Bioinformatics

Discrete Mathematics and Optimisation

Solutions to Problem Sheet Iterative Methods

Exercise 1. We have $f(x) = x^2 + \sqrt{x} - 15$ and $f'(x) = 2x + 1/(2\sqrt{x})$.

(a) We start at $x_0 = 4$. We get

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 4 - \frac{3}{33/4} = 4 - \frac{12}{33} = \frac{40}{11}.$$

And

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} \approx \frac{40}{11} - \frac{0.1301}{8.1019} \approx 3.619102.$$

Lastly, we have

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} \approx 3.619063.$$

(b) We start at $x_0 = 1$. We have

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 1 - \frac{-13}{3/2} = \frac{31}{5}.$$

And furthermore

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} \approx 4.142197.$$

Finally, we get

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} \approx 3.650637.$$

Exercise 2.

(a) Notice that we can rewrite $f(x, y)$ as the following sum of two functions

$$f(x, y) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \ln(xy).$$

The first summand of $f(x, y)$ is a quadratic function whose Hessian matrix is given by $\begin{pmatrix} 2 & -4 \\ -4 & 10 \end{pmatrix}$. Its second leading principal minor (the determinant) is equal to 4 while its leading principal minor is equal to 2, hence they are both positive so that by Sylvester's criterion $\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$ is a strictly convex function in C. The second summand is the negation of the logarithmic function defined over positive numbers xy (as $x > 0$ and $y > 0$), which is a concave function, and is thus striclty convex in C. Hence, $f(x, y)$ is also strictly convex in C, as the sum of two strictly convex functions.

(b) We first compute the gradient of $f(x,y)$

$$\nabla f(x,y) = \left(2x - 4y - \frac{1}{x}, 10y - 4x - \frac{1}{y}\right).$$

So that its Hessian matrix is given by

$$H_f(x,y) = \begin{pmatrix} 2 + x^{-2} & -4 \\ -4 & 10 + y^{-2} \end{pmatrix}.$$

Starting at $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, the next iteration of the Newton method gives

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - H_f(x_0,y_0)^{-1} \nabla f(x_0,y_0),$$

$$\begin{aligned} \text{where} \quad \nabla f(1,1) = \begin{pmatrix} -3 \\ 5 \end{pmatrix}, \quad \text{while} \quad H_f(1,1)^{-1} &= \begin{pmatrix} 3 & -4 \\ -4 & 11 \end{pmatrix}^{-1} \\ &= \frac{\mathrm{Adj}(H_f(1,1))}{\det(H_f(1,1))} \\ &= \frac{1}{17}\begin{pmatrix} 11 & 4 \\ 4 & 3 \end{pmatrix}. \end{aligned}$$

Thus we obtain

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{17}\begin{pmatrix} 11 & 4 \\ 4 & 3 \end{pmatrix}\begin{pmatrix} -3 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{17}\begin{pmatrix} -13 \\ 3 \end{pmatrix} = \begin{pmatrix} 30/17 \\ 14/17 \end{pmatrix} \approx \begin{pmatrix} 1.76 \\ 0.82 \end{pmatrix}.$$

Similarly, the second iteration gives

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} 30/17 \\ 14/17 \end{pmatrix} - \frac{176400}{1875761}\begin{pmatrix} 2249/196 & 4 \\ 4 & 2089/900 \end{pmatrix}\begin{pmatrix} -169/510 \\ -9/238 \end{pmatrix} \approx \begin{pmatrix} 2.14 \\ 0.96 \end{pmatrix}.$$

Exercise 3.

(a) The gradient and the Hessian matrix of f at any point $(x,y) \in \mathbb{R}^2$ are given by

$$\nabla f(x,y) = \left(2x, 2ye^{y^2}\right) \quad \text{and} \quad H_f(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 2e^{y^2}(2y^2+1) \end{pmatrix}.$$

The leading princpals minors of $H_f(x,y)$ are

$$\Delta_1 = 2 > 0 \quad \text{and} \quad \Delta_2 = 4e^{y^2}(2y^2+1) \geq 5 > 0, \quad \text{as } \forall y \in \mathbb{R}, \quad y^2 \geq 0 \text{ and } e^{y^2} \geq 1.$$

Thus, by Sylvester's criterion, $H_f(x,y)$ is definite positive for any $(x,y) \in \mathbb{R}^2$. This means that $f(x,y)$ is strictly convex in $\mathbb{R}^2$. And together with the fact that $\mathbb{R}^2$ has no boundary point, this implies that $f(x,y)$ admits a unique global minimum in $\mathbb{R}^2$.

