0. Read sections 14.1, 14.2, 14.3, and 14.5 in the book.

1. *Problem 2.2(i) on page 489.* Determine the minimizers/maximizers of the following functions subject to the given constraints.

$$f(x_1, x_2) = x_1 x_2^3 \text{ subject to } 2x_1 + 3x_2 = 4.$$

Solution. For this problem we will use the $x = \bar{x} + Zv$ method and minimize $\phi(v)$. So we will need to determine an $\bar{x}$, find Z, and determine values for v.

$$A = \begin{pmatrix} 2 & 3 \end{pmatrix} \sim \begin{pmatrix} 1 & \frac{3}{2} \end{pmatrix} \text{ and thus,}$$
$$Z = \begin{pmatrix} -\frac{3}{2} & 1 \end{pmatrix}^T.$$

Next, let $\bar{x} = \begin{pmatrix} 2 & 0 \end{pmatrix}^T$ (it is a solution to $Ax = b$). Then,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x = \bar{x} + Zv = \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \begin{pmatrix} -\frac{3}{2} \\ 1 \end{pmatrix} v = \begin{pmatrix} 2 - \frac{3}{2}v \\ v \end{pmatrix}$$

Note that v is a scalar. Plugging x into f gives

$$\phi(v) = f(\bar{x} + Zv) = (2 - \frac{3}{2}v)(v)^3 = 2v^3 - \frac{3}{2}v^4.$$

Now, we have reduced the constrained optimization problem to an unconstrained optimization problem. We simply take the derivative and set it equal to zero, like in calculus. For a more general problem , we might need to use an iterative solver like Newton's method or BFGS, but this problem is nice enough that we can solve it with algebra.

$$0 \overset{\text{set}}{=} \nabla\phi(v) = 6v^2 - 6v^3 \Rightarrow 6v^2 = 6v^3 \Rightarrow v = 0 \text{ or } v = 1.$$

From here we can see that when $v = 0$ we will get $x_1 = 2$ and $x_2 = 0$. On the other hand when $v = 1$ we will get $x_1 = \frac{1}{2}$ and $x_2 = 1$. Next, we'll use these values to determine the type of stationary points these values may be.

$$\nabla f(x) = \begin{pmatrix} x_2^3 \\ 3x_1x_2^2 \end{pmatrix} \text{ and } \nabla^2 f(x) = \begin{pmatrix} 0 & 3x_2^2 \\ 3x_2^2 & 6x_1x_2 \end{pmatrix}$$

Next, we'll use Lemma 14.2 to check the stationary point $x_* = \begin{pmatrix} 2 & 0 \end{pmatrix}^T$. So we need to check that $Z^T\nabla f(x_*) = 0$ and $Z^T\nabla^2 f(x_*)Z$ is positive definite.

$$Z^T\nabla f(x_*) = \begin{pmatrix} -\frac{3}{2} & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$$
$$Z^T\nabla^2 f(x_*)Z = \begin{pmatrix} -\frac{3}{2} & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} -\frac{3}{2} \\ 1 \end{pmatrix} = 0$$

Thus, the reduced Hessian is only semidefinite, and we cannot conclude that $x_* = (2, 0)$ is a local minimizer. (Lemma 14.2 only gives a necessary, but not sufficient condition.)

Next, we'll use Lemma 14.3 to check the stationary point $x_* = \begin{pmatrix} \frac{1}{2} & 1 \end{pmatrix}^T$. For Lemma 14.3 we need to check that $Ax_* = b$, $Z^T \nabla f(x_*) = 0$, and $Z^T \nabla^2 f(x_*) Z$ is positive definite.

$$A x_* = \begin{pmatrix} 2 & 3 \end{pmatrix} \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \end{pmatrix} = b$$

$$Z^T \nabla f(x_*) = \begin{pmatrix} -\frac{3}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ \frac{3}{2} \end{pmatrix} = 0$$

$$Z^T \nabla^2 f(x_*) Z = \begin{pmatrix} -\frac{3}{2} & 1 \end{pmatrix} \begin{pmatrix} 0 & 3 \\ 3 & 3 \end{pmatrix} \begin{pmatrix} -\frac{3}{2} \\ 1 \end{pmatrix} = -6$$

Although we can see that $Z^T \nabla^2 f(x_*) Z$ is not positive definite, it is negative definite which implies that $x_* = \begin{pmatrix} \frac{1}{2} & 1 \end{pmatrix}^T$ is a strict local maximizer of f.

