1 Consider the problem

$$\min f(\mathbf{x}) \quad \text{subject to} \quad \mathbf{x} \in \Omega,$$

where $f(\mathbf{x}) = 2x_1 + 3$ and $\Omega = \{\mathbf{x} \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \geq 1\}$.

a. Find all point(s) satisfying the FONC.

b. Which of the point(s) in part (a) satisfy the SONC?

c. Which of the point(s) in part (a) are local minimizers?

Solution a. We need to consider the points on the interior of Ω and the boundary separately.
First, note that

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \quad \text{and} \quad H_f(\mathbf{x}) = 0$$

Boundary of Ω:

We want the points such that $\nabla f(\mathbf{x}) \cdot \mathbf{d} \geq 0$ for all feasible $\mathbf{d}$. If we take $\mathbf{d}$ to be a unit vector, we can write $\mathbf{d} = (\cos\theta, \sin\theta)^\top$. The inner product is

$$\nabla f(\mathbf{x}) \cdot \mathbf{d} = 2\cos\theta$$

which is greater than or equal to 0 whenever $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$. Since the boundary is a circle, the only point whose feasible directions satisfy that condition is $(1,0)^\top$.

Interior of Ω

We want the points such that $\nabla f(\mathbf{x}) \cdot \mathbf{d} = 0$. Taking $\mathbf{d}$ to be a unit vector as above gives the inner product $2\cos\theta$, which is 0 if $\theta \in \{\frac{\pi}{2}, -\frac{\pi}{2}\}$. But in the interior, all $\mathbf{d}$ are feasible directions, so the FONC cannot be satisfied in the interior of Ω.

b. The only point we need to consider is $(1,0)^\top$. $H_f(\mathbf{x}) = 0 \succeq 0$, so the SONC is satisfied.

c. $(1,0)^\top$ is not a local minimizer. If we take any neighborhood centered at $(1,0)^\top$, it would include a point that's above and to the left of it. Points to the left (i.e., points with $x_1 < 1$) will have a smaller value of f since $\frac{\partial f}{\partial x_1} = 2 > 0$. Thus, $(0,1)^\top$ cannot be a local minimizer.

---

2 Consider the unconstrained optimization problem

$$\min f(\mathbf{x}) = \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|^2,$$

where $A \in \mathbb{R}^{n\times n}$, $m \geq n$, and $\mathbf{b} \in \mathbb{R}^n$.

a. Show that $f(\mathbf{x})$ is a quadratic function of the form $\frac{1}{2}\mathbf{x}^\top Q\mathbf{x} - \mathbf{p}^\top\mathbf{x} + c$ by specifying Q, $\mathbf{p}$, and c.

b. Find the gradient $\nabla f(\mathbf{x})$ and Hessian matrix $H_f(\mathbf{x})$.

c. Suppose $A = \begin{pmatrix} 5 & 4 \\ 0 & 3 \end{pmatrix}$. Find the upper bound for α such that gradient descent with the fixed step size α converges to the solution.

Solution a. By the definition of an inner product, $\|A\mathbf{x} - \mathbf{b}\|^2 = (A\mathbf{x} - \mathbf{b}) \cdot (A\mathbf{x} - \mathbf{b})$. Thus,

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{2}(A\mathbf{x} - \mathbf{b}) \cdot (A\mathbf{x} - \mathbf{b}) \\ &= \frac{1}{2}(\mathbf{x}^\top A^\top A\mathbf{x} - 2\mathbf{b}^\top A\mathbf{x} + \|\mathbf{b}\|)^2 \\ &= \frac{1}{2}\mathbf{x}^\top(A^\top A)\mathbf{x} - (A^\top\mathbf{b})^\top\mathbf{x} + \frac{1}{2}\|\mathbf{b}\|^2. \end{aligned}$$