The next point of the Newton method on $f(x,y)$ starting at $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - H_f(x_0,y_0)^{-1} \nabla f(x_0,y_0).$$

We have $\nabla f(x_0,y_0) = \begin{pmatrix} 2 \\ 2e \end{pmatrix}$, where $e = e^1$ is the *Euler constant.* And

$$H_f(x_0,y_0)^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 6e \end{pmatrix}^{-1} = \frac{\mathrm{Adj}(H_f(x_0,y_0))}{\det(H_f(x_0,y_0))} = \frac{1}{12e} \begin{pmatrix} 6e & 0 \\ 0 & 2 \end{pmatrix}.$$

This gives

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{12e} \begin{pmatrix} 6e & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 2e \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{12e} \begin{pmatrix} 12e \\ 4e \end{pmatrix} = \begin{pmatrix} 0 \\ 2/3 \end{pmatrix}.$$

(b) The first part is similar to (a). The gradient and the Hessian matrix of g at any point $(x,y) \in \mathbb{R}^2$ are given by

$$\nabla g(x,y) = (4x^3, 4y) \quad \text{and} \quad H_g(x,y) = \begin{pmatrix} 12x^2 & 0 \\ 0 & 4 \end{pmatrix}.$$

The leading princpals minors of $H_g(x,y)$ are

$$\Delta_1 = 12x^2 > 0, \quad \text{when } x \neq 0, \quad \text{and} \quad \Delta_2 = 48x^2 > 0, \quad \text{when } x \neq 0.$$

Thus, by Sylvester's criterion, $H_g(x,y)$ is definite positive for any $(x,y) \in (\mathbb{R} \setminus \{0\}) \times \mathbb{R}$. Hence, $g(x,y)$ is strictly convex in $(\mathbb{R} \setminus \{0\}) \times \mathbb{R}$ and it admits a unique strict global minimum. In fact, $g(x,y) > 0 = g(0,0)$ whenever $(x,y) \neq (0,0)$ so that $(0,0)$ is the unique strict global minimum of $g(x,y)$ in $\mathbb{R}^2$.

For the second part, we compute the first iteration of the steepest descent method on $g(x,y)$ starting at $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. It is given by the point

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - t_0 \nabla g(x_0,y_0),$$

where $t_0 \in \mathbb{R}$ minimises the function

$$\phi_0(t) = g\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right) = g\left(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - t\nabla g(x_0,y_0)\right).$$

We first compute $\nabla g(1,2) = \begin{pmatrix} 4 \\ 8 \end{pmatrix}$. Second, notice that t_0 is determined by the equation $\phi_0'(t) = 0$, such that $\phi_0''(t) > 0$ holds. Using the chain rule, we have

$$\begin{aligned} \phi_0'(t) = -\nabla g(1,2) \cdot \nabla g\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix} - t\nabla g(1,2)\right) &= -\begin{pmatrix} 4 \\ 8 \end{pmatrix} \cdot \nabla g\left(\begin{pmatrix} 1-4t \\ 2-8t \end{pmatrix}\right) \\ &= -\begin{pmatrix} 4 \\ 8 \end{pmatrix} \cdot \begin{pmatrix} 4(1-4t)^3 \\ 8(1-4t) \end{pmatrix} \\ &= -16(1-4t)((1-4t)^2+4). \end{aligned}$$

Thus

$$\phi_0'(t) = 0 \Leftrightarrow -16(1-4t)((1-4t)^2+4) = 0.$$

The above polynomial equation in t has three roots: one real $1/4$ and two complex roots with non-negative imaginary parts. This means that $t_0 = 1/4$ is the unique real critical point of $\phi_0(t)$. By defintion of ϕ_0, it is in fact a strict global minimum as $\phi_0(t)$ is strictly convex due to the fact that $g(x,y)$ is. In any case, t_0 is a strict local minimum of $\phi_0(t)$, because $\phi_0''(t_0) = 256 > 0$. This gives

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \frac{1}{4}\begin{pmatrix} 4 \\ 8 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

We have reached the minimum in one step.