2. *Problem 2.2(vii) on page 490.* Determine the minimizers/maximizers of the following functions subject to the given constraints.

$$f(x_1, x_2) = \frac{1}{3} x_1^3 + x_2 \text{ subject to } x_1^2 + x_2^2 = 1$$

Solution. This problem has a nonlinear constraint, so the method used in the previous problem does not apply. Instead, we will use the Lagrangian function in its nonlinear form, namely, $\mathcal{L}(x, \lambda) = f(x) - \lambda^T \nabla g(x)$ (note that since we only have one constraint in this problem, λ is a scalar, so $\lambda^T = \lambda$).

$$\mathcal{L}(x_1, x_2, \lambda) = \frac{1}{3} x_1^3 + x_2 - \lambda(x_1^2 + x_2^2 - 1) = \frac{1}{3} x_1^3 + x_2 - \lambda x_1^2 - \lambda x_2^2 + \lambda$$

Our task is to find x_* and λ_* satisfying $\nabla \mathcal{L}(x_*, \lambda_*) = 0$. Thus, we proceed as follows.

$$\vec{0} \overset{\text{set}}{=} \nabla \mathcal{L} = \begin{pmatrix} x_1^2 - 2\lambda x_1 \\ 1 - 2\lambda x_2 \\ -x_1^2 - x_2^2 + 1 \end{pmatrix} \Rightarrow \begin{array}{c} x_1^2 = 2\lambda x_1 \\ 1 = 2\lambda x_2 \\ 1 = x_1^2 + x_2^2 \end{array}$$

Here we can see we have a few choices that we can make. First, **assuming** x_1 **is not zero**, we can solve for x_1 and x_2 in terms of λ to arrive at:

$$x_1 = 2\lambda$$
$$x_2 = \frac{1}{2\lambda}$$
$$(2\lambda)^2 + \left(\frac{1}{2\lambda}\right)^2 = 1 \;\Rightarrow\; 4\lambda^2 + \frac{1}{4\lambda^2} = 1 \;\Rightarrow\; 16\lambda^4 - 4\lambda^2 + 1 = 0.$$

Note that in general, solving a 4^{th} degree polynomial equation can be hard. However, there is a special form to this polynomial. Namely, it is 2^{nd} degree in λ^2:

$$\begin{aligned} 0 &= 16\lambda^4 - 4\lambda^2 + 1 \\ &= 16(\lambda^2)^2 - 4(\lambda^2) + 1 \end{aligned}$$

This means we can use the quadratic formula to solve for λ^2!

$$\lambda^2 = \frac{4 \pm \sqrt{4^2 - 4 \cdot 16 \cdot 1}}{2 \cdot 16} = \frac{4 \pm \sqrt{16 - 64}}{32}.$$

We have a negative under the square root, so there are no real solutions for λ. Moreover, since $x_1 = 2\lambda$ and $x_2 = \frac{1}{2\lambda}$, if λ is complex, x_1 and x_2 will be complex as well, but they are assumed to be real, so we must choose different values. Thus, the only possibility is that $x_1 = 0$.

