1. Let $f_1, f_2, \ldots, f_n$ be linearly independent functionals on a vector space $X$. Show that there are $n$ elements $x_1, x_2, \ldots, x_n$ in $X$ such that the $n \times n$ matrix $[f_i(x_j)]$ is nonsingular.

[This is problem 11 in page 210 of the book. This result was used in the proof of Theorem 1, page 187.]

**Solution:** We will prove this using induction on $n$.

For $n = 1$, since $f_1$ is not identically equal to 0, there exists an $x_1$ such that $f_1(x_1) \neq 0$.

Now, assume that the given statement is true for $n$. We will show that it is true for $n + 1$.

Choose $x_{n+1}$ such that $f_{n+1}(x_{n+1}) \neq 0$. Such a $x_{n+1}$ must exist. Now, for $i = 1, \ldots, n$, define

$$g_i = - \left( \frac{f_i(x_{n+1})}{f_{n+1}(x_{n+1})} \right) f_{n+1} + f_i$$

Note that this is a linear combination of the functionals $f_{n+1}$ and $f_i$, where $- \left( \frac{f_i(x_{n+1})}{f_{n+1}(x_{n+1})} \right)$ is a well defined constant. Choose $x_1, x_2, \ldots, x_n$ such that $[g_i(x_j)]$ is non-singular. Such $x$’s exist by the induction hypothesis.

Now, we will prove that $[f_i(x_j)]$, $i, j \in \{1, 2, \ldots, n + 1\}$ is non-singular. Suppose not, there exists $\lambda_1, \ldots, \lambda_{n+1}$ such that

$$f_i(x_{n+1}) = \sum_{j=1}^{n} \lambda_j f_i(x_j) \quad \text{for } i = 1, 2, \ldots, n + 1 \quad (1)$$

$$\frac{f_i(x_{n+1})}{f_{n+1}(x_{n+1})} f_{n+1}(x_{n+1}) = \sum_{j=1}^{n} \lambda_j f_i(x_j) \quad \text{for } i = 1, 2, \ldots, n \quad (2)$$

$$\frac{f_i(x_{n+1})}{f_{n+1}(x_{n+1})} \sum_{j=1}^{n} \lambda_j f_{n+1}(x_j) = \sum_{j=1}^{n} \lambda_j f_i(x_j) \quad \text{for } i = 1, 2, \ldots, n \quad (3)$$

$$\sum_{j=1}^{n} \lambda_j \left( - \frac{f_i(x_{n+1})}{f_{n+1}(x_{n+1})} f_{n+1}(x_j) + f_i(x_j) \right) = 0 \quad \text{for } i = 1, 2, \ldots, n \quad (4)$$

$$\sum_{j=1}^{n} \lambda_j g_i(x_j) = 0 \quad \text{for } i = 1, 2, \ldots, n \quad (5)$$

Equation 2 is well-defined because $f_{n+1}(x_{n+1}) \neq 0$ and (3) follows from (1) with $i = n + 1$. Note that (5) means that $[g_i(x_j)]$ is non-singular, which is a contradiction.

2. Lagrange Multipliers Method and Duality with Equality constraints: Exercise 7 in page 236 of the book shows that Theorem 1 on page 217 can be generalized to problems with a
finite number of linear equality constraints along with inequality constraints. In this problem, we will assume the result in Exercise 7, and show that duality and sufficiency hold with a finite number of linear equality constraints.

Let \( \mu_0 = \inf \{ f(x) : x \in \Omega, G(x) \leq 0, H(x) = 0 \} \) as defined in Exercise 7. For parts (i) and (ii) below, also assume all the conditions in Exercise 7 and Theorem 1 hold. Define

\[
L(x, z^*, y^*) = f(x) + \langle G(x), z^* \rangle + \langle H(x), y^* \rangle
\]

where \( z^* \in \mathbb{Z}^*, z^* \geq 0 \) and \( y^* \in \mathbb{Y}^* \). Define \( D(z^*, y^*) = \inf_{x \in \Omega} L(x, z^*, y^*) \).