Exercise 4.

(a) Let $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$. If we set $Q = \begin{pmatrix} 2 & -1/2 \\ -1/2 & 1 \end{pmatrix}$, $b = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $c = 4$, then we obtain

$$f(x,y) = \mathbf{x}^T Q \mathbf{x} + \mathbf{x}^T b + c.$$

(b) From the notation in (a), we get the following

$$\nabla f(x,y) = 2Q\mathbf{x} + b = \begin{pmatrix} 4x - y + 2 \\ 2y - x + 1 \end{pmatrix} \quad \text{and} \quad H_f(x,y) = 2Q = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}.$$

In particular, the leading principal minors of the Hessian matrix of $f(x,y)$ satisfy

$$\Delta_1(H_f(x,y)) = 4 > 0 \quad \text{and} \quad \Delta_2(H_f(x,y)) = \det(2Q) = 7 > 0.$$

By Sylvester's criterion this implies that $H_f(x,y)$ is definite positive $\forall (x,y) \in \mathbb{R}^2$, which in turns means that $f(x,y)$ is strictly convex in $\mathbb{R}^2$. Thus, as $\mathbb{R}^2$ has no boundary point, $f(x,y)$ admits a unique global minimum in $\mathbb{R}^2$.

The first iteration of the Newton method with starting point $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$ is given by the point

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - H_f(x_0,y_0)^{-1}\nabla f(x_0,y_0),$$

where $\nabla f(-1,-1) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$, while
$$H_f(-1,-1)^{-1} = (2Q)^{-1} = \frac{\mathrm{Adj}(2Q)}{\det(2Q)} = \frac{1}{7}\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}.$$

Thus

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix} - \frac{1}{7}\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}\begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix} - \frac{1}{7}\begin{pmatrix} -2 \\ -1 \end{pmatrix} = \begin{pmatrix} -5/7 \\ -6/7 \end{pmatrix}.$$

(c) Let again $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$ be the initial point of the steepest descent method. Then the next point is given by

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - t_0 \nabla f(x_0, y_0),$$

where $t_0 \in \mathbb{R}$ is such that it minimises the function

$$\phi_0(t) = f\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right) = f\left(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - t\nabla f(x_0, y_0)\right).$$

Remark that t_0 is determined by the equation $\phi_0'(t) = 0$, where using the chain rule

$$\begin{aligned} \phi_0'(t) = -\nabla f\left(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}\right) \cdot \nabla f\left(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - t\nabla f(x_0, y_0)\right) &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cdot \nabla f\left(\begin{pmatrix} t-1 \\ -1 \end{pmatrix}\right) \\ &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 4(t-1)+3 \\ -t \end{pmatrix}. \end{aligned}$$

So that we have

$$\phi_0'(t) = 0 \Leftrightarrow 4(t-1) + 3 = 0 \Rightarrow t_0 = 1/4.$$

In particular, $t_0 = 1/4$ is a critical point of $\phi_0(t)$. Furthermore, by definition of ϕ_0 and because $f(x, y)$ is strictly convex in $\mathbb{R}^2$, so is $\phi_0(t)$ in $\mathbb{R}$. This means that t_0 is the strict global minimum of $\phi_0(t)$ in $\mathbb{R}$. In any case, the fact that t_0 is a strict local minimum of $\phi_0(t)$ can also be proven by verifying that $\phi_0''(t_0) = 4 > 0$. And we finally get

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix} - \frac{1}{4}\begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -3/4 \\ -1 \end{pmatrix}.$$

Exercise 5.

