Project Report

Functional Analysis (MA2212)

Applications of Functional Analysis in Economics, Finance, and Machine Learning

Course Instructor & Mentor: Dr. Meraj Alam

Submitted By:

Madhav Basur
Roll No: SE23UCAM009
Pulipaka Prem Siddharth
Roll No: SE23UCAM033
Sanket Motagi
Roll No: SE23UCAM020
Tupakula Savith Chandrahas
Roll No: SE23UCAM026

Mahindra University
Department of Mathematics

May 20, 2025

Abstract

Functional Analysis, a branch of mathematical analysis, plays a foundational role in numerous modern applications, including economic theory, financial modeling, and machine learning. This report explores key concepts from Functional Analysis such as convex sets, convex functions, and Lagrange duality, and illustrates their practical use in optimization problems across domains. We aim to provide both theoretical insights and real-world relevance to these mathematical tools, emphasizing their significance in decision-making, risk management, and computational algorithms.

1 Introduction

Functional Analysis provides the framework for studying infinite-dimensional vector spaces and linear operators, and serves as the mathematical backbone for optimization and variational problems. Its techniques have proven particularly effective in handling constrained optimization problems, which are central to economic equilibrium analysis, portfolio optimization, and training of machine learning models.

In this report, we focus on the study of convexity and duality — two core ideas that not only simplify mathematical treatment of optimization problems but also lend themselves to efficient computational algorithms. We will present their definitions, properties, and practical interpretations in various applied settings.

2 Preliminary Concepts

To understand the applications discussed, we introduce key foundational notions from Functional Analysis and Optimization:

Vector Spaces and Norms

A vector space V over $\mathbb{R}$ is a set equipped with operations of vector addition and scalar multiplication satisfying standard axioms. A norm on V, denoted $\|x\|$, assigns a non-negative length to each vector and induces a metric.

Convex Sets

A set $C \subseteq \mathbb{R}^n$ is convex if for any $x, y \in C$ and $\lambda \in [0, 1]$,

$$\lambda x + (1 - \lambda)y \in C.$$

Convexity ensures that linear interpolations between feasible solutions remain feasible.

Convex Functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if its domain is convex and for all x, y and $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

Lagrange Multipliers

These are variables introduced to transform constrained optimization problems into unconstrained ones, via the construction of the Lagrangian function.

3 Main Content

Convex Sets

Convex sets form the foundation of convex optimization. This section presents formal definitions, examples, and key operations that preserve convexity.

Affine and Convex Sets

Affine Set: A set $C \subseteq \mathbb{R}^n$ is affine if for any $x_1, x_2 \in C$ and $\theta \in \mathbb{R}$,

$$\theta x_1 + (1-\theta)x_2 \in C$$

More generally, if $\sum_{i=1}^k \theta_i = 1$, then $\sum_{i=1}^k \theta_i x_i \in C$.

Convex Set: A set $C \subseteq \mathbb{R}^n$ is convex if for any $x_1, x_2 \in C$ and $\theta \in [0,1]$,

$$\theta x_1 + (1-\theta)x_2 \in C$$

Extension to Infinite Dimensions: In functional analysis, the notions of affine and convex sets extend naturally to infinite-dimensional spaces like Banach and Hilbert spaces. A set C in a real vector space X (possibly infinite-dimensional) is called affine or convex if it satisfies the same definitions, i.e., closure under affine or convex combinations. In particular, convex sets in Hilbert spaces inherit properties such as the existence of nearest points (projection theorem) due to the geometry induced by the inner product.

Some Important Examples of Convex Sets

- **Hyperplanes:** The set $\{x \in \mathbb{R}^n \mid a^T x = b\}$ represents a flat $(n-1)$-dimensional surface in $\mathbb{R}^n$. In $\mathbb{R}^2$, it's a line; in $\mathbb{R}^3$, it's a plane. It is both affine and convex, as it is closed under affine combinations.
  In a **Hilbert Space** H, a hyperplane is defined by $x \in H | \langle x, a \rangle = b$, where $a \neq 0$ and $\langle \cdot, \cdot \rangle$ is the inner product. This generalises the concept of hyperplanes from R^n to infinite dimensions.

- **Halfspaces:** The set $\{x \in \mathbb{R}^n \mid a^T x \leq b\}$ includes all points on one side of a hyperplane (and possibly the hyperplane itself). Halfspaces are convex because the linear inequality defines a region where convex combinations of any two points remain within the region.
  Similarly, a halfspace in a **Hilbert space** is defined by $x \in H | \langle x, a \rangle \leq b$. Halfspaces remain convex in infinite-dimensional spaces.