(i) Show that \( \mu_0 = \max_{z^* \geq 0, y^* \in \mathbb{Y}^*} D(z^*, y^*) \)

(ii) Moreover, if there is an \( x_0 \in \Omega \) s.t \( G(x_0) \leq 0, H(x_0) = 0 \) and \( \mu_0 = f(x_0) \), then,

\[
\langle G(x), z_0^* \rangle = 0,
\]

and \( x_0 \) solves

\[
\min_{x_0 \in \Omega} L(x, z_0^*, y_0^*),
\]

where \( (z_0^*, y_0^*) \) is the solution to part (i).

(iii) If there exists an \( x_0 \in \Omega, y_0^* \), and \( z_0^* \geq 0 \) such that \( L(x_0, z_0^*, y_0^*) \leq L(x, z_0^*, y_0^*) \) for all \( x \in \Omega \) and \( \langle G(x_0), z_0^* \rangle = 0 \), then \( x_0 \) must be a solution to \( \inf_x f(x) \) such that \( x \in \Omega, G(x) \leq 0, H(x) = 0 \).

**Solution:**

(i)

\[
D(z^*, y^*) = \inf_{x \in \Omega} L(x, z^*, y^*)
\]

\[
\leq \inf_{x \in \Omega, G(x) \leq 0, H(x) = 0} L(x, z^*, y^*)
\]

\[
= \inf_{x \in \Omega, G(x) \leq 0, H(x) = 0} f(x) + \langle G(x), z^* \rangle
\]

\[
\leq \inf_{x \in \Omega, G(x) \leq 0, H(x) = 0} f(x)
\]

since \( z^* \geq 0 \)

\[
= \mu_0.
\]

From Exercise 7, we know that there exists \( z_0^* \geq 0 \) and \( y_0^* \) s.t. \( D(z_0^*, y_0^*) = \mu_0 \). Therefore, we have that \( \mu_0 = \max_{z^* \geq 0, y^* \in \mathbb{Y}^*} D(z^*, y^*) \). This proves part (i).

(ii) Assume that there is an \( x_0 \) which solves the original problem. Then,

\[
\mu_0 = \inf_{x \in \Omega} f(x) + \langle G(x), z_0^* \rangle + \langle H(x), y_0^* \rangle
\]

\[
\leq f(x_0) + \langle G(x_0), z_0^* \rangle + \langle H(x_0), y_0^* \rangle
\]

\[
\leq f(x_0)
\]

since \( G(x_0) \leq 0 \) and \( H(x_0) = 0 \)

\[
= \mu_0
\]

by the result in Exercise 7.

Therefore, we have \( \langle G(x_0), z_0^* \rangle = 0 \) and \( x_0 \) minimizes \( L(x, z_0^*, y_0^*) \).

(iii) Suppose there exists an \( x_1 \in \Omega \) satisfying the constraints such that \( f(x_1) < f(x_0) \). Then,

\[
f(x_1) + \langle G(x_1), z_0^* \rangle + \langle H(x_1), y_0^* \rangle < f(x_0) + \langle G(x_1), z_0^* \rangle + \langle H(x_1), y_0^* \rangle
\]

\[
\leq f(x_0)
\]

\[
= f(x_0) + \langle G(x_0), z_0^* \rangle + \langle H(x_0), y_0^* \rangle,
\]

thus contradicting the assumption that \( L(x_0, z_0^*, y_0^*) \leq L(x, z_0^*, y_0^*) \).

Page 2
Remark: Note that Exercise 7 in page 236 of the book, and this problem, suggests the following two strategies to solve convex optimization problems.

- Form the Lagrangian \( L \), and solve for \( x \) by minimizing the Lagrangian over \( x \in \Omega \). This solution \( x \) will be a function of the variables \( z^*, y^* \). Since we know that the optimal solution \( x_0 \) to the original constrained optimization problem and the Lagrange multipliers \( z_0^*, y_0^* \) must satisfy
  \[
  \langle G(x_0), z_0^* \rangle = 0, \quad G(x_0) \leq 0, \quad H(x_0) = 0, \quad z_0^* \geq 0,
  \]
  we can substitute the \( x \) obtained as a function of \( z^* \) and \( y^* \) in these inequalities/equalities to obtain \( x_0, z_0^* \) and \( y_0^* \). Sufficiency ensures that this \( x_0 \) must be an optimal solution.

