**Problem 1.**

Consider the unconstrained optimisation problem

$$\min_{x,y} f(x, y),$$

where $f\colon \mathbb{R}^2 \to \mathbb{R}$ is defined as

$$f(x, y) = x^4 - 2x^2y + 2y^2 + 2x^2 + 4y - 3.$$

a) Find all local and global minimisers of $f$.
*(10 points)*

b) Determine whether the function $f$ is convex.
*(5 points)*

c) Perform one step of Newton's method with backtracking Armijo line search starting from the point $(x, y) = (1, -1)$. Use the parameters $c_1 = 0.1$ (sufficient decrease parameter) and $\rho = 0.5$ (contraction factor).
*(10 points)*

**Solution.**

a) We start by computing the gradient of $f$: We have

$$\nabla f(x, y) = \begin{pmatrix} 4x^3 - 4xy + 4x \\ -2x^2 + 4y + 4 \end{pmatrix}.$$

The first order optimality condition $\nabla f(x, y)$ thus becomes

$$\begin{aligned} 4x^3 - 4xy + 4x &= 0, \\ -2x^2 + 4y + 4 &= 0. \end{aligned}$$

The first equation can be rewritten as

$$4x(x^2 - y + 1) = 0.$$

From this we see that either $x = 0$ or $x^2 - y + 4 = 0$ needs to hold.

- We first consider the case $x = 0$. In this case, the second equation reduces to $4y + 4 = 0$, which implies that $y = -1$. Thus we obtain a critical point $(x, y) = (0, -1)$.
- Now we consider the case $x^2 - y + 1 = 0$, or $x^2 = y - 1$. Inserting this in the second equation, we obtain that $-2(y-1) + 4y + 4 = 0$, or $2y + 6 = 0$. Thus $y = -3$, which would imply that $x^2 = y - 1 = -4$. Since this is impossible, we have no solution in this case.

In total, the only critical point of $f$ is the point $(x, y) = (0, -1)$.

We note now that we can write

$$f(x, y) = (x^2 - y)^2 + (y + 2)^2 + 2x^2 - 7,$$

which is clearly a coercive function (we can for instance estimate $f(x, y) \geq (y+2)^2 + 2x^2 - 7$). Thus the problem $\min_{x,y} f(x, y)$ attains a global minimum. Since $(0, -1)$ is the only candidate for a minimum, it follows that this point is indeed the unique global and local minimum of $f$.

b) The Hessian of $f$ is

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

For $(x, y) = (0, 2)$, the Hessian is clearly indefinite (it is a diagonal matrix with diagonal entries $-4$ and $+4$). Thus $f$ is not convex.

c) At the point $(1, -1)$ we have

$$\nabla f(1, -1) = \begin{pmatrix} 12 \\ -2 \end{pmatrix}$$

and

$$H_f(1, -1) = \begin{pmatrix} 20 & -4 \\ -4 & 4 \end{pmatrix}.$$

The search direction for the line search is therefore

$$p = -H_f(1, -1)^{-1} \nabla f(1, -1) = -\begin{pmatrix} 20 & -4 \\ -4 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 12 \\ -2 \end{pmatrix} = -\frac{1}{64} \begin{pmatrix} 4 & 4 \\ 4 & 20 \end{pmatrix} \begin{pmatrix} 12 \\ -2 \end{pmatrix} = -\frac{1}{64} \begin{pmatrix} 40 \\ 8 \end{pmatrix} = -\frac{1}{8} \begin{pmatrix} 5 \\ 1 \end{pmatrix}.$$

Moreover,

$$\langle p, \nabla f(1, -1) \rangle = -\frac{29}{4}$$

and $f(1, -1) = 0$. The Armijo condition with $\alpha = 1$ and $c_1 = 0.1$ thus requires that

$$f(3/8, -9/8) \leq -\frac{29}{40}.$$

We compute

$$f(3/8, -9/8) = \frac{81}{4096} - \frac{81}{256} + \frac{81}{64} + \frac{9}{32} - \frac{9}{2} - 3,$$

which is easily seen to be smaller than $-29/40$ (the value is approximately $-4.35$). Thus the Armijo condition is satisfied, and the Newton step is accepted.