- **Euclidean Balls:** The ball $B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\}$ includes all points within a distance r of center x_c. The 2-norm is convex, and its sublevel sets (like

this one) are convex. Balls are symmetric and represent isotropic distances.
In **Banach** or **Hilbert spaces**, balls are defined by the norm of the space: $B(x_c, r) = x \| \|x - x_c\| \leq r$. Convexity of balls holds regardless of dimensionality.

- **Ellipsoids:** Generalizing balls, ellipsoids are defined by $\{x \mid (x-x_c)^T P^{-1}(x-x_c) \leq 1\}$ for a positive definite matrix P. They represent stretched or rotated balls. The eigenvectors of P determine the axes, and eigenvalues control the scaling. The set is convex as it is a sublevel set of a convex quadratic form.
Ellipsoids can be generalized to Hilbert spaces by replacing matrix P with a bounded, self-adjoint, positive definite operator. The set remains convex.

- **Polyhedra:** A polyhedron is any set that can be represented as $\{x \mid Ax \leq b,\ Cx = d\}$, i.e., an intersection of a finite number of halfspaces and hyperplanes. Examples include cubes, prisms, and pyramids. Convexity follows from the fact that convexity is preserved under intersection.
In **infinite-dimensional spaces**, polyhedra are less commonly used since they rely on finitely many linear inequalities. Their direct analog is the intersection of finitely many halfspaces, but they may fail to be closed or bounded under weak topology.

- **Simplex:** The standard simplex is $\{x \in \mathbb{R}^n \mid x_i \geq 0,\ \sum x_i = 1\}$. It is the set of all convex combinations of the standard basis vectors. This set is used to model probability distributions, where components are nonnegative and sum to one. The simplex is convex by definition.
In **infinite-dimensional settings**, the simplex generalizes to probability measures on measurable spaces, or to convex sets in l_1 or L^1 spaces satisfying similar "sum to 1" conditions.

- **Convex Hull:** Given a set S, its convex hull is the smallest convex set that contains S:
$$\mathrm{conv}(S) = \left\{ \sum \theta_i x_i \mid x_i \in S,\ \theta_i \geq 0,\ \sum \theta_i = 1 \right\}$$
It includes all convex combinations of points in S, essentially filling in the "shape" formed by those points.
The convex hull of a set S in a **Banach space** is defined the same way, but care must be taken about closure: in **infinite dimensions**, convex hulls are often considered along with their closure in weak or norm topology.

Operations That Preserve Convexity

Several operations allow us to build new convex sets from existing ones. These are crucial when modeling constraints in optimization problems.

- **Intersection:** If $C_1, C_2, \ldots, C_k$ are convex sets, then their intersection $C = C_1 \cap C_2 \cap \cdots \cap C_k$ is also convex. This follows directly from the definition of convexity — any convex combination of points in all sets will also lie in the intersection.

- **Affine Images:** If $C \subseteq \mathbb{R}^n$ is convex, and we apply an affine transformation $x \mapsto Ax + b$, then the resulting set $AC + b = \{Ax + b \mid x \in C\}$ is also convex. Affine transformations preserve straight lines and weighted averages.

- **Preimages under Affine Maps:** If $D \subseteq \mathbb{R}^m$ is convex and $f(x) = Ax + b$, then the preimage $f^{-1}(D) = \{x \mid f(x) \in D\}$ is convex. This is important when constraints are imposed on a transformed variable.
- **Cartesian Products:** If $C_1 \subseteq \mathbb{R}^{n_1}$ and $C_2 \subseteq \mathbb{R}^{n_2}$ are convex, then the product set $C_1 \times C_2 = \{(x, y) \mid x \in C_1, y \in C_2\} \subseteq \mathbb{R}^{n_1+n_2}$ is convex. The idea is that convexity in separate spaces carries over to the product space.
- **Perspective Transformations:** Let $C \subseteq \mathbb{R}^n \times \mathbb{R}_{++}$ be convex. The perspective function $P(x, t) = x/t$ maps $(x, t) \mapsto x/t$. Then the image of C under this map is also convex. This operation is used in geometric programming and fractional programming.

These operations(intersection, affine image, preimage, Cartesian product, perspective transformation) also preserve convexity in **Banach** and **Hilbert spaces**, providedd the mapping are continuous and the topology is taken into account.

