Bioinformatics

Discrete Mathematics and Optimisation

Solutions to Problem Sheet Convexity and Newton method

Exercise 1.

(a) The Hessian matrix of $f_1(x, y, z)$ is given by

$$H_{f_1}(x, y, z) = \begin{pmatrix} 12x^2 + 4 & -2 & 0 \\ -2 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Furthermore, we have

$$\begin{aligned}
\Delta_1 &= \det\left(12x^2 + 4\right) = 12x^2 + 4 > 0, \quad \text{as } x^2 \geq 0 \quad \forall x \in \mathbb{R}, \\
\Delta_2 &= \det\begin{pmatrix} 12x^2 + 4 & -2 \\ -2 & 2 \end{pmatrix} = 24x^2 + 4 > 0, \quad \text{as } x^2 \geq 0 \quad \forall x \in \mathbb{R}, \\
\Delta_3 &= \det\begin{pmatrix} 12x^2 + 4 & -2 & 0 \\ -2 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} = 2\Delta_2 > 0, \quad \text{as } \Delta_2 > 0.
\end{aligned}$$

Hence, by Sylvester's criterion $H_{f_1}(x, y, z)$ is definite positive for any $(x, y, z) \in \mathbb{R}^3$. Therefore $f_1(x, y, z)$ is stricIty convex in $\mathbb{R}^3$.

(b) The function x^2 is strictly convex in $\mathbb{R}$ as it is a monomial with an even power, so that the same holds for $5x^2$ as the scalar $5 > 0$. Similarly, $9y^4$ is strictly convex in $\mathbb{R}$. Furthermore, both e^x and e^{-z} are strictly convex in $\mathbb{R}$ by definition of the exponential function.

Therefore, $f_2(y, y, z)$ is the sum of functions that are strictly convex and is hence strictly convex in $\mathbb{R}^3$.

(c) By arguments analogue to those in (b), the function $x^2 + y^2$ is convex in $\mathbb{R}^2$. Thus $f_3(x, y) = -(x^2 + y^2)$ is concave in $\mathbb{R}^2$.

(d) The Hessian matrix of $f_4(x, y)$ is given by

$$H_{f_4}(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

And we have $\det(H_{f_4}(x, y)) = -4 < 0$. Therefore by Sylvester's criterion, $H_{f_4}(x, y)$ is indefinite and $f_4(x, y)$ is neither convex nor concave in $\mathbb{R}^2$.

However, notice that when restricting its domain to $\mathbb{R} \times \{0\}$, the function $f_4(x, y)$ is convex. While when restricting its domain to $\{0\} \times \mathbb{R}$, it is concave.

Exercise 2. $f(x,y) = (x-y)^2 + (x+2y+1)^2 - 8xy = 2x^2 - 6xy + 5y^2 + 2x + 4y + 1$.

(a) Let $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$. If we set $Q = \begin{pmatrix} 2 & -3 \\ -3 & 5 \end{pmatrix}$, $b = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$ and $c = 1$, then we obtain

$$\begin{aligned} f(x,y) = (x-y)^2 + (x+2y+1)^2 - 8xy &= 2x^2 - 6xy + 5y^2 + 2x + 4y + 1 \\ &= \mathbf{x}^T Q \mathbf{x} + \mathbf{x}^T b + c. \end{aligned}$$

(b) The Hessian matrix of $f(x,y)$ is given by

$$H_f(x,y) = 2Q = \begin{pmatrix} 4 & -6 \\ -6 & 10 \end{pmatrix}.$$

Furthermore, we have $\Delta_1 = 4 > 0$ and $\Delta_2 = \det(2Q) = 4 > 0$. So that by Sylvester's criterion, $H_f(x,y)$ is a definite positive matrix for any $(x,y) \in \mathbb{R}^2$. This implies that $f(x,y)$ is strictly convex in $\mathbb{R}^2$.

(c) The critical points of $f(x,y)$ are determined by

$$\nabla f(x,y) = 2Q\mathbf{x} + b = \begin{pmatrix} 4 & -6 \\ -6 & 10 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 4x - 6y + 2 \\ 10y - 6x + 4 \end{pmatrix} = 0.$$

The only solution is $(x,y) = (-11,-7)$. Since $f(x,y)$ is strictly convex in $\mathbb{R}^2$, which admits no boundary point, $(-11,-7)$ is the global minimum of $f(x,y)$ in $\mathbb{R}^2$.

Exercise 3.

