CI_logo.png

Optimization Methods **2024**

Student: Jeferson Morales Mariciano

Discussed with:

Assignment 4 **Due date:** Monday, 3 June 2024, 12:00 AM

---

1. Exercise (20/100)

Consider the quadratic function $f : \mathbb{R}^2 \to \mathbb{R}$ defined as:

$$f(\mathbf{x}) = 7x^2 + 4xy + y^2 \tag{1}$$

where $\mathbf{x} = (x, y)^T$.

1. Write this function in canonical form, i.e. $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$, where A is a symmetric matrix.
2. Describe briefly how the Conjugate Gradient (CG) Method works and discuss whether it is suitable to minimize f from equation 1. Explain your reasoning in detail (max. 30 lines).

1. Answer

The function written in canonical form correspond to:

$$\begin{aligned}
f(\mathbf{x}) &= 7x^2 + 4xy + y^2 \\
&= \begin{bmatrix} 7x + 2y & 2x + y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\
&= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 7 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\
&= \frac{1}{2} \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 14 & 4 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\
&= \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}
\end{aligned}$$

With $\mathbf{b} = \mathbf{0}$, $c = 0$, and A being clearly a symmetric matrix:

$$\mathbf{A} = \begin{bmatrix} 14 & 4 \\ 4 & 2 \end{bmatrix}$$

Let's verify if A is positive define as required by the quadratic form:

$$\det(\lambda I - \mathbf{A}) = \begin{vmatrix} \lambda - 14 & -4 \\ -4 & \lambda - 2 \end{vmatrix}$$
$$= \lambda^2 - 16 + 12$$
$$\Rightarrow \lambda_{1,2} = 8 \pm 2\sqrt{13} > 0$$

Finally, since all eigenvalues are positive, $\mathbf{A}$ is SPD.

2. Answer

The CG method is an iterative algorithm for solving a linear system of equations $Ax = b$ where $A \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix. whose A matrix is symmetric and **positive definite**. For (1) we proved that A is SPD, which already perfectly fits the requirements. The **iterativeness** of the method builds up a solution over a series of steps, each of which improves the approximation to the exact solution. The performance of the linear CG method is determined by the **distribution of the eigenvalues** of the coefficient matrix, which are two, so it is already a good candidate anyways, thus reaching immediately the exact solution before mentioned. The method exploits the **conjugate directions** of the matrix A, i.e. $\langle p_i, p_j \rangle = 0, \ i \neq j$. Such properties allows the method to **converge** in at most n iterations, where n is the dimension of the problem. The **convergence rate** depends on the distance between the current eigval and the largest as the formula suggests: $\|x_{k+1} - x^*\|_A^2 \leq \left(\frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1}\right)^2 \|x_0 - x^*\|_A^2$ if A has eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Particularly, the CG method allows to compute directions as a **linear combination** of the residual $r_{k+1} = Ax_{k+1} - b$ and the previous direction p_k, ensuring that the new search directions are conjugate with respect to A, thus significantly reducing the number of iterations needed to converge to the minimum when compared **to other gradient-based methods** like steepest descent. Such advantage becomes strongly evident especially in high-dimensional spaces. The CG method is appropriate and effective for minimizing the quadratic function. In fact, its use case perfectly matches the context of optimization for finding the minimum of **quadratic forms**.

2. Exercise (20/100)

Consider the following constrained minimization problem for $\mathbf{x} = (x, y, z)^T$

$$\begin{aligned} &\min_{\mathbf{x}} f(\mathbf{x}) := -3x^2 + y^2 + 2z^2 + 2(x + y + z) \\ &\text{subject to} \quad c(\mathbf{x}) = x^2 + y^2 + z^2 - 1 = 0 \end{aligned} \tag{2}$$

Write down the Lagrangian function and derive the KKT conditions for (2).