Separating and Supporting Hyperplanes

Separating Hyperplane Theorem: If $A \cap B = \emptyset$ and both sets are convex, then:

$$\exists\, a, b \text{ such that } a^T x \leq b,\ \forall x \in A, \text{ and } a^T y \geq b,\ \forall y \in B$$

This means there's a hyperplane that separates the two sets with no overlap.
Supporting Hyperplane Theorem: For any $x_0 \in \partial C$, there exists a hyperplane that passes through x_0 and keeps C entirely on one side. Supporting hyperplanes are used in subgradient and duality theory.
In i**infinite-dimensional spaces**, the Hahn–Banach Separation Theorem generalizes the separating hyperplane theorem: any closed convex set and a point outside it can be separated by a continuous linear functional. Likewise, at every boundary point of a closed convex set in a Banach space, there exists a supporting hyperplane defined by a continuous linear functional.

Dual Cones and Generalized Inequalities

Dual Cone: Given a cone K, the dual cone is:

$$K^* = \{y \in \mathbb{R}^n \mid y^T x \geq 0,\ \forall x \in K\}$$

Examples:

- $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$
- $(S^n_+)^* = S^n_+$

Dual cones are used in defining generalized inequalities, constructing dual problems in optimization, and proving optimality conditions.
In functional analysis, the dual cone K^* is defined in the continuous dual space X^* of a Banach Space X:

$$K^* := \{f \in X^* | f(x) \geq 0 \forall x \in K\}$$

This is foundational in duality theory and generalized inequalities in infinite-dimensional spaces.

Optimisation and the Lagrange Dual Function

The optimization problem, generalized to a functional analysis setting, can be written as:

$$\begin{array}{rll} \text{minimize} & J(u) & \\ \text{subject to} & G_i(u) \leq 0, & i = 1, \ldots, m \\ & E_j(u) = 0, & j = 1, \ldots, p, \end{array} \tag{1}$$

with variable $u \in \mathcal{X}$, where $\mathcal{X}$ is a real Banach space. Here, $J : \mathcal{X} \to \mathbf{R}$ is the objective **functional**, $G_i : \mathcal{X} \to \mathbf{R}$ are inequality constraint **functionals**, and $E_j : \mathcal{X} \to \mathbf{R}$ are equality constraint **functionals**. We assume its domain $\mathcal{D} = \operatorname{dom}(J) \cap \bigcap_{i=1}^m \operatorname{dom}(G_i) \cap \bigcap_{j=1}^p \operatorname{dom}(E_j)$ is nonempty, and denote the optimal value by p^*. We do not initially assume convexity of these functionals.

The Lagrangian

We define the Lagrangian $L : \mathcal{X} \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ associated with the problem (1) as:

$$L(u, \lambda, \nu) = J(u) + \sum_{i=1}^m \lambda_i G_i(u) + \sum_{j=1}^p \nu_j E_j(u)$$

with $\operatorname{dom}(L) = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$. The basic idea in Lagrangian duality is to take the constraints into account by augmenting the objective functional with a weighted sum of the constraint functionals.

We refer to λ_i as the Lagrange multiplier associated with the i^{th} inequality constraint $G_i(u) \leq 0$, and similarly ν_j is the Lagrange multiplier associated with the j^{th} equality constraint $E_j(u) = 0$. The vectors $\lambda = (\lambda_1, \ldots, \lambda_m)$ and $\nu = (\nu_1, \ldots, \nu_p)$ are also called the dual variables.

The Lagrange Dual Function

We define the **Lagrange dual function** (or just **dual function**) $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ as the minimum value (infimum) of the Lagrangian over u: for $\lambda \in \mathbf{R}^m$, $\nu \in \mathbf{R}^p$,

$$g(\lambda, \nu) = \inf_{u \in \mathcal{D}} L(u, \lambda, \nu) = \inf_{u \in \mathcal{D}} \left(J(u) + \sum_{i=1}^m \lambda_i G_i(u) + \sum_{j=1}^p \nu_j E_j(u) \right).$$

When the Lagrangian is unbounded below in u, the dual function takes on the value $-\infty$. Since the dual function is the pointwise infimum of a family of **functionals of** (λ, ν) **that are affine** (for each fixed u, $L(u, \cdot, \cdot)$ is an affine functional of (λ, ν)), g is concave, even when the original problem (functionals J, G_i, E_j) is not convex.