Comparing with a quadratic form, we see that

$$\begin{aligned} Q &= A^\top A \\ \mathbf{p} &= A^\top\mathbf{b} \\ c &= \frac{1}{2}\|\mathbf{b}\|^2 \end{aligned}$$

b. Note that Q is symmetric, so

$$\begin{aligned} \nabla f(\mathbf{x}) &= Q\mathbf{x} - \mathbf{p} = A^\top A\mathbf{x} - A^\top \mathbf{b} = A^\top (A\mathbf{x} - \mathbf{b}) \\ H_f(\mathbf{x}) &= Q\mathbf{x} = A^\top A \end{aligned}$$

c. By a theorem, we need $\alpha \in (0, \frac{2}{\lambda_{\max}})$. By inspection, the eigenvalues of $A^\top A$ are $\lambda_1 = 25$ and $\lambda_2 = 9$. Hence, the upper bound for α is $\frac{2}{25}$.

---

3 Let $(x_1, y_1)^\top, \ldots, (x_n, y_n)^\top$, $n \geq 2$ be points on the $\mathbb{R}^2$ plane. We wish to find the straight line of "best fit" through these points ("best in the sense that the average squared error is minimized); that is, we wish to find $a, b \in \mathbb{R}$ to minimize

$$f(a,b) = \frac{1}{n}\sum_{i=1}^{n}(ax_i + b - y_i)^2.$$

a. Let

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \qquad \overline{XY} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$$
$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \overline{Y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2$$

Show that $f(a, b)$ can be written in the form $\mathbf{z}^\top Q\mathbf{z} - 2\mathbf{x}^\top \mathbf{z} + d$, where $\mathbf{z} = (a, b)^\top$, $Q = Q^\top \in \mathbb{R}^{2\times 2}$, $\mathbf{x} \in \mathbb{R}^2$ and $d \in \mathbb{R}$, and find expressions for Q, $\mathbf{c}$, and d in terms of $\overline{X}, \overline{Y}, \overline{X^2}, \overline{Y^2}$, and $\overline{XY}$.

b. Assume that the x_i, $i = 1, \ldots, n$ are not all equal. Find the parameters a^* and b^* for the line of best fit in terms of $\overline{X}, \overline{Y}, \overline{X^2}, \overline{Y^2}$, and $\overline{XY}$. Show that the point $(a^*, b^*)^\top$ is the only local minimizer of f.
Hint: $\overline{X^2} - (\overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2$.

c. Show that if a^* and b^* are the parameters of the line of best fit, then $\overline{Y} = a^*\overline{X} + b^*$ (and hence once we have computed a^*, we can compute b^* using the formula $b^* = \overline{Y} - a^*\overline{X}$).

Solution a. Expanding $f(a, b)$,

$$\begin{aligned} f(a,b) &= \frac{1}{n}\sum_{i=1}^{n}(ax_i + b - y_i)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(a^2x_i^2 + abx_i - ax_iy_i + abx_i + b^2 - by_i - ax_iy_i - by_i + y_i^2) \\ &= \frac{1}{n}\sum_{i=1}^{n}(a^2x_i^2 + b^2 + 2abx_i - 2ax_iy_i - 2by_i + y_i^2) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{z}^\top \begin{pmatrix} x_i^2 & x_i \\ x_i & 1 \end{pmatrix}\mathbf{z} - 2(x_iy_i, y_i)^\top \mathbf{z} + y_i^2\right) \\ &= \mathbf{z}^\top \begin{pmatrix} \overline{X^2} & \overline{X} \\ \overline{X} & 1 \end{pmatrix}\mathbf{z} - 2\left(\overline{XY}, \overline{Y}\right)^\top \mathbf{z} + \overline{Y^2} \end{aligned}$$

Thus, by comparison,

$$\begin{aligned} Q &= \begin{pmatrix} \overline{X^2} & \overline{X} \\ \overline{X} & 1 \end{pmatrix} \\ \mathbf{x} &= \left(\overline{XY}, \overline{Y}\right) \\ d &= \overline{Y^2} \end{aligned}$$

b. We first find the points that satisfy the FONC. Since our feasible set is $\mathbb{R}^2$, we only need to consider the interior case. I.e., we want $\nabla f(\mathbf{z}^*) = 0$.