Answer

The constrained optimization problem can be written as:

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \quad \text{subject to} \quad \begin{cases} c_i(\mathbf{x}) = 0 & i \in \mathcal{E} \\ c_i(\mathbf{x}) \geq 0 & i \in \mathcal{I} \end{cases} \tag{3}$$

where the **objective function** f and the **constraint functions** on the variables c_i are all smooth and real-valued defined on a subset of $\mathbb{R}^n$. The problem defines two finite sets of indices: $\mathcal{I}$ for the **equality constraints** and $\mathcal{E}$ for the **inequality contraints**. In addition, the set of points $\mathbf{x}$ that satisfy the constraints is defined as the **feasible region** Ω:

$$\Omega = \{\mathbf{x} \mid c_i(\mathbf{x}) = 0, \ i \in \mathcal{E}; \ c_i(\mathbf{x}) \geq 0, \ i \in \mathcal{I}\} \tag{4}$$

Allowing to coincisely write the constrained optimization problem as:

$$\min_{\mathbf{x}\in\Omega} f(\mathbf{x}) \tag{5}$$

Then, the **active set** $\mathcal{A}(\mathbf{x})$ at any feasible $\mathbf{x}$ consists of the equality constraints indices from $\mathcal{E}$ together with the indices of the inequality constraints i for which $c_i(\mathbf{x}) = 0$:

$$\mathcal{A}(\mathbf{x}) = \mathcal{E} \cup \{i \in \mathcal{I} \mid c_i(\mathbf{x}) = 0\} \tag{6}$$

So, at a feasible point $\mathbf{x}$, the inequality constraint $i \in \mathcal{I}$ is said to be active if $c_i(\mathbf{x}) = 0$ and inactive if the strict inequality $c_i(\mathbf{x}) > 0$ is satisfied.
Assuming a single equality scenario part of the active set, at the solution $\mathbf{x}^*$, the constraint normal $\nabla c_1(\mathbf{x}^*)$ is parallel to $\nabla f(\mathbf{x}^*)$, meaning that there is a scalar λ_1^* called **Lagrangian multiplier** such that:

$$\nabla f(\mathbf{x}^*) = \lambda_1^* \nabla c_1(\mathbf{x}^*) \tag{7}$$

Finally, the **Langragian function** and its gradient for the general problem are defined as:

$$\begin{aligned}\mathcal{L}(\mathbf{x}, \lambda) &= f(\mathbf{x}) - \sum_{i\in\mathcal{E}\cup\mathcal{I}} \lambda_i c_i(\mathbf{x}) \\ \nabla_x \mathcal{L}(\mathbf{x}, \lambda) &= \nabla f(\mathbf{x}) - \sum_{i\in\mathcal{E}\cup\mathcal{I}} \lambda_i \nabla c_i(\mathbf{x})\end{aligned} \tag{8}$$

If assuming a single equality constraint scenario part of the active set, note that $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda_1) = \nabla f(\mathbf{x}) - \lambda_1 \nabla c_1(\mathbf{x})$, allowing to write the condition (7) equivalently as follows:

$$\text{at solution } \mathbf{x}^*,\ \exists\ \lambda_1^*:\ \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \lambda_1^*) = \mathbf{0} \tag{9}$$

This observation suggests that we can search for solutions of the equality-constrained problem (2) by seeking stationary points of the Lagrangian function. The conditions (7) and (9) are equivalent and necessary conditions for an optimal solution of the problem (2), but clearly not sufficient.

An important constraint qualification condition is **LICQ**: given the point $\mathbf{x}$ and the active set $\mathcal{A}(\mathbf{x})$ defined in (6), we say that the Linear Independence Contraint Qualification (LICQ) holds if the set of active constraint gradients $\{\nabla c_i(\mathbf{x}) \mid i \in \mathcal{A}(\mathbf{x})\}$ is linearly independent. In general, if LICQ holds, none of the active constraints gradients can be zero.