**Problem 2.**

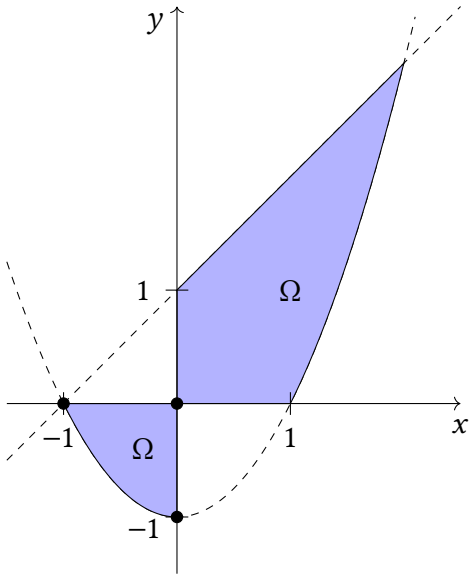
Consider the optimisation problem

$$\min_{x,y} 2x^2 - y^2$$

subject to the constraint $(x, y) \in \Omega$, where $\Omega \subset \mathbb{R}^2$ is given by the inequalities

$$xy \geq 0, \qquad y \geq x^2 - 1, \qquad \text{and} \qquad x \geq y - 1.$$

Here is a sketch of the set $\Omega$:




a) Determine for each point in $\Omega$ whether the linear independence constraint qualification (LICQ) holds.
*(5 points)*

b) Determine the tangent cone $T_\Omega(x, y)$ and the cone of linearised feasible directions $\mathcal{F}(x, y)$ to the set $\Omega$ in the points $(x, y) = (0, -1)$, $(x, y) = (0, 0)$, and $(x, y) = (-1, 0)$. (The points are marked in the sketch.)
*(10 points)*

c) Find all KKT points and all local minimisers for this optimisation problem.
*(15 points)*

**Solution.**

As a preparation, we define

$$f(x, y) = 2x^2 - y^2$$

and

$$c_1(x, y) = xy, \qquad c_2(x, y) = y - x^2 + 1, \qquad c_3(x, y) = x - y + 1,$$

and we compute

$$\nabla f(x, y) = \begin{pmatrix} 4x \\ -2y \end{pmatrix}, \quad \nabla c_1(x, y) = \begin{pmatrix} y \\ x \end{pmatrix}, \quad \nabla c_2(x, y) = \begin{pmatrix} -2x \\ 1 \end{pmatrix}, \quad \nabla c_3(x, y) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

a) We see immediately that $\nabla c_2(x, y) \neq 0$ and $\nabla c_3(x, y) \neq 0$ for all $x$, $y$. Moreover, the only point where $\nabla c_3(x, y) = 0$ is the point $(x, y) = (0, 0)$, which is feasible. Thus LICQ fails at $(0, 0)$, but holds at all other points where none or only a single constraint is active.

It remains to check the points where more than one constraint is active:

- At $(x, y) = (-1, 0)$, we have three active constraints, and thus LICQ fails.
- At $(x, y) = (0, -1)$, the constraints $c_1$ and $c_2$ are active, and $\nabla c_1(0, -1) = (-1, 0)$ and $\nabla c_2(0, -1) = (0, 1)$. These are obviously linearly independent and thus LICQ holds.
- At $(x, y) = (1, 0)$, the constraints $c_1$ and $c_2$ are active, and $\nabla c_1(1, 0) = (0, 1)$ and $\nabla c_2(1, 0) = (-2, 1)$. These are obviously linearly independent and thus LICQ holds.
- At $(x, y) = (0, 1)$, the constraints $c_1$ and $c_3$ are active, and $\nabla c_1(0, 1) = (1, 0)$ and $\nabla c_3(1, 0) = (1, -1)$. These are obviously linearly independent and thus LICQ holds.
- At $(x, y) = (2, 3)$, the constraints $c_2$ and $c_3$ are active, and $\nabla c_2(2, 3) = (-4, 6)$ and $\nabla c_3(2, 3) = (1, -1)$. Again, these are obviously linearly independent, and again LICQ holds.