$$
\begin{aligned}
\nabla f(\mathbf{z}^*) &= 2Q\mathbf{z}^* - 2\mathbf{x} \\
&= 2\begin{pmatrix} \overline{X^2} & \overline{X} \\ \overline{X} & 1 \end{pmatrix}\begin{pmatrix} a^* \\ b^* \end{pmatrix} - 2\begin{pmatrix} \overline{XY} \\ \overline{Y} \end{pmatrix} \\
&= \begin{pmatrix} a^*\overline{X^2} + b^*\overline{X} \\ a^*\overline{X} + b^* \end{pmatrix} - \begin{pmatrix} \overline{XY} \\ \overline{Y} \end{pmatrix} = 0 \\
\implies \begin{pmatrix} a^* \\ b^* \end{pmatrix} &= \begin{pmatrix} \frac{\overline{XY} - (\overline{X})(\overline{Y})}{\overline{X^2} - (\overline{X})^2} \\ \overline{Y} - a^*\overline{X} \end{pmatrix}
\end{aligned}
$$

This is the only point that satisfies the FONC.

Next, we check the SONC. We want $Q \succeq 0$, so it suffices to show that the eigenvalues of Q are non-negative.

$$
\begin{aligned}
\lambda_1 + \lambda_2 &= \operatorname{tr} Q = \overline{X^2} + 1 \\
\lambda_1 \lambda_2 &= \det Q = \overline{X^2} - (\overline{X})^2
\end{aligned}
$$

$\operatorname{tr} Q > 0$ since $\overline{X^2}$ involves a sum of non-negative terms. Additionally, $\det Q \geq 0$ since by the hint, it is also a sum of non-negative terms. Moreover, since not all the x_i are equal, there exists x_i such that $x_i \neq \overline{X}$. Otherwise, we would have $x_i = \overline{X}$ for all i, which is a contradiction. Thus, $\det Q > 0$, so the eigenvalues of Q are both positive.

Since Q is positive definite, this implies that the point satisfying the FONC is the unique strict local minimizer of $f(a, b)$ in $\mathbb{R}^2$.

c. From (b), we have that $b^* = \overline{Y} - a^*\overline{X}$ as desired.

---

4 Consider the function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = \frac{1}{2}(x - c)^2$, $c \in \mathbb{R}$. We are interested in computing the minimizer of f using the iterative algorithm

$$x^{(k+1)} = x^{(k)} - \alpha_k f'(x^{(k)}),$$

where f' is the derivative of f and α_k is a step size satisfying $0 < \alpha_k < 1$.

a. Derive a formula relating $f(x^{(k+1)})$ with $f(x^{(k)})$, involving α_k.

b. Show that the algorithm is globally convergent if and only if

$$\sum_{k=0}^{\infty} \alpha_k = \infty.$$

Hint: Use part (a) and the fact that for any sequence $\{a_k\}_{k \geq 1} \subseteq (0, 1)$, we have

$$\prod_{k=0}^{\infty} (1 - \alpha_k) = 0 \iff \sum_{k=0}^{\infty} \alpha_k = \infty.$$

Solution a. Note that $f'(x) = x - c$, so $f'(x^{(k)}) = x^{(k)} - c$. Thus,

$$
\begin{aligned}
f(x^{(k+1)}) &= \frac{1}{2}(x^{(k+1)} - c)^2 \\
&= \frac{1}{2}\Big(x^{(k)} - \alpha_k(x^{(k)} - c) - c\Big)^2 \\
&= \frac{1}{2}\Big[(x^{(k)} - c)(1 - \alpha_k)\Big]^2 \\
&= f(x^{(k)})(1 - \alpha_k)^2
\end{aligned}
$$

b. First note that the only point satisfying the FONC is $x = c$. This is because the problem is unconstrained, so we need $f'(x) = 0 \implies x = c$.

