rangle}{\langle \boldsymbol{y} - B\boldsymbol{s}, \boldsymbol{s} \rangle}.$$

(we omit the subscript k's on the RHS for clarity). We also have:

$$\begin{aligned}
\langle \boldsymbol{u}\boldsymbol{u}^T \boldsymbol{s}, \boldsymbol{e}_i \rangle &= \left[\boldsymbol{u}\boldsymbol{u}^T \boldsymbol{s}\right]_i, \\
&= \sum_j [\boldsymbol{u}\boldsymbol{u}^T]_{ij} s_j, \\
&= \sum_j u_i u_j s_j, \\
&= u_i \langle \boldsymbol{u}, \boldsymbol{s} \rangle, \\
&= \langle \boldsymbol{u}, \boldsymbol{e}_i \rangle \langle \boldsymbol{u}, \boldsymbol{s} \rangle.
\end{aligned}$$

Putting this all together, we have:

$$\begin{aligned}
\langle B_{k+1}\boldsymbol{s}_k, \boldsymbol{e}_i \rangle &= \langle B\boldsymbol{s}, \boldsymbol{e}_i \rangle + \frac{\langle (\boldsymbol{y} - B\boldsymbol{s}), \boldsymbol{e}_i \rangle \langle (\boldsymbol{y} - B\boldsymbol{s}), \boldsymbol{s} \rangle}{\langle \boldsymbol{y} - B\boldsymbol{s}, \boldsymbol{s} \rangle}, \\
&= \frac{\langle B\boldsymbol{s}, \boldsymbol{e}_i \rangle \langle \boldsymbol{y} - B\boldsymbol{s}, \boldsymbol{s} \rangle + \langle (\boldsymbol{y} - B\boldsymbol{s}), \boldsymbol{e}_i \rangle \langle (\boldsymbol{y} - B\boldsymbol{s}), \boldsymbol{s} \rangle}{\langle \boldsymbol{y} - B\boldsymbol{s}, \boldsymbol{s} \rangle}.
\end{aligned}$$

Hence,

$$\langle B_{k+1}\boldsymbol{s}_k, \boldsymbol{e}_i \rangle =$$
$$\frac{\cancel{\langle B\boldsymbol{s}, \boldsymbol{e}_i \rangle \langle \boldsymbol{y}, \boldsymbol{s} \rangle} - \cancel{\langle B\boldsymbol{s}, \boldsymbol{e}_i \rangle \langle B\boldsymbol{s}, \boldsymbol{s} \rangle} + \langle \boldsymbol{y}, \boldsymbol{e}_i \rangle \langle \boldsymbol{y}, \boldsymbol{s} \rangle - \cancel{\langle B\boldsymbol{s}, \boldsymbol{e}_i \rangle \langle \boldsymbol{y}, \boldsymbol{s} \rangle} - \langle \boldsymbol{y}, \boldsymbol{e}_i \rangle \langle B\boldsymbol{s}, \boldsymbol{s} \rangle + \cancel{\langle B\boldsymbol{s}, \boldsymbol{e}_i \rangle \langle B\boldsymbol{s}, \boldsymbol{s} \rangle}}{\langle \boldsymbol{y} - B\boldsymbol{s}, \boldsymbol{s} \rangle}$$

Continuing thus, we have:

$$\begin{aligned}
\langle B_{k+1}\boldsymbol{s}_k, \boldsymbol{e}_i \rangle &= \frac{\langle \boldsymbol{y}, \boldsymbol{e}_i \rangle \left[\langle \boldsymbol{y}, \boldsymbol{s} \rangle - \langle B\boldsymbol{s}, \boldsymbol{s} \rangle\right]}{\langle \boldsymbol{y} - B\boldsymbol{s}, \boldsymbol{s} \rangle}, \\
&= \langle \boldsymbol{y}, \boldsymbol{e}_i \rangle.
\end{aligned}$$

As this is true for all $i \in \{1, 2, \cdots, n\}$ it follows that:

$$B_{k+1}\boldsymbol{s}_k = \boldsymbol{y}_k,$$

and thus, the symmetric rank-1 formula satisfies the Secant equation.

For Part (c), a direct application of the Sherman–Morrison formula (with the subscript k's suppressed on the RHS) gives:

$$\begin{aligned} H_{k+1} &= H - \frac{H(\boldsymbol{y} - B\boldsymbol{s})[H(\boldsymbol{y} - B\boldsymbol{s})^T]/\langle \boldsymbol{y} - B\boldsymbol{s}, \boldsymbol{s}\rangle}{1 + \langle \boldsymbol{y} - B\boldsymbol{s}, H(\boldsymbol{y} - B\boldsymbol{s})\rangle/\langle \boldsymbol{y} - B\boldsymbol{s}, \boldsymbol{s}\rangle}, \\ &= H - \frac{(H\boldsymbol{y} - \boldsymbol{s})(H\boldsymbol{y} - \boldsymbol{s})^T}{\langle \boldsymbol{y} - B\boldsymbol{s}, \boldsymbol{s}\rangle + \langle \boldsymbol{y} - B\boldsymbol{s}, H(\boldsymbol{y} - B\boldsymbol{s})\rangle}. \end{aligned}$$

We work on the denominator:

$$\begin{aligned} \langle \boldsymbol{y} - B\boldsymbol{s}, \boldsymbol{s}\rangle + \langle \boldsymbol{y} - B\boldsymbol{s}, H(\boldsymbol{y} - B\boldsymbol{s})\rangle &= \langle \boldsymbol{y}, \boldsymbol{s}\rangle - \langle B\boldsymbol{s}, \boldsymbol{s}\rangle + \langle \boldsymbol{y} - B\boldsymbol{s}, H\boldsymbol{y} - \boldsymbol{s}\rangle, \\ &= \cancel{\langle \boldsymbol{y}, \boldsymbol{s}\rangle} - \cancel{\langle B\boldsymbol{s}, \boldsymbol{s}\rangle} + \langle \boldsymbol{y}, H\boldsymbol{y}\rangle - \cancel{\langle \boldsymbol{y}, \boldsymbol{s}\rangle} - \langle B\boldsymbol{s}, H\boldsymbol{y}\rangle + \cancel{\langle B\boldsymbol{s}, \boldsymbol{s}\rangle} \\ &= \langle \boldsymbol{y}, H\boldsymbol{y}\rangle - \langle \boldsymbol{s}, \boldsymbol{y}\rangle = \langle \boldsymbol{y}, H\boldsymbol{y} - \boldsymbol{s}\rangle. \end{aligned}$$

Thus,

$$H_{k+1} = H - \frac{(H\boldsymbol{y} - \boldsymbol{s})(H\boldsymbol{y} - \boldsymbol{s})^T}{\langle \boldsymbol{y}, H\boldsymbol{y} - \boldsymbol{s}\rangle}$$

We now restore the subscript k's on the RHS and apply $(-1)(-1) = 1$ in various places to obtain the standard formula:

$$H_{k+1} = H_k + \frac{(\boldsymbol{s}_k - H_k\boldsymbol{y}_k)(\boldsymbol{s}_k - H_k\boldsymbol{y}_k)^T}{\langle \boldsymbol{s}_k - H_k\boldsymbol{y}_k, \boldsymbol{y}_k\rangle}.$$

2. Write a code (in whatever programming langauge) that uses the Trust-Region method (Dogleg method) to solve the Rosenbrock problem

$$f = 10(x_2 - x_1^2)^2 + (1 - x_1)^2.$$

TBC