To summarise, LICQ holds at every point except $(x, y) = (0, 0)$ and $(x, y) = (-1, 0)$.

b) At $(0, -1)$, the constraints $c_1$ and $c_2$ are active, and $\nabla c_1(0, -1) = (-1, 0)$ and $\nabla c_2(0, -1) = (0, 1)$. Since LICQ holds, we have

$$T_\Omega(0, -1) = \mathcal{F}(0, -1) = \big\{(p, q) \in \mathbb{R}^2 : p \leq 0 \text{ and } q \geq 0\big\}.$$

At $(0, 0)$, the constraint $c_1$ is active and $\nabla c_1(0, 0) = (0, 0)$. Thus

$$\mathcal{F}(0, 0) = \mathbb{R}^2.$$

Locally around $(0, 0)$, the feasible domain $\Omega$ consists of the two parts

$$\Omega_1 = \{(x, y) : x \geq 0 \text{ and } y \geq 0\} \quad \text{and} \quad \Omega_2 = \{(x, y) : x \leq 0 \text{ and } y \leq 0\}.$$

Thus

$$T_\Omega(0,0) = T_{\Omega_1}(0,0) \cup T_{\Omega_2}(0,0) = \{(p,q) : p \geq 0 \text{ and } q \geq 0\} \cup \{(p,q) : p \leq 0 \text{ and } q \leq 0\}.$$

At $(-1,0)$, all constraints are active and $\nabla c_1(-1,0) = (0,-1)$, $\nabla c_2(-1,0) = (-2,1)$, and $\nabla c_3(-1,0) = (1,-1)$. Thus

$$\mathcal{F}(-1,0) = \left\{(p,q) \in \mathbb{R}^2 : q \leq 0,\ 2p+q \geq 0,\ p-q \geq 0\right\}.$$

Note here that the last inequality $p - q \geq 0$ is superfluous (if $q \leq 0$ and $2p + q \geq 0$, then also $p = (2p+q)/2 - q/2 \geq 0$). Thus we can simplify this to

$$\mathcal{F}(-1,0) = \left\{(p,q) \in \mathbb{R}^2 : q \leq 0,\ 2p+q \geq 0\right\}.$$

Next we note that the constraint $c_3$, while active at $(-1,0)$, does not contribute to the local shape of $\Omega$. Thus it can be ignored for the computation of the tangent cone $T_\Omega(-1,0)$. As a consequence, we obtain that

$$T_\Omega(-1,0) = \mathcal{F}(-1,0) = \left\{(p,q) \in \mathbb{R}^2 : q \leq 0,\ 2p+q \geq 0\right\}.$$

c) The KKT conditions for this problem are

$$\begin{aligned}
4x &= \lambda y - 2\mu x + \nu, \\
-2y &= \lambda x + \mu - \nu, \\
xy &\geq 0, \\
y &\geq x^2 - 1, \\
x &\geq y - 1, \\
\lambda &\geq 0, \\
\mu &\geq 0, \\
\nu &\geq 0, \\
\lambda x y &= 0, \\
\mu(y - x^2 + 1) &= 0, \\
\nu(x - y + 1) &= 0.
\end{aligned}$$

For finding the KKT points of the problem, we consider all possible active sets:

- At $(-1,0)$, all constraints are active, and we obtain the conditions

  $$-4 = 2\mu + \nu \qquad \text{and} \qquad 0 = -\lambda + \mu - \nu.$$

  The first equation cannot be satisfied by non-negative Lagrange multipliers $\mu \geq 0$ and $\nu \geq 0$, and thus this is no KKT point.

- The constraints $c_1$ and $c_2$ are satisfied at the points $(0,-1)$ and $(1,0)$. At the point $(1,0)$ we obtain the conditions
$$4 = -2\mu \qquad \text{and } 0 = \lambda,$$
which cannot hold for non-negative Lagrange multipliers. Thus $(1,0)$ is no KKT point. At the point $(0,-1)$ we obtain the conditions
$$0 = -\lambda \qquad \text{and} \qquad 2 = \mu.$$
Thus $(0,-1)$ is a KKT point with Lagrange multipliers $\lambda = 0$ and $\mu = 2$.
- The constraints $c_1$ and $c_3$ are satisfied at the point $(0,1)$. Here we obtain the conditions
$$0 = \lambda + \nu \qquad \text{and} \qquad -2 = -\nu.$$
This implies that $\nu = 2$ and $\lambda = -2 < 0$, and thus $(0,1)$ is no KKT point.
- The constraints $c_2$ and $c_3$ are satisfied at the point $(2,3)$. Here we obtain the conditions
$$8 = -4\mu + \nu \qquad \text{and} \qquad -6 = \mu - \nu.$$
Summing these two conditions, we obtain that $2 = -3\mu$, which implies that $\mu < 0$. Thus $(2,3)$ is no KKT point.
- If only the constraint $c_1$ is active, we obtain the conditions
$$4x = \lambda y, \qquad -2y = \lambda x, \qquad \text{and} \qquad xy = 0.$$
The last condition implies that at least one of the equations $x = 0$ or $y = 0$ is satisfied, and now the first two conditions imply that the other equation is satisfied as well. Thus $(0,0)$ is a KKT point with Lagrange multiplier $\lambda = 0$.
- If only the constraint $c_2$ is active, we obtain the conditions
$$4x = -2\mu x, \qquad -2y = \mu, \qquad \text{and} \qquad y = x^2 - 1.$$
The first equation implies that either $x = 0$ or $\mu = -2$. The latter is impossible as $\mu$ needs to be non-negative, and the former implies that also $c_1$ is active, which we have already covered.
- If only the constraint $c_3$ is active, we obtain the conditions
$$4x = \nu, \qquad -2y = -\nu, \qquad \text{and} \qquad x = y - 1.$$
Thus
$$4x = \nu = 2y = 2x + 2,$$
and thus $x = 1$, $y = 2$, and $\nu = 4$. Thus $(1,2)$ is a KKT point with Lagrange multiplier $\nu = 4$.