The following necessary conditions defined are called first-order conditions because they are concerned with properties of the gradients (first-derivative vectors) of the objective and constraint functions. **First-Order Necessary Conditions**: suppose that $\mathbf{x}^*$ is a local solution of (2), that the functions f and c_i in (2) are continuously differentiable, and that the LICQ holds at $\mathbf{x}^*$. Then there is a Lagrange multiplier vector λ^*, with components $\lambda_i^*,\ i \in \mathcal{E} \cup \mathcal{I}$, such that the following conditions are satisfied at $(\mathbf{x}^*, \lambda^*)$:

$$\begin{aligned}\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \lambda^*) &= \mathbf{0} \\ c_i(\mathbf{x}^*) &= \mathbf{0} \quad \forall i \in \mathcal{E} \\ c_i(\mathbf{x}^*) &\geq \mathbf{0} \quad \forall i \in \mathcal{I} \\ \lambda_i^* &\geq 0 \quad \forall i \in \mathcal{I} \\ \lambda_i^* c_i(\mathbf{x}^*) &= \mathbf{0} \quad \forall i \in \mathcal{E} \cup \mathcal{I} \quad \text{(complementary conditions)}\end{aligned} \tag{10}$$

The conditions (10) are often known as the Karush-Kuhn-Tucker conditions, or **KKT conditions** for short. The last row in (10) contains the complementary conditions, they imply that either constraint i is active or $\lambda_i^* = 0$, or possibly both. In particular, the Lagrange multipliers corresponding to inactive inequality constraints are zero, we can omit the terms for indices $i \notin \mathcal{A}(\mathbf{x}^*)$ from the first condition in (10) and rewrite it as:

$$\mathbf{0} = \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \lambda^*) = \nabla f(\mathbf{x}^*) - \sum_{i\in\mathcal{A}(\mathbf{x}^*)} \lambda_i^* \nabla c_i(\mathbf{x}^*) \tag{11}$$

The derived Lagrangian function for the problem (2) is:

$$\begin{aligned}
\mathcal{L}(\mathbf{x},\lambda) &= f(\mathbf{x}) - \sum_{i\in\mathcal{E}\cup\mathcal{I}} \lambda_i c_i(\mathbf{x}) \qquad \mathcal{E}=\{1\},\ \mathcal{I}=\emptyset \\
&= f(\mathbf{x}) - \lambda c(\mathbf{x}) \\
&= -3x^2 + y^2 + 2z^2 + 2(x+y+z) - \lambda\left(x^2+y^2+z^2-1\right) \\
&= (-3-\lambda)x^2 + (1-\lambda)y^2 + (2-\lambda)z^2 + 2(x+y+z) + \lambda
\end{aligned}$$

Let $\mathbf{x}^* = \left[x^*, y^*, z^*\right]^T$ be a local solution, then the derived KKT conditions are:

$$\begin{aligned}
\nabla f(\mathbf{x}^*) = \begin{bmatrix} -6x^*+2 \\ 2y^*+2 \\ 4z^*+2 \end{bmatrix}, &\quad \nabla c = \begin{bmatrix} 2x^* \\ 2y^* \\ 2z^* \end{bmatrix} \\
\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*,\lambda^*) = \mathbf{0} \Rightarrow \begin{bmatrix} -6x^*+2-2\lambda^*x^* \\ 2y^*+2-2\lambda^*y^* \\ 4z^*+2-2\lambda^*z^* \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\
c(\mathbf{x}^*) = 0 \Rightarrow (x^*)^2 + (y^*)^2 + (z^*)^2 - 1 &= 0 \\
\lambda^* c(\mathbf{x}^*) = 0 \Rightarrow \lambda^*\left((x^*)^2 + (y^*)^2 + (z^*)^2 - 1\right) &= 0
\end{aligned}$$

Note that if equality condition $c(\mathbf{x}^*) = 0$ holds, then condition $\lambda^* c(\mathbf{x}^*) = 0$ is also satisfied.

3. Exercise (60/100)

1. Read the chapter on Simplex method, in particular the section 13.3 The Simplex Method, in Numerical Optimization, Nocedal and Wright. Explain how the method works, with a particular attention to the search direction.