"$\Longleftarrow$"

Suppose $\prod_{k=0}^{\infty}(1-\alpha_k) = 0$. Then

$$\begin{aligned}\prod_{k=0}^{\infty}(1-\alpha_k) = 0 &\implies \sum_{k=0}^{\infty}\log(1-\alpha_k) = -\infty \\ &\implies \sum_{k=0}^{\infty}2\log(1-\alpha_k) = \sum_{k=0}^{\infty}\log(1-\alpha_k)^2 = -\infty \\ &\implies \prod_{k=0}^{\infty}(1-\alpha_k)^2 = 0\end{aligned}$$

Hence,

$$f(x^{(k)}) = \prod_{i=0}^{k}(1-\alpha_i)^2 \xrightarrow{k\to\infty} 0 \implies x^{(k)} \xrightarrow{k\to\infty} c$$

for all $x^{(0)}$ since f is continuous. So, the algorithm is globally convergent.

"$\Longrightarrow$"

Suppose the algorithm is globally convergent, i.e., $x^{(k)} \xrightarrow{k\to\infty} c \implies f(x^{(k)}) \xrightarrow{k\to\infty} 0$. Then

$$\begin{aligned}\lim_{k\to\infty} f(x^{(k+1)}) = 0 &\implies \lim_{k\to\infty}\prod_{i=0}^{k}(1-\alpha_i)^2 = 0 \\ &\implies \sum_{k=0}^{\infty}2\log(1-\alpha_k) = -\infty \\ &\implies \sum_{k=0}^{\infty}\log(1-\alpha_k) = -\infty \\ &\implies \prod_{k=0}^{\infty}(1-\alpha_k) = 0 \\ &\implies \sum_{k=0}^{\infty}\alpha_k = \infty\end{aligned}$$

as desired.

---

5 Consider a function $f\colon \mathbb{R}^n \to \mathbb{R}^n$ given by $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ where $A \in \mathbb{R}^{n\times n}$ and $\mathbf{b} \in \mathbb{R}^n$. Suppose that A is invertible and $\mathbf{x}^*$ is the zero of f (i.e., $f(\mathbf{x}^*) = 0$). We wish to compute $\mathbf{x}^*$ using the iterative algorithm

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha f(\mathbf{x}^{(k)}),$$

where $\alpha \in \mathbb{R}$, $\alpha > 0$. We say that the algorithm is *globally monotone* if for any $\mathbf{x}^{(0)}$, $\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\| \leq \|\mathbf{x}^{(k)} - \mathbf{x}^*\|$ for all k.

a. Assume that all the eigenvalues of A are real. Show that a necessary condition for the algorithm above to be *globally monotone* is that all the eigenvalues of A are nonnegative.

   Hint: Use contraposition.

b. Suppose that

$$A = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}.$$

   Find the largest range of values of α for which the algorithm is *globally convergent* (i.e., $\mathbf{x}^{(k)} \to \mathbf{x}^*$ for all $\mathbf{x}^{(0)}$).

Solution a. Since $\mathbf{x}^*$ is a zero of f and A is invertible, we have that $\mathbf{x}^* = -A^{-1}\mathbf{b}$.

