1 Consider the least squares problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \coloneqq \|A\mathbf{x}-\mathbf{b}\|^2,$$

where $A \in \mathbb{R}^{m\times n}$ is NOT necessarily of full-rank. Then the FONC points can be non-unique. Show that any FONC point $\mathbf{x}^*$ is a global minimizer, i.e., $f(\mathbf{x}) \geq f(\mathbf{x}^*)$, $\forall \mathbf{x} \in \mathbb{R}^n$.

Solution Note that we can write $f(\mathbf{x})$ as

$$f(\mathbf{x}) = \mathbf{x}^\top(A^\top A)\mathbf{x} - 2(A^\top \mathbf{b})^\top \mathbf{x} + \|\mathbf{b}\|^2$$

which is a quadratic form.

Note that

$$0 \leq \|A\mathbf{x}\|^2 = \mathbf{x}^\top A^\top A\mathbf{x}$$

so $A^\top A \succeq 0$.

Thus, if $\mathbf{x}^*$ is an FONC point, it is a global minimizer since f is a quadratic form with a positive semi-definite matrix.

---

2 Let $A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, $m \geq n$, and $\operatorname{rank} A = n$. Consider the constrained optimization problem

$$\min \frac{1}{2}\mathbf{x}^\top\mathbf{x} - \mathbf{x}^\top\mathbf{b} \quad \text{s.t.} \quad \mathbf{x} \in \operatorname{im} A$$

where $\operatorname{im} A$ denotes the range of A. Derive an expression for the global minimizer of this problem in terms of A and $\mathbf{b}$.

Solution For any $\mathbf{x} \in \operatorname{im} A$, which is of full column rank, there exists a unique $\mathbf{y} \in \mathbb{R}^n$ such that $A\mathbf{y} = \mathbf{x}$. Thus, we can rewrite the objective function as

$$f(\mathbf{y}) = \frac{1}{2}\mathbf{y}^\top A^\top A\mathbf{y} - (A^\top\mathbf{b})^\top\mathbf{y}$$

with $\operatorname{rank}(A^\top A) = n$. Since $A^\top A \in \mathbb{R}^{n\times n}$, it is invertible.

As f is a quadratic form with a positive definite matrix, it has a unique global minimizer, which is the unique FONC point $\mathbf{y}^* = (A^\top A)^{-1}A^\top\mathbf{b}$. Thus, the global minimizer of our original problem is

$$\mathbf{x}^* = A\mathbf{y}^* = A(A^\top A)^{-1}A^\top\mathbf{b}.$$

---

3 Given $A \in \mathbb{R}^{m\times n}$, $m \geq n$, $\operatorname{rank} A = n$, and $\mathbf{b}_1, \ldots, \mathbf{b}_p \in \mathbb{R}^m$, consider the problem

$$\min\left(\|A\mathbf{x}-\mathbf{b}_1\|^2 + \|A\mathbf{x}-\mathbf{b}_2\|^2 + \cdots + \|A\mathbf{x}-\mathbf{b}_p\|^2\right).$$

Suppose that $\mathbf{x}_i^*$ is a solution to the problem

$$\min\|A\mathbf{x}-\mathbf{b}_i\|^2,$$

where $i = 1, \ldots, p$. Write the solution to the problem in terms of $\mathbf{x}_1^*, \ldots, \mathbf{x}_p^*$.

Solution As A is of full column rank, each $\mathbf{x}_i^*$ is the solution to a least squares problem, so

$$\mathbf{x}_i^* = (A^\top A)^{-1}A^\top\mathbf{b}_i.$$

Notice that we can write our objective function as

$$\begin{aligned} f(\mathbf{x}) &= \sum_{i=1}^{p}\left(\mathbf{x}^\top A^\top A\mathbf{x} - 2(A^\top\mathbf{b}_i)^\top\mathbf{x} + \|\mathbf{b}_i\|^2\right) \\ &= p\mathbf{x}^\top A^\top A\mathbf{x} - 2(A^\top\sum_{i=1}^{p}\mathbf{b}_i)^\top\mathbf{x} + \sum_{i=1}^{p}\|\mathbf{b}_i\|^2 \end{aligned}$$

which is a quadratic form with $Q = pA^\top A \succ 0$ and $\mathbf{b} = -2(A^\top \sum_{i=1}^p \mathbf{b}_i)^\top$. Hence, our global minimizer is our FONC point. So,

$$\begin{aligned}
\mathbf{0} = \nabla f(\mathbf{x}^*) &= 2pA^\top A\mathbf{x}^* - 2(A^\top \sum_{i=1}^p \mathbf{b}_i)^\top \\
\implies \mathbf{x}^* &= \frac{1}{p}(A^\top A)^{-1}A^\top \sum_{i=1}^p \mathbf{b}_i \\
\implies \mathbf{x}^* &= \frac{1}{p}\sum_{i=1}^p \mathbf{x}_i^*.
\end{aligned}$$