(a) Let $x, y \in C$ and $0 \le \lambda \le 1$. First, because C is a convex set we have $\lambda x + (1-\lambda)y \in C$. Furthermore, since both functions $f(x)$ and $g(x)$ are convex in C we also have

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y) \text{ and } g(\lambda x + (1-\lambda)y) \le \lambda g(x) + (1-\lambda) g(y).$$

It follows that

$$\begin{aligned} h(\lambda x + (1-\lambda)x) &= \max\big(f(\lambda x + (1-\lambda)y), g(\lambda x + (1-\lambda)y)\big) \\ &\le \max\big(\lambda f(x) + (1-\lambda)f(y), \lambda g(x) + (1-\lambda)g(y)\big) \\ &\le \max\big(\lambda f(x), \lambda g(x)\big) + \max\big((1-\lambda)f(y), (1-\lambda)g(y)\big) \\ &= \lambda \max\big(f(x), g(x)\big) + (1-\lambda)\max\big(f(y), g(y)\big) \\ &\le \lambda h(x) + (1-\lambda)h(y). \end{aligned}$$

Hence, by definition, $h(x)$ is also convex in C.

(b) Let $x' \in C$ be a global minimum of $h(x)$ in C, i.e. $h(x') \le h(x)$ for all $x \in C$.

First, if $g(x') = f(x')$ then we are done. So only the case $g(x') \ne f(x')$ remains. We can then assume without loss of generality that $g(x') < f(x') = h(x')$. And we will show next that x' is also a global minimum of $f(x)$ in C.

Assume to a contradiction that there exists some $x'' \in C$ such that $f(x'') < f(x')$. Since $h(x'') \ge h(x')$, it follows that $g(x'') = h(x'') \ge h(x') = f(x')$. This means that there exists some $0 < \lambda_0 < 1$ satisfying

$$\lambda_0 g(x') + (1-\lambda_0) g(x'') = \lambda_0 f(x') + (1-\lambda_0) f(x'').$$

Furthermore, by convexity of C, we have $\lambda_0 x' + (1-\lambda_0)x'' \in C$. But, by convexity of f and g in C, we also have that both

$$f(\lambda_0 x' + (1-\lambda_0)x'') \leq \lambda_0 f(x') + (1-\lambda_0) f(x'') < f(x') = h(x'),$$

and

$$\begin{aligned} g(\lambda_0 x' + (1-\lambda_0)x'') \leq \lambda_0 g(x') + (1-\lambda_0) g(x'') &= \lambda_0 f(x') + (1-\lambda_0) f(x'') \\ &< f(x') = h(x'). \end{aligned}$$

This gives

$$h(\lambda_0 x' + (1-\lambda_0)x'') < h(x').$$

This contradicts the assumption that x' is a global minimum of $h(x)$ in C.

Exercise 4.

(a) The gradient of $g(x,y)$ is given by

$$\nabla g(x,y) = (4x^3 + 4x - y, 4y^3 + 2y - x).$$

While its Hessian matrix is

$$H_g(x,y) = \begin{pmatrix} 12x^2+4 & -1 \\ -1 & 12y^2+2 \end{pmatrix}.$$

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

$$\begin{aligned} \Delta_1 &= 12x^2 + 4 \geq 4 > 0, \quad \text{as } x^2 \geq 0 \quad \forall x \in \mathbb{R}, \\ \Delta_2 &= (12x^2+4)(12y^2+2) - 1 \\ &= 144x^2y^2 + 24x^2 + 48y^2 + 7 \geq 7 > 0, \quad \text{as } x^2, y^2 \geq 0 \quad \forall x, y \in \mathbb{R}. \end{aligned}$$

Hence by Sylvester's criterion $H_g(x,y)$ is a definite positive matrix for any $(x,y) \in \mathbb{R}^2$. This implies that $g(x,y)$ is a strictly convex function in $\mathbb{R}^2$. And because the exponential function is increasing over $\mathbb{R}$, this further means that $f(x,y) = e^{g(x,y)}$ is also a strictly convex function in $\mathbb{R}^2$.

(c) Another consequence of the fact that the exponential function is increasing over $\mathbb{R}$ is that $f(x,y)$ will reach its (global as f is convex) minimum when $g(x,y)$ is minimum.

From (b), we know that $g(x,y)$ is convex in $\mathbb{R}^2$, thus its (unique) critical point will be a global minimum in $\mathbb{R}^2$. The latter is determined by the equation $\nabla g(x,y) = 0$, whose unique solution in $\mathbb{R}^2$ is given by $(x,y) = (0,0)$.

Thus the global minimum of $f(x,y)$ in $\mathbb{R}^2$ is $(0,0)$.