2. Consider the following contrained minimization problem, $\mathbf{x} = (x_1, x_2)^T$;

$$\min_{\mathbf{x}} f(\mathbf{x}) := 4x_1 + 3x_2 \tag{12}$$

subect to:

$$\begin{aligned}
6 - 2x_1 - 3x_2 &\geq 0 \\
3 + 3x_1 - 2x_2 &\geq 0 \\
5 - 2x_2 &\geq 0 \\
4 - 2x_1 - x_2 &\geq 0 \\
x_2 &\geq 0 \\
x_1 &\geq 0
\end{aligned} \tag{13}$$

   a) Sketch the feasible region for this problem.
   b) Which are the basic feasible points of the problem (12)? Compute them by hand using the geometrical interpretation and find the optimal point $\mathbf{x}^*$ that minimizes f subject to the constraints.
   c) Prove that the first order necessary conditions holds for the optimal point.

1. Answer

Before explaining the Simplex method, let's first introduce the concept of linear programming. Despite real-case situation models are often nonlinear, linear programming is appealing because of the advanced state of the software, guaranteed convergence to global minimum, and the fact that uncertainty in the model makes a linear model more appropriate than an overly complex nonlinear model. Linear programs have a **linear objective function** and **linear constraints**, which may include both equalities and inequalities. The feasible set is **polytope**: a convex, connected set with flat, polygonal faces. The contours of the objective function are planar. The solution can be unique e.g. a single vertex of the polytope, or non unique e.g. an entire edge. In higher dimensions, the set of optimal points can be a single vertex, an edge or face, or even the entire feasible set. **Infeasible case**: the problem has no solution if the feasible set is empty. **Unbounded case**: the problem has no solution if the objective function is unbounded below on the feasible region. Linear programs are analyzed in *standard form*:

$$\min c^T x \quad \text{subject to} \quad \begin{cases} Ax = b \\ x \geq 0 \end{cases} \tag{14}$$

where $c, x \in \mathbb{R}^n, \ b \in \mathbb{R}^m, \ A \in \mathbb{R}^{m \times n}$. For the standard formulation of (14), we will assume throughout that $m \leq n$ for simplicity. Simple devices can be used to transform any linear program to this form. For instance, given the problem (15) without any bounds on x:

$$\min c^T x \quad \text{subject to } Ax \leq b \tag{15}$$

we can convert the inequality constraints to equalities by introducing a vector of **slack variables** z and writing:

$$\min c^T x \quad \text{subject to} \begin{cases} Ax + z = b \\ z \geq 0 \end{cases} \tag{16}$$

It's not yet in standard form since not all the variables are constrained to be nonnegative. We deal with this by splitting x into its nonnegative and nonpositive parts:

$$x = x^+ - x^-, \quad x^+ = \max(x, 0) \geq 0, \quad x^- = \max(-x, 0) \geq 0$$

The problem (16) can be written as:

$$\min \begin{bmatrix} c \\ -c \\ 0 \end{bmatrix}^T \begin{bmatrix} x^+ \\ x^- \\ z \end{bmatrix}, \quad \text{subject to} \begin{bmatrix} A & -A & I \end{bmatrix} \begin{bmatrix} x^+ \\ x^- \\ z \end{bmatrix} = b, \quad \begin{bmatrix} x^+ \\ x^- \\ z \end{bmatrix} \geq 0$$

which clearly has the same form as (14). For inequality constraints of the form $Ax \geq b$, we subtract the variables from the left hand side as in (17), such variables are known as **surplus variables**.

$$Ax \geq b \Leftrightarrow Ax - y = b, \quad y \geq 0 \tag{17}$$

The optimality conditions for problem (14) are only the first-order conditions alias **KKT conditions**. The convexity of the problem ensures that these conditions are sufficient for a global minimum. The KKT conditions requires linear independence of the active constraint gradients (LICQ). However, the result continues to hold for dependent constraints provided they are linear, as is the case here.