---

4 This problem derives the so-called *projected gradient descent* algorithm. Consider the following constrained problem:

$$\min_{x \in \mathbb{R}^n} f(\mathbf{x}) \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b},$$

where $A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, $m < n$, and $\operatorname{rank} A = m$.

a. Consider minimization of the following quadratic approximation to $f(\mathbf{x})$ around $\mathbf{x}^k$ without the constraint:

$$\mathbf{x}^{k+1} = \arg\min_{x \in \mathbb{R}^n} f(\mathbf{x}^k) + (\nabla f(\mathbf{x}^k))^\top(\mathbf{x} - \mathbf{x}^k) + \frac{1}{2\alpha_k}\|\mathbf{x} - \mathbf{x}^k\|^2$$

for some $\alpha_k > 0$. Show that $\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha_k \nabla f(\mathbf{x}^k)$.

b. For any $\mathbf{y} \in \mathbb{R}^n$, show that the problem

$$\min_{x \in \mathbb{R}^n} \|\mathbf{x} - \mathbf{y}\|^2 \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b}$$

has a unique solution given by $\mathbf{x}^* = \Pi(\mathbf{y})$, where Π is a linear function on $\mathbb{R}^n$ defined as

$$\Pi\colon \mathbf{x} \mapsto (I_n - A^\top(AA^\top)^{-1}A)\mathbf{x} + A^\top(AA^\top)^{-1}\mathbf{b}$$

with I_n being the identity matrix of order n.

c. Consider minimization of the following quadratic approximation to $f(\mathbf{x})$ around $\mathbf{x}^k$ under the same constraint:

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}^k) + (\nabla f(\mathbf{x}^k))^\top(\mathbf{x} - \mathbf{x}^k) + \frac{1}{2\alpha_k}\|\mathbf{x} - \mathbf{x}^k\|^2 \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b}$$

for some $\alpha_k > 0$. Show that $\mathbf{x}^{k+1} = \Pi(\mathbf{x}^k - \alpha_k \nabla f(\mathbf{x}^k))$. This gives the iteration for projected gradient descent.

Solution a. Note that if we let $\mathbf{y} = \mathbf{x} - \mathbf{x}^k$, then $f(\mathbf{y}) = f(\mathbf{x}^k) + (\nabla f(\mathbf{x}^k))^\top \mathbf{y} + \frac{1}{2\alpha_k}\|\mathbf{y}\|^2$ is a quadratic form with

$$Q = \frac{1}{\alpha_k} > 0 \quad \text{and} \quad \mathbf{b} = \nabla f(\mathbf{x}^k)$$

so its gradient is

$$\nabla g(\mathbf{x}) = \frac{1}{\alpha_k}\mathbf{x} + \nabla f(\mathbf{x}^k).$$

Note that

$$\nabla g(-\alpha_k \nabla f(\mathbf{x}^k)) = \mathbf{0}.$$

Since f is a quadratic form with a positive definite matrix, $\mathbf{y}^* = -\alpha_k \nabla f(\mathbf{x}^k)$. Hence,

$$\mathbf{y}^* = -\alpha_k \nabla f(\mathbf{x}^k) \implies \mathbf{x}^* = \mathbf{x}^k - \alpha_k \nabla f(\mathbf{x}^k),$$

as desired.

b. Consider the change in variable $\mathbf{z} = \mathbf{x} - \mathbf{y}$. Then the problem becomes

$$\min_{\mathbf{z}\in\mathbb{R}^n} \|\mathbf{z}\|^2 \quad \text{s.t.} \quad A\mathbf{z} = \mathbf{b} - A\mathbf{y}$$

which, since A is full rank, has the unique solution

$$\begin{aligned}
\mathbf{z}^* &= A^\top(AA^\top)^{-1}(\mathbf{b} - A\mathbf{y}) \\
\implies \mathbf{x}^* &= \mathbf{y} + A^\top(AA^\top)^{-1}(\mathbf{b} - A\mathbf{y}) \\
&= I_n\mathbf{y} - A^\top(AA^\top)^{-1}A\mathbf{y} + A^\top(AA^\top)^{-1}\mathbf{b} \\
&= (I_n - A^\top(AA^\top)^{-1}A)\mathbf{y} + A^\top(AA^\top)^{-1}\mathbf{b} \\
&= \Pi(\mathbf{y}).
\end{aligned}$$