Lower bounds on optimal value

The dual function yields lower bounds on the optimal value p^* of the problem (1): For any $\lambda \succeq 0$ (meaning $\lambda_i \geq 0$ for all i) and any ν, we have

$$g(\lambda, \nu) \leq p^*. \tag{2}$$

This important property is easily verified. Suppose $\tilde{u}$ is a feasible point for problem (1), i.e., $G_i(\tilde{u}) \le 0$ and $E_j(\tilde{u}) = 0$, and $\lambda \ge 0$. Then we have

$$\sum_{i=1}^{m} \lambda_i G_i(\tilde{u}) + \sum_{j=1}^{p} \nu_j E_j(\tilde{u}) \le 0,$$

since each term in the first sum is nonpositive (as $\lambda_i \ge 0, G_i(\tilde{u}) \le 0$), and each term in the second sum is zero. Therefore,

$$L(\tilde{u}, \lambda, \nu) = J(\tilde{u}) + \sum_{i=1}^{m} \lambda_i G_i(\tilde{u}) + \sum_{j=1}^{p} \nu_j E_j(\tilde{u}) \le J(\tilde{u}).$$

Hence,

$$g(\lambda, \nu) = \inf_{u \in \mathcal{D}} L(u, \lambda, \nu) \le L(\tilde{u}, \lambda, \nu) \le J(\tilde{u}).$$

Since $g(\lambda, \nu) \le J(\tilde{u})$ holds for every feasible point $\tilde{u}$, the inequality (2) follows by taking the infimum of $J(\tilde{u})$ over all feasible $\tilde{u}$, which is p^*.

The dual function gives a nontrivial lower bound on p^* only when $\lambda \ge 0$ and $(\lambda, \nu) \in \mathrm{dom}(g)$, i.e., $g(\lambda, \nu) > -\infty$. We refer to a pair (λ, ν) with $\lambda \ge 0$ and $(\lambda, \nu) \in \mathrm{dom}(g)$ as **dual feasible**.

The Lagrange dual problem

For each pair (λ, ν) with $\lambda \ge 0$, the Lagrange dual function gives us a lower bound on the optimal value p^* of the optimization problem. A natural question is: What is the best lower bound that can be obtained from the Lagrange dual function?

This leads to the optimization problem:

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \ge 0. \end{array} \tag{3}$$

This problem is called the **Lagrange dual problem** associated with problem (1). In this context the original problem is sometimes called the **primal problem**. We refer to $(\lambda \cdot \nu^)$ as **dual optimal** or **optimal Lagrange multipliers** if they are optimal for problem (3).

The Lagrange dual problem is a convex optimization problem, since the objective $g(\lambda, \nu)$ to be maximized is concave and the constraint $\lambda \ge 0$ defines a convex set. This is the case whether or not the primal problem is convex.

Weak Duality

The optimal value of the Lagrange dual problem, which we denote d, is, by definition, the best lower bound on p that can be obtained from the Lagrange dual function. Thus, we have

$$d^* \le p^*$$

which holds even if the original problem is not convex. This property is called **weak duality**.

We refer to the difference $p^* - d$ as the *optimal duality gap* of the original problem. The optimal duality gap is always nonnegative. The bound d can sometimes be used to find a lower bound on the optimal value of a problem that is difficult to solve, since the dual problem is always convex and may be easier to solve.

Strong Duality and Slater's Condition

If the equality

$$d^* = p^* \tag{4}$$

holds, i.e., the optimal duality gap is zero, then we say that *strong duality* holds. This means that the best bound that can be obtained from the Lagrange dual function is tight.

Strong duality does not, in general, hold. But if the primal problem is convex, i.e., of the form

$$\begin{aligned} \text{minimize} \quad & J(u) \\ \text{subject to} \quad & G_i(u) \leq 0, \quad i = 1, \dots, m, \\ & Au = b, \end{aligned} \tag{5}$$

where $J, G_1, \dots, G_m$ are convex **functionals**, and the equality constraints are affine, represented by $Au = b$ (where $A : \mathcal{X} \to \mathcal{Z}$ is a bounded linear operator, $\mathcal{Z}$ is a Banach space, and $b \in \mathcal{Z}$), we usually (but not always) have strong duality. There are many results that establish conditions on the problem, beyond convexity, under which strong duality holds. These conditions are called *constraint qualifications*.

Slater's Condition

One simple constraint qualification is *Slater's condition*: For a convex problem as in (5), there exists a point $u_0 \in \operatorname{relint} \mathcal{D}$ (the relative interior of the domain $\mathcal{D}$) such that

$$G_i(u_0) < 0, \quad i = 1, \dots, m, \quad Au_0 = b. \tag{6}$$

Such a point is sometimes called *strictly feasible*. Slater's theorem states that if Slater's condition holds (and the problem is convex), then strong duality holds. (Note: In infinite-dimensional spaces, the definition of $\operatorname{relint} \mathcal{D}$ and the precise form of Slater's condition may require more care, sometimes using $\operatorname{int} \mathcal{D}$ or specific conditions on the affine constraints).