Thus, we assume $x_1 = 0$ and similarly from above we can solve for λ and x_2 as follows.

$$x_1 = 0$$
$$1 = (0)^2 + (x_2)^2 \;\Rightarrow\; x_2 = \pm 1 \;\; (\textit{from constraint})$$
$$1 = 2\lambda(\pm 1) \;\Rightarrow\; \lambda = \pm\frac{1}{2}$$

Thus, we can check these two possibilities $x_1 = 0$, $x_2 = 1$, and $\lambda = \frac{1}{2}$ and $x_1 = 0$, $x_2 = -1$, and $\lambda = -\frac{1}{2}$. We will use Theorem 14.16, so we'll need to check that $\nabla_x \mathcal{L}(x_*, \lambda_*) = 0$ and $Z(x_*)^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) Z(x_*)$ is positive definite.

First, we'll consider the first stationary point $x_* = \begin{pmatrix} 0 & 1 \end{pmatrix}^T$ and $\lambda_* = \frac{1}{2}$.

$$\nabla g(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \quad \text{and} \quad \nabla g(x_*) = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \quad \text{so let} \quad Z = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$
$$\nabla^2_{xx}\mathcal{L}(x,\lambda) = \begin{pmatrix} 2x_1 & 0 \\ 0 & 0 \end{pmatrix} - \lambda \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 2x_1 - 2\lambda & 0 \\ 0 & -2\lambda \end{pmatrix}$$
$$\text{so} \quad \nabla^2_{xx}\mathcal{L}(x_*,\lambda_*) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$
$$Z(x_*)^T \nabla^2_{xx}\mathcal{L}(x_*,\lambda_*) Z(x_*) \;\Rightarrow\; \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = -1$$

Although $Z(x_*)^T \nabla^2_{xx}\mathcal{L}(x_*,\lambda_*) Z(x_*)$ is not positive definite, it is negative definite and thus $x_* = \begin{pmatrix} 0 & 1 \end{pmatrix}^T$ is a local maximizer of f. Finally, we'll consider the other stationary point $x_* = \begin{pmatrix} 0 & -1 \end{pmatrix}^T$ and $\lambda_* = -\frac{1}{2}$.

$$\nabla g(x_*) = \begin{pmatrix} 0 \\ -2 \end{pmatrix} \quad \text{so let} \quad Z = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$
$$\nabla^2_{xx}\mathcal{L}(x_*,\lambda_*) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$
$$Z(x_*)^T \nabla^2_{xx}\mathcal{L}(x_*,\lambda_*) Z(x_*) \;\Rightarrow\; \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1$$

The stationary point $x_* = \begin{pmatrix} 0 & -1 \end{pmatrix}^T$ and associated $\lambda_* = -\frac{1}{2}$ satisfies all the conditions for Theorem 14.16 and thus x_* is a strict local minimizer of f.

3. *Problem 5.2 on page 509.* Solve the problem

$$\begin{aligned} \text{minimize} \quad & f(x) = c^T x \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i = 0 \\ & \sum_{i=1}^{n} x_1^2 = 1 \end{aligned}$$

Solution. We solve this problem using the Lagrangian method. First, we'll make the constraints easier to use in the function by defining them in a different way.

$$\text{Let } \ w = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}^T.$$
$$\text{Note that:} \quad w^T x = \sum_{i=1}^{n} x_i = 0$$
$$\text{and } \ x^T x = \sum_{i=1}^{n} x_1^2 = 1.$$

Next, we write down the Lagrangian function and set its gradient equal to zero

$$\mathcal{L} = c^T x - \lambda^T \begin{pmatrix} w^T x \\ x^T x - 1 \end{pmatrix} = c^T x - \lambda_1 w^T x - \lambda_2 x^T x + \lambda_2$$
$$0 \overset{\text{set}}{=} \nabla \mathcal{L} = c - \lambda_1 w - \lambda_2 x$$
$$\Rightarrow \ x = \frac{1}{\lambda_2}(c - \lambda_1 w)$$