- Find the dual function, solve the dual problem to compute \( z_0^*, y_0^* \). Then, solve for \( x_0^* \) by minimizing \( L(x, z_0^*, y_0^*) \) over all \( x \in \Omega \).

3. (Horse Racing Problem)

What is the best way to place a bet totaling \( x_0 \) dollars in a race involving \( n \) horses? Assume we know \( p_i \), the probability that the \( i \)-th horse wins, and \( s_i \), the amount that the rest of the public is betting on the \( i \)-th horse. The track keeps a proportion \( 1 - C \) of the total amount bet (\( 0 < 1 - C < 1 \)) and distributes the rest among the public in proportion to the amounts bet on the winning horse. [This is Example 2 on page 203 in the book, solved using Lagrange multipliers.]

**Solution:** Let \( x_i \) denote the bet on \( i \)-th horse. Then, one receives \( C(x_0 + \sum_i s_i) \frac{x_i}{s_i + x_i} \) if the \( i \)-th horse wins. Then the total expected return \( R \) is

\[
R(x_1, \ldots, x_n) = C(x_0 + \sum_i s_i) \sum_i \frac{p_i x_i}{s_i + x_i} - x_0
\]

So, the problem is to maximize \( R(x_1, \ldots, x_n) \) s.t \( \sum_i x_i = x_0 \) and \( x_i \geq 0 \) for all \( i = 1, 2, \ldots, n \). Equivalently,

\[
\min \quad -\sum_{i=1}^{n} \frac{p_i x_i}{s_i + x_i}
\]

s.t

\[
\sum_{i=1}^{n} x_i = x_0
\]

\[
-x_i \leq 0 \quad \text{for } i = 1, \ldots, n
\]

Using the first strategy outlined at the end of Problem 2, we first solve

\[
\min_{x \in \mathbb{R}^n} -\sum_{i=1}^{n} \frac{p_i x_i}{s_i + x_i} + \lambda \left( \sum_{i=1}^{n} x_i - x_0 \right) + \sum_i \mu_i x_i
\]

Differentiating w.r.t \( x_i \) and setting it equal to zero, we get

\[
\frac{p_i s_i}{(s_i + x_i)^2} = \lambda + \mu_i
\]
\[ s_i + x_i = \sqrt{\frac{p_is_i}{\lambda + \mu_i}} \]  

(7)

From the second part of the previous problem (6), we know that \(\mu_i x_i = 0\).

So, whenever \(x_i > 0\), from (7), \(x_i = \sqrt{\frac{p_is_i}{\lambda}} - s_i\). But since \(x_i > 0\), we must have \(\lambda < \frac{p_i}{s_i}\).

Further, if \(\lambda \geq \frac{p_i}{s_i}\), \(x_i = 0\) and \(\mu_i \geq 0\) satisfy (7).

Without loss of generality, assume that the horses are numbered such that \(\frac{p_1}{s_1} \geq \frac{p_2}{s_2} \ldots \frac{p_n}{s_n}\). So, depending on the value of \(\lambda\), either all the \(x_i\)'s are positive or there an index \(m\) s.t.

\[
x_i = \begin{cases} 
\sqrt{\frac{p_is_i}{\lambda}} - s_i & \text{for } i = 1, 2, \ldots, m \\
0 & \text{for } i = m + 1, \ldots, n 
\end{cases}
\]

We can now solve for \(\lambda\) by noting that the constraint \(\sum_{i=1}^n x_i = x_0\) must be satisfied. Summing (7) for \(i = 1\) to \(m\), we get

\[
\sum_{i=1}^m s_i + x_0 = \frac{\sum_{i=1}^m \sqrt{p_i s_i}}{\sqrt{\lambda}}
\]

In other words, \(\lambda\) should be chosen so that

\[
S(\lambda) := \sum_{i: p_i/s_i > \lambda} \left( \sqrt{\frac{p_is_i}{\lambda}} - s_i \right) = x_0.
\]

Clearly \(S(0) = \infty\) and \(S(\infty) = 0\), and it is easy to see that \(S(\lambda)\) is continuous in \(\lambda\). So, the above equation has a solution.

4. Find a \(u \in L_2[0,1]\) that minimizes \(f(u) = \frac{1}{2} \int_0^1 u^2(t)dt\) while satisfying the linear constraints \(Ku = c\) where \(K : L_2[0,1] \to E^n\).