We partition the Lagrange multipliers for the problem (14) into two vectors $\lambda \in \mathbb{R}^m$ multiplier vector for equality constraints $Ax = b$, and $s \in \mathbb{R}^n$ multiplier vector for the bound constraints $x \geq 0$. The **Lagrangian function** for the problem (14) is:

$$\mathcal{L}(x, \lambda, s) = c^T x + \lambda^T (Ax - b) - s^T x \tag{18}$$

Applying Theorem (10), we find that the first-order necessary conditions for x^* to be a solution of (14) are:

$$\begin{aligned}
\exists\, \lambda, s:\ A^T\lambda + s &= c,\\
Ax &= b,\\
x &\geq 0,\\
s &\geq 0,\\
x_i s_i &= 0,\ i = 1, 2, \ldots, n
\end{aligned} \tag{19}$$

The multipliers λ, s indicate the sensitivity of the optimal objective value to perturbations in the constraints and the process of finding (λ, s) for a given optimal x is called **sensitivity analysis**.

We assume for the remainder of the discussion that the matrix A in (14) has full row rank. Each iterate generated by the simplex method is a **basic feasible point** of (14). A vector x is a basic feasible point if it is feasible and if there exists a subset $\mathcal{B}$ of the index set $\{1, 2, \ldots, n\}$ such that:

- $\mathcal{B}$ contains exactly m indices
- $i \notin \mathcal{B} \Rightarrow x_i = 0$, meaning the bound $x_i \geq 0$ can be inactive only if $i \in \mathcal{B}$
- The $B \in \mathbb{R}^{m\times m},\ B = [A_i]_{i\in\mathcal{B}}$ matrix is nonsingular, where A_i is the ith column of A

A set $\mathcal{B}$ satisfying these properties is called a **basis** for the problem (14). The corresponding matrix B is called the **basis matrix**. The **simplex method** strategy is to examine only basic feasible points leading to the convergence of the solution of (14). Such strategy holds thanks to the following theorem:

Theorem 3.1 (Fundamental Theorem of Linear Programming)**.**

1. *If (14) has a nonempty feasible region, then there is at least one basic feasible point*
2. *If (14) has solutions, then at least one such solution is a basic optimal point*
3. *If (14) is feasible and bounded, then it has an optimal solution*

The feasible set defined by the linear constraints is a polytope, and the **vertices of the polytope** are the points that do not lie on a straght line between two other points in the set. Geometrically, they are visible as the corners of the polytope in Figure (1). Algebraically, the vertices are exactly the basic feasible points defined above.

Theorem 3.2. *All basic feasible points for (14) are vertices of the feasible polytope $\{x \mid Ax = b, x \geq 0\}$, and vice versa.*

The **Simplex method** explaination provided in the book is known as the **revised simplex method**. All iterates of the simplex method are basic feasible points for the problem (14), and therefore vertices of the feasible polytope. Most steps consist of a move from one vertex to an adjacent one for which the basis $\mathcal{B}$ differs in exactly one component. The major issue at each simplex iteration is to decide which index to remove from the basis $\mathcal{B}$. Unless the step is a **direction** of unboundedness, a single index must be removed from $\mathcal{B}$ and replaced by another from outside $\mathcal{B}$. More details regarding this exchange of index procedure is presented: from $\mathcal{B}$ and (19), we can derive values for not just.

Theorem 3.3 (Optimal Solution of a Linear Programming Problem)**.** *If a linear programming problem has a solution, it must occur at a vertex of the set of feasible solutions. If the problem has more than one solution, then at least one of them must occur at a vertex of the set of feasible solutions. In either case, the value of the objective function is unique.*

A **basic solution** of a linear programming problem in standard form is a solution $(x_1, x_2, \ldots, x_n, s_1, s_2, \ldots, s_m$ of the constraint equations in which at most m variables are nonzero. The nonzero variables are called **basic variables**. A basic solution for which all variables are nonnegative is called a **basic feasible solution**.