We take the dot-product of the last relationship with w, x, and c, to obtain:

$$(1) \qquad \begin{cases} w^T x = \dfrac{1}{\lambda_2}(w^T c - \lambda_1 w^T w) \\ x^T x = \dfrac{1}{\lambda_2}(x^T c - \lambda_1 x^T w) \\ c^T x = \dfrac{1}{\lambda_2}(c^T c - \lambda_1 c^T w) \end{cases}$$

From the constraints, we know that $w^T x = 0$, and $x^T x = 1$. Also, note that $c^T c = \|c\|^2, w^T w = n$, and $c^T w = \sum_{i=1}^n c_i = n \frac{1}{n} \sum_{i=1}^n c_i = n\bar{c}$, where $\bar{c} = \frac{1}{n} \sum_{i=1}^n c_i$ denotes the average value of the components of the vector c. Using these relations, we obtain

$$(2) \qquad \begin{cases} 0 = \dfrac{1}{\lambda_2}(n\bar{c} - n\lambda_1) \\ 1 = \dfrac{1}{\lambda_2} x^T c \\ c^T x = \dfrac{1}{\lambda_2}(\|c\|^2 - \lambda_1 n\bar{c}) \end{cases}$$

From the first equation, we see that $\lambda_1 = \bar{c}$. Note that $x^T c = x \cdot c = c \cdot x = c^T x$, and that, from the second equation, $\lambda_2 = x^T c$. Thus, we can combine this with the third equation to obtain:

$$\begin{aligned} \lambda_2 &= \frac{1}{\lambda_2}(\|c\|^2 - \lambda_1 n\bar{c}) \\ \Rightarrow \lambda_2^2 &= \|c\|^2 - \lambda_1 n\bar{c} \\ &= \|c\|^2 - n\bar{c}^2 && \text{(using } \lambda_1 = \bar{c}\text{)} \\ &= n\left(\frac{1}{n}\sum_{i=1}^{n} c_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} c_i\right)^2\right) && \text{(by definitions and algebra)} \end{aligned}$$

Now, we would like to take a square root to find λ_2, but how do we know the right-hand side is non-negative? Note that the question of whether the right-hand side is non-negative is the question of *whether the average of the squares is greater than or equal to the square of the average.* To prove that it is (you are not required to for your homework), we can use the Cauchy-Schwarz inequality:

$$|\vec{x} \cdot \vec{y}| = \|\vec{x}\|\|\vec{y}\||\cos(\theta)| \leq \|\vec{x}\|\|\vec{y}\|$$

Letting $\vec{x} = c$ and $\vec{y} = w$, we find that

$$|n\bar{c}| = \left|c^T w\right| = |c \cdot w| \leq \|c\|\|w\| = \sqrt{n}\|c\|$$

Squaring both sides and dividing by n, we find,

$$n\bar{c}^2 \leq \|c\|^2$$

so that $\|c\|^2 - n\bar{c}^2 \geq 0$. Thus, we can indeed take a square root, and we find

$$\lambda_2 = \pm\sqrt{\|c\|^2 - n\bar{c}^2}.$$

Substituting λ_1 and λ_2 back into x, we obtain

$$x = \frac{1}{\lambda_2}(c - \lambda_1 w) = \pm\frac{c - \bar{c}w}{\sqrt{\|c\|^2 - n\bar{c}^2}}.$$

Therefore, at these two points, we find

$$f(x) = c^T x = \pm\frac{\|c\|^2 - n\bar{c}^2}{\sqrt{\|c\|^2 - n\bar{c}^2}} = \pm\sqrt{\|c\|^2 - n\bar{c}^2}$$

One point is positive, the other is negative. Therefore, we have found that the maximum is $\sqrt{\|c\|^2 - n\bar{c}^2}$ and occurs at $x = \frac{c - \bar{c}w}{\sqrt{\|c\|^2 - n\bar{c}^2}}$, and the minimum is $-\sqrt{\|c\|^2 - n\bar{c}^2}$ and occurs at $x = -\frac{c - \bar{c}w}{\sqrt{\|c\|^2 - n\bar{c}^2}}$.