- Finally, if no constraint is satisfied, we obtain the equations

$$4x = 0 \qquad \text{and} \qquad -2y = 0,$$

  which implies that $(x, y) = (0, 0)$, and thus, again, $c_1$ is satisfied.

To summarise, we have the following KKT points:

- The point $(0, 0)$, where $c_1$ is active and all Lagrange parameters are 0.
- The point $(1, 2)$, where $c_3$ is active and $\nu = 4$.
- The point $(0, -1)$, where $c_1$ and $c_2$ are active, and $\lambda = 0$ and $\mu = 2$.

The point $(0, 0)$ is clearly no local minimiser, as the function values decrease in both directions along the $y$-axis.

At the point $(-1, 0)$, LICQ fails, and thus it might be a possible candidate for a local minimiser. However, as $T_\Omega(-1, 0) = \mathcal{F}(-1, 0)$, this is not the case. (If the tangent cone and the cone of linearised feasible directions agree, the KKT conditions are necessary optimality conditions.)

For the two points $(1, 2)$ and $(0, -1)$, we consider the second order sufficient conditions. The Hessian of the Lagrangian is

$$H_{\mathcal{L}}(x, y; \lambda, \mu, \nu) = \begin{pmatrix} 4 + 2\mu & -\lambda \\ -\lambda & -2 \end{pmatrix}.$$

At the point $(1, 2)$, the constraint $c_3$ is active and $\nu > 0$. Thus the critical cone at $(1, 2)$ is

$$C(1, 2) = \big\{(p, q) \in \mathbb{R}^2 : \langle \nabla c_3(1, 2), (p, q)\rangle = 0\big\} = \big\{(p, q) \in \mathbb{R}^2 : p - q = 0\big\}.$$

Assume now that $(p, q) \in C(1, 2) \setminus \{(0, 0)\}$. Then $p = q \neq 0$ and thus

$$\langle (p, q), H_{\mathcal{L}}(1, 2; 0, 0, 4)(p, q)\rangle = \langle (p, p), (4p, -2p)\rangle = 2p^2 > 0.$$

Thus the second order sufficient condition is satisfied and thus the point $(1, 2)$ is a strict local minimum.

At the point $(0, -1)$, the constraints $c_1$ and $c_2$ are active, and $\lambda = 0$, $\mu = 2 > 0$. Thus the critical cone at $(0, 1)$ is

$$\begin{aligned} C(0, -1) = \big\{(p, q) \in \mathbb{R}^2 : \langle \nabla c_1(0, -1), (p, q)\rangle \geq 0 \text{ and } \langle \nabla c_2(0, -1), (p, q)\rangle = 0\big\} \\ = \big\{(p, q) \in \mathbb{R}^2 : p \leq 0 \text{ and } q = 0\big\}. \end{aligned}$$

Assume now that $(p, q) \in C(0, -1) \setminus \{(0, 0)\}$. Then $p > 0$ and $q = 0$ and thus

$$\langle (p, q), H_{\mathcal{L}}(0, -1; 1, 2; 0, 0, 4)(p, q)\rangle = \langle (p, 0), (6p, 0)\rangle = 6p^2 > 0.$$

Thus the second order sufficient condition is satisfied and thus the point $(0, -1)$ is a strict local minimum.