[This is Example 3 on page 205 in the book, solved using Lagrange multipliers.]

**Solution:** We will use the duality result in Problem 2. We will leave it as an exercise to verify that the assumptions of that problem are satisfied. Now, consider the dual function \(D(y^*)\) defined for all \(y^* \in E^n\) as follows.

\[
D(y^*) = \inf_{u \in L_2[0,1]} \left( \frac{1}{2} \int_0^1 u^2(t)dt + \langle Ku - c, y^* \rangle \right)
\]

\[
= \inf_{u \in L_2[0,1]} \left( \frac{1}{2} \int_0^1 u^2(t)dt + \langle u, K^* y^* \rangle \right) - \langle c, y^* \rangle \quad \text{where } K^* \text{ is the adjoint of } K
\]

\[
= \inf_{u \in L_2[0,1]} \left( \int_0^1 \left( \frac{1}{2} u^2(t) + u(t)K^* y^*(t) \right) dt \right) - \langle c, y^* \rangle
\]

It is easy to solve the above problem to get \(u = -K^* y^*\). Note that \(y^* \in E^n\) and \(K^* : E^n \to L_2[0,1]\) and so \(K^* y^* \in L_2[0,1]\). Therefore, we get

\[
D(y^*) = -\frac{1}{2} \langle K^* y^*, K^* y^* \rangle - \langle c, y^* \rangle
\]
Now the dual problem is a finite dimensional unconstrained optimization problem:

$$\max_{y^* \in E^n} -\frac{1}{2} \langle KK^*y^*, y^* \rangle - \langle c, y^* \rangle$$

where $KK^*$ is a $n \times n$ matrix. This is solved by a $y_0^*$ s.t $KK^*y_0^* = -c$. Using the second strategy outlined at the end of Problem 2, we have that solving the original problem is same as solving

$$\inf_{u \in L^2[0,1]} \left( \int_0^1 \left( \frac{1}{2} u^2(t) + u(t)K^*y_0^*(t) \right) dt \right) - \langle c, y_0^* \rangle.$$

We already know that the solution to this problem is $u_0 = -K^*y_0^*$. In summary, the optimal $u(t)$ for the original problem is $u_0 = -K^*y_0^*$ where $y_0^* \in E^n$ is the solution to $KK^*y_0^* = -c$.

5. Consider the fixed-point (FP) equation

$$x(t) = \frac{1}{2} t^3 + \alpha \sin \pi x(t)$$

defined over the interval $t \in [-2, 2]$, with $x \in C[-2, 2]$, and $\alpha$ a positive constant (a parameter). For what values of $\alpha$ does there exist a unique continuous function $x(\cdot)$ on $[-2, 2]$ which solves the FP equation. Show (prove) that for these values of $\alpha$ a solution indeed exists and is unique.

**Hint:** Use a contraction mapping type argument, applied to a subset of $C[-2, 2]$, which comprises all uniformly bounded functions, such as functions satisfying the bound $|x(t)| \leq \beta$, for some $\beta$.

**Solution:** $x(t) = T(x)(t) = \frac{1}{2} t^3 + \alpha \sin \pi x(t); \quad -2 \leq t \leq 2$.

Note that

$$|T(x)(t)| \leq \frac{1}{2} \max_t |t^3| + \alpha \max_x |\sin \pi x(t)| \leq 4 + \alpha.$$

Let $S = \{ x \in C[-2, 2] : |x(t)| \leq 4 + \alpha \}$. Clearly $T : S \to S$, and $S$ is complete (being a closed and bounded subset of another complete metric space, $C[-2, 2]$).

Now, $|T(x)(t) - T(y)(t)| = \alpha \sin \pi x(t) - \sin \pi y(t) | \leq 2\alpha \left| \sin \frac{\pi x(t) - \pi y(t)}{2} \right| \cdot \left| \cos \frac{\pi x(t) + \pi y(t)}{2} \right|$

$$\leq \alpha \pi |x(t) - y(t)| \left| \frac{\cos \frac{\pi x(t) + \pi y(t)}{2}}{2} \right| \leq 1.$$

Therefore, $\|T(x) - T(y)\| \leq \alpha \pi \max_t |x(t) - y(t)| = \alpha \pi \|x - y\|$

$\Rightarrow$ For $\alpha \pi < 1$, $T$ is a contraction mapping $\Rightarrow \exists$ a unique $x^0 \in S \ni T(x^0) = x^0$.

6. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a transformation defined by $T(x) = Ax + b$, where $A$ is an $n \times n$ matrix with real entries $a_{ij}$, and $b$ is a given vector in $\mathbb{R}^n$, with components $b_i$'s.
(a) Under what conditions on the $a_{ij}$’s and $b_i$’s is $T$ a contraction when the norm on $\mathbb{R}^n$ is the Euclidean one, that is $\|x\| = \left\{ \sum_{i=1}^{n} (x_i)^2 \right\}^{1/2}$?

(b) Under what conditions on the $a_{ij}$’s and $b_i$’s is $T$ a contraction when the norm on $\mathbb{R}^n$ is the maximum norm, that is $\|x\| = \max_i |x_i|$? Are these conditions more or less restrictive than the ones obtained in part (a)?

(c) We now wish to compute a fixed point of $T$ by using the iteration (successive approximation) 

$$x_{(i+1)} = Ax_{(i)} + b, \quad i = 0, 1, \ldots$$

where $x_{(0)}$ is an arbitrary starting point. Based on the results you obtained in parts (a) and (b) above, what can you deduce as the conditions on the $a_{ij}$’s and $b_i$’s for this sequence to converge to a fixed point of $T$.

Solution:

We are given the linear mapping $T : \mathbb{R}^n \to \mathbb{R}^n$, $T(x) = Ax + b$

(a) Under the Euclidean norm,

$$\|T(x) - T(y)\|^2 = \|A(x - y)\|^2 = (x - y)^T A^T A (x - y) \leq \lambda_{\text{max}}(A^T A) (x - y)^T (x - y)$$

where $\lambda_{\text{max}}(A^T A)$ is the maximum eigenvalue of the nonnegative-definite (symmetric) matrix $A^T A$. Hence, $T$ is a contraction if

$$\lambda_{\text{max}}(A^T A) < 1 \iff I - A^T A > 0$$

(b) $T$ is a contraction under the metric $\rho(x, y) = \max_{1 \leq i \leq n} |x_i - y_i|$, if

$$\sum_{j=1}^{n} |a_{ij}| < 1 \quad \forall i = 1, \ldots, n$$

where $a_{ij}$ is the $ij$-th entry of the matrix $A$. This follows because:

$$\rho(T(x), T(y)) = \max_i \left| \sum_j a_{ij} (x_j - y_j) \right| \leq \max_i \sum_j |a_{ij}| \max_j |x_j - y_j| = \max_i \sum_j |a_{ij}| \rho(x, y)$$

The two conditions (in part (a) and in part (b) above) do not admit any ordering as there are cases where the first one is more restrictive than the second, and also cases where it is the other way around. Consider, for example, the $2 \times 2$ matrix $A$ where $a_{11} = 0.8, a_{12} = a_{21} = 0.3, a_{22} = 0.2$, which is symmetric and positive definite. Both of its eigenvalues are less than 1 (and hence those of $A^T A = A^2$), but the first row sum is 1.1 and hence larger than 1. Thus, in this case condition (a) is satisfied, but (b) is not. To show that the reverse can also happen, take the elements of $A$ to be: $a_{11} = a_{21} = 0.76, a_{12} = a_{22} = 0.19$. In this case, the maximum row sum is 0.95 (and hence less than 1), but the maximum eigenvalue of $A^T A$ (which is the only positive eigenvalue) is $1.2274 > 1$. 

Page 6
Under either of the two contraction conditions above, the successive approximation algorithm converges for any starting point, since $\mathbb{R}^n$ is a Banach space. But as we have seen, neither one implies the other. The question is whether there exists one (obtained under a different norm, and/or using some powers of $T$) which would be less restrictive than these two (and in fact least restrictive). The answer is yes! There exists a least restrictive condition, which is (from linear system theory):

All eigenvalues of $A$ should have their absolute values less than 1, i.e., $|\lambda(A)| < 1$. 