Suppose the eigenvalues of A are negative, so that A is negative definite. Choose $\mathbf{x}^{(0)} = \mathbf{x}^* + \mathbf{v}$, where $\mathbf{v}$ is an eigenvector of A with eigenvalue $\lambda < 0$. Then $A\mathbf{v} = \lambda\mathbf{v}$ and $A^{-1}\mathbf{v} = \frac{1}{\lambda}\mathbf{v}$. So,

$$\begin{aligned}
\|\mathbf{x}^{(0)} - \mathbf{x}^*\| &= \|\mathbf{v}\| \\
\|\mathbf{x}^{(1)} - \mathbf{x}^*\| &= \|\mathbf{x}^{(0)} - \alpha\lambda\mathbf{v} - \mathbf{x}^*\| \\
&= \|\mathbf{v} - \alpha\lambda\mathbf{v}\| \\
&= \|\mathbf{v}(1 - \alpha\lambda)\|
\end{aligned}$$

Since $\lambda < 0$, $-\alpha\lambda > 0 \implies 1 - \alpha\lambda > 1$ for all $\alpha > 0$. Hence,

$$\|\mathbf{v}(1 - \alpha\lambda)\| > \|\mathbf{v}\| \implies \|\mathbf{x}^{(0)} - \mathbf{x}^*\| > \|\mathbf{x}^{(0)} - \mathbf{x}^*\|$$

so the algorithm is not globally monotone. Hence, if the algorithm is globally monotone, all of its eigenvalues must be non-negative.

b. Notice that the eigenvalues of A are $\lambda = 1$ and $\lambda = 5$. Thus, by (a), the algorithm is globally monotone. Since we have two distinct eigenvalues in $\mathbb{R}^2$, we can form an eigenbasis. In particular, its eigenvectors are $\mathbf{v}_1 = (1,1)^\top$ and $\mathbf{v}_2 = (1,-1)^\top$.

Fix $\mathbf{x}^{(0)}$. Then there exist $c_1, c_2 \in \mathbb{R}$ such that $\mathbf{x}^{(0)} = \mathbf{x}^* + c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \coloneqq \mathbf{x}^* + \mathbf{v}$. We claim that $\mathbf{x}^{(k)} = \mathbf{x}^* + (I - \alpha A)^k\mathbf{v}$.

Base step:

This is clearly true for $k = 0$.

Inductive step:

Suppose $\mathbf{x}^{(k)}$ can be written as the above. Then we wish to show that $\mathbf{x}^{(k+1)}$ can be also. By definition,

$$\begin{aligned}
\mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} - \alpha(A\mathbf{x}^{(k)} + b) \\
&= \mathbf{x}^* + (I - \alpha A)^k\mathbf{v} - \alpha(A(\mathbf{x}^* + (I - \alpha A)^k\mathbf{v}) + b) \\
&= \mathbf{x}^* + (I - \alpha A)^k\mathbf{v} - \alpha\Big((A\mathbf{x}^* + b) + A(I - \alpha A)^k\mathbf{v}\Big) \\
&= \mathbf{x}^* + (I - \alpha A)^k\mathbf{v} - \alpha A(I - \alpha A)^k\mathbf{v} \\
&= \mathbf{x}^* + (I - \alpha A)(I - \alpha A)^k\mathbf{v} \\
&= \mathbf{x}^* + (I - \alpha A)^{k+1}\mathbf{v}
\end{aligned}$$

Thus, the formula holds by induction.

Next, we want to find α so that $\|\mathbf{x}^{(k)} - \mathbf{x}^*\| \xrightarrow{k\to\infty} 0$, which implies that $\mathbf{x}^{(k)} \xrightarrow{k\to\infty} \mathbf{x}^*$.

$$\begin{aligned}
\|\mathbf{x}^{(k)} - \mathbf{x}^*\| &= \|(I - \alpha A)^{k+1}\mathbf{v}\| \\
&= \|(I - \alpha A)^{k+1}(c_1\mathbf{v}_1 + c_2\mathbf{v}_2)\| \\
&= \|c_1(1 - \alpha)^{k+1}\mathbf{v}_1 + c_2(1 - 5\alpha)^{k+1}\mathbf{v}_2\|
\end{aligned}$$

For this to go to 0, we need $|1-\alpha| < 1$ and $|1-5\alpha| < 1$. The values of α which satisfy this are $0 < \alpha < \frac{2}{5}$. If α steps outside this range, then the sequence of norms will not converge to 0.