To summarise, we have the following KKT points and local minimisers:

- The KKT point $(0, 0)$, which is no local minimiser.
- The KKT point and local minimiser $(1, 2)$.
- The KKT point and local minimiser $(0, -1)$.

**Problem 3.**

Assume that $f_i \colon \mathbb{R}^d \to \mathbb{R}$, $i = 1, \dots, n$, are continuously differentiable, convex functions. Define the function $f \colon \mathbb{R}^d \to \mathbb{R}$,

$$f(x) := \max\{f_1(x), \dots, f_n(x)\}.$$

We consider the problem

$$\min_{x \in \mathbb{R}} f(x). \tag{1}$$

a) Show that the problem (1) is equivalent to the problem

$$\min_{(x,t) \in \mathbb{R}^d \times \mathbb{R}} t \qquad \text{s.t. } t \geq f_i(x) \qquad \text{for all } i = 1, \dots, n. \tag{2}$$

in the sense that $x^*$ solves (1), if and only if $(x^*, t^*)$ solves (2) and $t^* = f(x^*)$.

*(5 points)*

b) Use part a) to show that $x^* \in \mathbb{R}^d$ solves (1), if and only if there exist $\lambda_i^* \geq 0$, $i \in A(x^*)$, such that

$$\sum_{i \in A(x^*)} \lambda_i^* \nabla f_i(x^*) = 0 \qquad \text{and} \qquad \sum_{i \in A(x^*)} \lambda_i^* = 1,$$

where $A(x^*) := \{i : f_i(x^*) = f(x^*)\}$.

*(10 points)*

**Solution.**

a) Assume that $x^*$ solves (1). By construction,

$$f(x^*) = \max_i f_i(x^*) \geq f_j(x^*) \qquad \text{for all } j = 1, \dots, d,$$

and thus $(x^*, f(x^*))$ is feasible for (2). Assume now that $(x,t) \in \mathbb{R}^d \times \mathbb{R}$ are such that $f_i(x) \geq t$ for all $i$. Then

$$f(x^*) \leq f(x) = \max_i f_i(x) \leq t,$$

which shows that $(x^*, f(x^*))$ solves (2).

Conversely, assume that $(x^*, t^*)$ solves (2). Then clearly $t^* = f(x^*)$; else we could decrease the function value further. Now let $x \in \mathbb{R}^d$ be arbitrary. Then $(x, f(x))$ is feasible for (2) and thus

$$f(x) \geq f(x^*).$$

This shows that $x^*$ solves (1).

b) The problem (2) is a convex optimisation problem with concave constraints (the function $(x,t) \mapsto t - f_i(x)$ are concave). Moreover, Slater's constraint qualification is satisfied (e.g. by the point $(0,t)$ with $t = f(0)+1$, as clearly $f(0)+1 > f_i(0)$ for all $i$). Thus the KKT conditions are necessary and sufficent optimality conditions for (1). In order to formulate the optimality condition, we define the functions

$$\begin{aligned} g\colon \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}, \qquad & c_i\colon \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}, \\ g(x,t) = t, \qquad & c_i(x,t) = t - f_i(x). \end{aligned}$$

Then the KKT conditions read

$$\begin{aligned} \nabla g(x,t) &= \sum_i \lambda_i \nabla c_i(x,t), \\ t &\geq f_i(x), \\ \lambda_i &\geq 0, \\ \lambda_i &= 0 \qquad \text{if } t > f_i(x). \end{aligned}$$

Now the first equation can be decomposed into the two parts (by separating derivation w.r.t. the $x$ and $t$ variables)

$$\begin{aligned} 0 &= -\sum_i \lambda_i \nabla f_i(x), \\ 1 &= \sum_i \lambda_i. \end{aligned}$$

From the first part of the problem, we know that every solution $(x^*, t^*)$ satisfies $t^* = f(x^*)$. Thus the active set for such a solution is $A(x^*) = \{i : f_i(x^*) = f(x^*)\}$. Now the complementarity condition implies that any Lagrange multiplier for this problem satisfies $\lambda_i = 0$ whenever $i \notin A(x^*)$, and thus the summations over the $\lambda_i$'s need to be taken only for $i \in A(x^*)$. Thus

the KKT conditions are equivalent to the conditions

$$\begin{aligned} \sum_{i\in A(x)} \lambda_i \nabla f_i(x) &= 0, \\ \sum_{i\in A(x)} \lambda_i &= 1, \\ \lambda_i &\geq 0, \qquad i \in A(x), \end{aligned}$$

where $A(x) = \{i : f_i(x) = f(x)\}$, which was to show.