Algorithm 1 Simplex Method Iteration

Require: $\mathcal{B}$: the basic index set known as the basis of the problem
Require: $\mathcal{N}$: the nonbasic index set, complement of $\mathcal{B}$
Require: $x_B = B^{-1}b \geq 0$: m-element vector of basic variables
Require: $x_N = 0$: m-n element vector of nonbasic variables

Solve $B^T\lambda = c_B$ for λ;
Compute $s_N = c_N - N^T\lambda$; ▷ pricing

if $s_N \geq 0$ **then**
  stop; ▷ optimal point found
end if

Select $q \in \mathcal{N}$, with $s_q < 0$ as entering index;
Solve $Bd = A_q$ for d;

if $d \leq 0$ **then**
  stop; ▷ problem is unbounded
end if

Calculate $x_q^+ = \min_{i \mid d_i > 0} \frac{(x_B)_i}{d_i}$, and use p to denote the minimizing i;
Update $x_B^+ = x_B - dx_q^+$, $x_N^+ = (0, \ldots, 0, x_q^+, 0, \ldots, 0)^T$;
Change $\mathcal{B}$ by adding q and removing the basic variable corresponding to column p of B;

The Simplex method is ...

The algorithm works by ...

Algorithm block ...

The **search direction** move each iteration among the vertices of the polytope, meaning that they correspond to the direction from one vertex to the next choosen, i.e. their directed edge.

2. Answer

a)

The sketch of the feasible region in Figure 1 was done with an online tool using geogebra as backend.

b)

The basic feasible points are the vertices of the feasible region described by the polytope visible in Figure 1. The vertices can be computed by solving the system of equations following the geometrical interpretation:
Let $\bar{\mathbf{x}}_1 := x_1 \geq 0 \ \cap \ x_2 \geq 0$, the result is trivial since $x_1 = 0$ is the y-axis and $x_2 = 0$ is the x-axis, so the result is the origin O. Anyways, the intersection is computed by solving the system:

$$\begin{cases} x_1 = 0 \\ x_2 = 0 \end{cases} \qquad \bar{\mathbf{x}}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = O$$

Let $\bar{\mathbf{x}}_2 := x_1 \geq 0 \ \cap \ 3 + 3x_1 - 2x_2 \geq 0$. The intersection is computed by solving the system:

$$\begin{cases} x_1 = 0 \\ 3 + 3x_1 - 2x_2 = 0 \end{cases} \Rightarrow \begin{cases} x_1 = 0 \\ x_2 = \frac{3}{2} \end{cases} \qquad \bar{\mathbf{x}}_2 = \begin{bmatrix} 0 \\ \frac{3}{2} \end{bmatrix}$$

Let $\bar{\mathbf{x}}_3 := 3 + 3x_1 - 2x_2 \geq 0 \ \cap \ 6 - 2x_1 - 3x_2 \geq 0$. The intersection is computed by solving the system:

./figures/ex3-feasible-region.png

Figure 1: Feasible region for the problem 12 and 13

$$\begin{cases} 3 + 3x_1 - 2x_2 = 0 \\ 6 - 2x_1 - 3x_2 = 0 \end{cases} \Rightarrow \begin{cases} x_1 = \frac{2}{3}x_2 - 1 \\ x_2 = \frac{24}{13} \end{cases} \qquad \overline{\mathbf{x}}_3 = \begin{bmatrix} \frac{3}{13} \\ \frac{24}{13} \end{bmatrix}$$

Let $\overline{\mathbf{x}}_4 := 6 - 2x_1 - 3x_2 \geq 0 \ \cap \ 4 - 2x_1 - x_2 \geq 0$. The intersection is computed by solving the system:

$$\begin{cases} 6 - 2x_1 - 3x_2 = 0 \\ 4 - 2x_1 - x_2 = 0 \end{cases} \Rightarrow \begin{cases} x_2 = 1 \\ 4 - 2x_1 - (1) = 0 \end{cases} \qquad \overline{\mathbf{x}}_4 = \begin{bmatrix} \frac{3}{2} \\ 1 \end{bmatrix}$$