c. For all $\mathbf{x} \in \mathbb{R}^n$, we can write $\mathbf{x} = \mathbf{y} + \mathbf{x}^k - \alpha_k\nabla f(\mathbf{x}^k)$ for some $\mathbf{y} \in \mathbb{R}^n$. Then we can rewrite the objective function as

$$\begin{aligned}
f(\mathbf{y}) &= f(\mathbf{x}^k) + (\nabla f(\mathbf{x}^k))^\top(\mathbf{y} - \alpha_k\nabla f(\mathbf{x}^k)) + \frac{1}{2\alpha_k}\|\mathbf{y} - \alpha_k\nabla f(\mathbf{x}^k)\|^2 \\
&= f(\mathbf{x}^k) + (\nabla f(\mathbf{x}^k))^\top(\mathbf{y}) + \frac{1}{2\alpha_k}\|\mathbf{y}\|^2 - (\nabla f(\mathbf{x}^k))^\top(\alpha_k\nabla f(\mathbf{x}^k)) + \frac{1}{2\alpha_k}\|\alpha_k\nabla f(\mathbf{x}^k)\|^2 - \frac{1}{\alpha_k}\alpha_k(\nabla f(\mathbf{x}^k))^\top\mathbf{y} \\
&= f(\mathbf{x}^k) + (\nabla f(\mathbf{x}^k))^\top(\mathbf{y}) + \frac{1}{2\alpha_k}\|\mathbf{y}\|^2 - \frac{1}{2\alpha_k}\|\alpha_k\nabla f(\mathbf{x}^k)\|^2 - (\nabla f(\mathbf{x}^k))^\top\mathbf{y} \quad (-\alpha_k\nabla f(\mathbf{x}^k) \text{ is the minimizer}) \\
&= f(\mathbf{x}^k) - \frac{1}{2}\alpha_k\|\nabla f(\mathbf{x}^k)\|^2 + \frac{1}{2\alpha_k}\|\mathbf{y}\|^2
\end{aligned}$$

which is strictly increasing with respect to $\|\mathbf{y}\|^2$. Hence, the closest point to $-\alpha_k\nabla f(\mathbf{x}^k)$ will minimize the function in this point, i.e., minimizing the problem in this part is equivalent to minimizing the distance from $\mathbf{x}$ to $-\alpha_k\nabla f(\mathbf{x}^k)$.

Thus, if we replace $\mathbf{y}$ in the equation in (b) with $-\alpha_k\nabla f(\mathbf{x}^k)$, we get

$$\mathbf{x}^* = \Pi(-\alpha_k\nabla f(\mathbf{x}^k)).$$

---

5 Convert the following problem into a standard form linear programming problem:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{c}^\top\mathbf{x} \quad \text{s.t.} \quad A\mathbf{x} \geq \mathbf{b}, \mathbf{x} \leq \mathbf{d},$$

where $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{c}, \mathbf{d} \in \mathbb{R}^n$.

Solution We can rewrite the constraints as $\mathbf{y}_1 \coloneqq \mathbf{d} - \mathbf{x} \geq \mathbf{0}$ and $\mathbf{y}_2 \coloneqq A\mathbf{x} - \mathbf{b} \geq \mathbf{0}$. Then note that

$$\begin{pmatrix} A & I_n \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = A\mathbf{d} - \mathbf{b}$$

and that $\mathbf{x} = \mathbf{d} - \mathbf{y}_1$, so our objective function becomes

$$\mathbf{c}^\top(\mathbf{d} - \mathbf{y}_1) = \mathbf{c}^\top\mathbf{d} + (-\mathbf{c}^\top\mathbf{y}_1) + \mathbf{0}^\top\mathbf{y}_2 = \mathbf{c}^\top\mathbf{d} + \begin{pmatrix} -\mathbf{c}^\top & \mathbf{0}^\top \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}.$$

Since the first term is a constant the problem is equivalent to minimizing the second term. Thus, our problem becomes

$$\min_{(\mathbf{y}_1,\mathbf{y}_2)^\top\in\mathbb{R}^{n+m}} \begin{pmatrix} -\mathbf{c}^\top & \mathbf{0}^\top \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \quad \text{s.t.} \quad \begin{pmatrix} A & I_n \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}, \ \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \geq 0$$

which is in standard form.