(a) For each fixed $(x_0, y_0) \in \mathbb{R}^2$, the function $(x-x_0)^2+(y-y_0)^2$ is convex in $\mathbb{R}^2$. Therefore, $f(x, y)$ is a convex function in $\mathbb{R}^2$ as a sum of three convex functions. To see that it is furthermore strictly convex, we will show next that the Hessian matrix of $f(x, y)$ is definite positive. The gradient and the Hessian matrix of $f(x, y)$ at any point $(x, y) \in \mathbb{R}^2$ are given by

$$\nabla f(x, y) = (6x - 10, 6y - 12) \quad \text{and} \quad H_f(x, y) = \begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix}.$$

The leading principal minors verify $\Delta_1 = 6 > 0$ and $\Delta_2 = 36 > 0$. By Sylvester's criterion, $H_f(x, y)$ is definite positive thus $f(x, y)$ is stirctly convex in $\mathbb{R}^2$.

(b) Since $f(x, y)$ is strictly convex in $\mathbb{R}^2$, it admits a unique global minimum $d = (d_1, d_2)$ which must be a critical point, i.e. it must satisfy

$$6d_1 - 10 = 0 \quad \text{and} \quad 6d_2 - 12 = 0 \quad \Longrightarrow \quad d = (5/3, 2).$$

(c) The first iteration of the steepest descent method starting at $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is given by the point

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - t_0 \nabla f(x_0, y_0),$$

where $t_0 \in \mathbb{R}$ minimises the function

$$\phi_0(t) = f\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right) = f\left(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - t\nabla f(x_0, y_0)\right).$$

Notice that t_0 is determined by the equation $\phi_0'(t) = 0$, such that $\phi_0''(t) > 0$ holds. Using the chain rule, this gives

$$\begin{aligned}
\phi_0'(t) = -\nabla f(1,1) \cdot \nabla f\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} - t\nabla f(1,1)\right) &= \begin{pmatrix} 4 \\ 6 \end{pmatrix} \cdot \nabla f\left(\begin{pmatrix} 1+4t \\ 1+6t \end{pmatrix}\right) \\
&= \begin{pmatrix} 4 \\ 6 \end{pmatrix} \cdot \begin{pmatrix} 6(1+4t) - 10 \\ 6(1+6t) - 12 \end{pmatrix} \\
&= 312t - 52.
\end{aligned}$$

So that

$$\phi_0'(t_0) = 0 \Leftrightarrow t_0 = \frac{1}{6}.$$

By defintion of ϕ_0, t_0 is in fact a strict global minimum as $\phi_0(t)$ is strictly convex due to the fact that $f(x, y)$ is. In any case, t_0 is a strict local minimum of $\phi_0(t)$, because $\phi_0''(t_0) = 312 > 0$. This gives

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{6}\begin{pmatrix} 4 \\ 6 \end{pmatrix} = \begin{pmatrix} 5/3 \\ 2 \end{pmatrix}.$$

We have reached the minimum in one step.

(d) The gradient and the Hessian matrix of g at a point $(x, y) \in \mathbb{R}^2$ are given by

$$\nabla g(x, y) = (3x^2y^3 + 6x - 10, 3x^3y^2 + 6y - 12), \quad H_g(x, y) = \begin{pmatrix} 6xy^3 + 6 & 9x^2y^2 \\ 9x^2y^2 & 6x^3y + 6 \end{pmatrix}.$$

Let furtermore $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, so that $\nabla g(1,1) = -\begin{pmatrix} 1 \\ 3 \end{pmatrix}$, and

$$H_g(1,1)^{-1} = \begin{pmatrix} 12 & 9 \\ 9 & 12 \end{pmatrix}^{-1} = \frac{\mathrm{Adj}(H_f(1,1))}{\det(H_f(1,1))} = \frac{1}{63}\begin{pmatrix} 12 & -9 \\ -9 & 12 \end{pmatrix} = \frac{1}{21}\begin{pmatrix} 4 & -3 \\ -3 & 4 \end{pmatrix}.$$

Thus, the next iteration of the Newton method starting at $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ is

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{21}\begin{pmatrix} 4 & -3 \\ -3 & 4 \end{pmatrix}\begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{21}\begin{pmatrix} -5 \\ 9 \end{pmatrix} = \begin{pmatrix} 16/21 \\ 10/7 \end{pmatrix}.$$