Let $\overline{\mathbf{x}}_5 := 4 - 2x_1 - x_2 \geq 0 \ \cap \ x_2 \geq 0$. The intersection is computed by solving the system:

$$\begin{cases} 4 - 2x_1 - x_2 = 0 \\ x_2 = 0 \end{cases} \Rightarrow \begin{cases} 4 - 2x_1 - (0) = 0 \\ x_2 = 0 \end{cases} \qquad \overline{\mathbf{x}}_5 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

To find the optimal point $\mathbf{x}^*$ that minimizes f subject to the constraints, we evaluate the function f at each vertex of the polytope in the feasible region:

$$\begin{aligned} f(\overline{\mathbf{x}}_1) &= 4 \cdot 0 + 3 \cdot 0 &&= 0 \\ f(\overline{\mathbf{x}}_2) &= 4 \cdot 0 + 3 \cdot \frac{3}{2} &&= \frac{9}{2} \\ f(\overline{\mathbf{x}}_3) &= 4 \cdot \frac{3}{13} + 3 \cdot \frac{24}{13} &&= \frac{84}{13} \\ f(\overline{\mathbf{x}}_4) &= 4 \cdot \frac{3}{2} + 3 \cdot 1 &&= 9 \\ f(\overline{\mathbf{x}}_5) &= 4 \cdot 2 + 3 \cdot 0 &&= 8 \end{aligned}$$

The optimal point $\mathbf{x}^*$ that minimizes f subject to the constraints is the point $\overline{\mathbf{x}}_\mathbf{1}$ vertex of the polytope in the feasible region.

$$\arg\min_{\mathbf{x}\in\overline{\mathcal{X}}} f(\mathbf{x}) = \overline{\mathbf{x}}_1 \qquad \overline{\mathcal{X}} = \{\overline{\mathbf{x}}_1, \overline{\mathbf{x}}_2, \overline{\mathbf{x}}_3, \overline{\mathbf{x}}_4, \overline{\mathbf{x}}_5\}$$

c)

To prove that the first order necessary conditions hold for the optimal points, we need to verify the KKT conditions from (10) for the optimal point $\overline{\mathbf{x}}_1 = \begin{bmatrix}0 & 0\end{bmatrix}^T$. The active set $\mathcal{A}(\overline{\mathbf{x}}_1) = \{5, 6\}$ where the constraints $c_5 := x_2 \geq 0$ and $c_6 := x_1 \geq 0$ from (13) are active, both $c_5, c_6 = 0 \therefore$ the point $\overline{\mathbf{x}}_1$ is at a polytope vertex. By checking the LICQ condition, the gradients of the active constraints are linearly independent:

$$\langle \nabla c_5, \nabla c_6 \rangle = \langle \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix} \rangle = 0$$

Find λ_5, λ_6:

$$\nabla f = \begin{bmatrix}4\\3\end{bmatrix}, \quad \nabla c_5 = \begin{bmatrix}0\\1\end{bmatrix}, \quad \nabla c_6 = \begin{bmatrix}1\\0\end{bmatrix}$$

$$\begin{bmatrix}4\\3\end{bmatrix} = \lambda_5 \begin{bmatrix}0\\1\end{bmatrix} + \lambda_6 \begin{bmatrix}1\\0\end{bmatrix} \Rightarrow \begin{cases}\lambda_6 = 4\\ \lambda_5 = 3\end{cases}$$

Finally, let's check that the gradient of the Lagrangian function is zero at the optimal point $\mathbf{x}^* = \overline{\mathbf{x}}_1$, as rewritten in (11):

$$\begin{aligned}\nabla_x \mathcal{L}(\mathbf{x}^*, \lambda^*) &= \nabla f - \sum_{i\in\mathcal{A}(\mathbf{x}^*)} \lambda_i^* \nabla c_i \\ &= \begin{bmatrix}4\\3\end{bmatrix} - 3\begin{bmatrix}0\\1\end{bmatrix} - 4\begin{bmatrix}1\\0\end{bmatrix} \\ &= \begin{bmatrix}0\\0\end{bmatrix} = \mathbf{0}\end{aligned}$$

Hence, the first order necessary conditions hold for the optimal point $\overline{\mathbf{x}}_1$.