**Problem 4.**

Let $d \gg 1$ and let $b \in \mathbb{R}^d$. Define the function $f\colon \mathbb{R}^d \to \mathbb{R}$,

$$f(x) := \sum_{i=1}^{d} (x_i - b_i)^2 + \sum_{i=1}^{d-1} \sqrt{1 + (x_{i+1} - x_i)^2}.$$

Propose a numerical method for the solution of the problem

$$\min_{x\in\mathbb{R}^d} f(x).$$

Justify your proposal!

*(10 points)*

**Solution.**

The function $f$ is strictly convex, and all eigenvalues of the Hessian of $f$ are uniformly bounded below by 2. Thus Newton's method (with line search) converges quadratically to the (unique) global minimum of $f$. Moreover, the Newton steps can be easily computed, as the Hessian of $f$ is a tridiagonal matrix. Thus Newton's method with some cheap line search (e.g. backtracking Armijo line search) would probably be the method of choice.

**Problem 5.**

Let $c \in \mathbb{R}^d$, $A \in \mathbb{R}^{m\times d}$, and $b \in \mathbb{R}^m$, and consider the constrained optimisation problem

$$\min_x \frac{1}{2}\|x\|_2^2 + \langle c, x\rangle \qquad \text{s.t. } Ax \geq b. \tag{3}$$

Find the Lagrangian dual of the problem (3).

*(10 points)*

**Solution.**

The Lagrangian of (3) is the function $\mathcal{L}\colon \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$,

$$\mathcal{L}(x,\lambda) = \frac{1}{2}\|x\|_2^2 + \langle c, x\rangle - \langle \lambda, Ax - b\rangle.$$

The dual objective function is thus

$$q(\lambda) := \min_x \Big( \frac{1}{2}\|x\|_2^2 + \langle c, x\rangle - \langle \lambda, Ax - b\rangle \Big).$$

For fixed $\lambda \in \mathbb{R}^m$, the solution of this (quadratic, strongly convex) optimisation problem is attained at the point $\hat{x} \in \mathbb{R}^d$ satisfying

$$\hat{x} + c - A^T\lambda = 0.$$

Thus

$$q(\lambda) = \mathcal{L}(\hat{x},\lambda) = \frac{1}{2}\|A^T\lambda - c\|_2^2 + \langle c, A^T\lambda - c\rangle - \langle \lambda, A(A^T\lambda - c) - b\rangle = -\frac{1}{2}\|A^T\lambda - c\|^2 + \langle \lambda, b\rangle.$$

The dual problem is therefore

$$\max_\lambda -\frac{1}{2}\|A^T\lambda - c\|^2 + \langle \lambda, b\rangle \qquad \text{s.t. } \lambda \geq 0.$$

**Problem 6.**

Consider the Fletcher–Reeves non-linear CG method for the solution of the optimisation problem $\min_x f(x)$. Assume that the step length $\alpha_k > 0$ in each step is chosen according to an *exact* line search. Show that $p_k$ is in each step a descent direction provided that $\nabla f(x_k) \neq 0$.

*(10 points)*

**Solution.**

In the first step of a non-linear CG method, we start with $p_0 = -\nabla f(x_0)$, which is a descent direction (unless $\nabla f(x_0) = 0$).

Now consider a step $k \geq 0$. Because the step length $\alpha_k > 0$ is chosen according to an exact line search, it follows that $\langle \nabla f(x_k + \alpha_k p_k), p_k\rangle = 0$. With $x_k + \alpha_k p_k = x_{k+1}$, this implies that

$$\langle \nabla f(x_{k+1}), p_k\rangle = 0.$$

Thus

$$\langle \nabla f(x_{k+1}, p_{k+1} \rangle = \langle \nabla f(x_{k+1}, -\nabla f(x_{k+1}) + \beta_{k+1} p_k \rangle = -\|\nabla f(x_{k+1})\|_2^2 + \beta_{k+1} \langle \nabla f(x_{k+1}), p_k \rangle = -\|\nabla f(x_{k+1})\|_2^2 < 0$$

unless $\nabla f(x_{k+1}) = 0$.