Optimality Conditions

Complementary Slackness

Suppose that the primal and dual optimal values are attained and equal (so strong duality holds). Let u be a primal optimal point and $(\lambda^{,}\nu^*)$ be a dual optimal point. This means that

$$\begin{aligned} J(u^{)} &= g(\lambda^{,}\nu^*) \\ &= \inf_{u \in \mathcal{D}} \left(J(u) + \sum_{i=1}^{m} \lambda_i^* G_i(u) + \sum_{j=1}^{p} \nu_j^* E_j(u) \right) \\ &\leq J(u^{)} + \sum_{i=1}^{m} \lambda_i^G{}_i(u^{)} + \sum_{j=1}^{p} \nu_j^E{}_j(u^*) \\ &\leq J(u^*). \end{aligned}$$

The first line states that the optimal duality gap is zero. The second line is the definition of the dual function. The third line follows since the infimum of the Lagrangian over u is less than or equal to its value at $u = u$. The last inequality follows from $\lambda_i^{\geq}0, G_i(u^{)} \leq 0$ for $i = 1, \ldots, m$, and $E_j(u^{)} = 0$ for $j = 1, \ldots, p$. We conclude that the two inequalities in this chain must hold with equality.

From this, we can draw several interesting conclusions. For example, since the inequality in the third line is an equality, we conclude that u minimizes $L(u, \lambda^{,}\nu^*)$ over $u \in \mathcal{D}$.

Another important conclusion is that

$$\sum_{i=1}^{m} \lambda_i^* G_i(u^*) = 0.$$

Since each term $\lambda_i^* G_i(u^{)}$ in this sum is nonpositive (as $\lambda_i^{\geq}0$ and $G_i(u^*) \leq 0$), we must have

$$\lambda_i^* G_i(u^*) = 0, \quad i = 1, \ldots, m. \tag{7}$$

This condition is known as *complementary slackness*. It holds for any primal optimal u and any dual optimal $(\lambda^{,}\nu^*)$ (when strong duality holds). We can express the complementary slackness condition as

$$\lambda_i^* > 0 \implies G_i(u^*) = 0,$$

or, equivalently,

$$G_i(u^{)} < 0 \implies \lambda_i^{=}0.$$

Roughly speaking, this means the i-th optimal Lagrange multiplier is zero unless the i-th constraint is active at the optimum.

KKT Optimality Conditions

We now assume that the functionals $J, G_1, \ldots, G_m, E_1, \ldots, E_p$ are Fréchet differentiable (and therefore have open domains for their arguments in $\mathcal{X}$). We make no assumptions yet about convexity.

KKT conditions for nonconvex problems

As above, let u and $(\lambda^{,}\nu^{)}$ be any primal and dual optimal points with zero duality gap. Since u minimizes $L(u, \lambda^{,}\nu^{)}$ over $u \in \mathcal{D}$, it follows that if u is an interior point of $\mathcal{D}$ (or if constraints are handled appropriately for boundary points), its Fréchet derivative with respect to u must vanish at u. This derivative, an element of the dual space $\mathcal{X}^*$, is:

$$\mathrm{D}_u L(u^{,}\lambda^{,}\nu^{)} = \mathrm{D}J(u^{)} + \sum_{i=1}^{m} \lambda_i^* \mathrm{D}G_i(u^{)} + \sum_{j=1}^{p} \nu_j^{\mathrm{D}} E_j(u^*) = 0_{\mathcal{X}^*}.$$

Here, $\mathrm{D}J(u^{)})$, $\mathrm{D}G_i(u^{)})$, and $\mathrm{D}E_j(u^*)$ are the Fréchet derivatives (which are continuous linear functionals in $\mathcal{X}^*$). Thus we have the *Karush-Kuhn-Tucker (KKT) conditions*:

$$\begin{aligned} G_i(u^*) &\leq 0, & i &= 1,\dots,m \\ E_j(u^*) &= 0, & j &= 1,\dots,p \\ \lambda_i^* &\geq 0, & i &= 1,\dots,m \\ \lambda_i^* G_i(u^*) &= 0, & i &= 1,\dots,m \end{aligned}$$

$$\mathrm{D}J(u^{)}) + \sum_{i=1}^{m} \lambda_i^{\mathrm{D}} G_i(u^{)}) + \sum_{j=1}^{p} \nu_j^{\mathrm{D}} E_j(u^*) = 0_{\mathcal{X}^*}. \tag{8}$$

To summarize, for any optimization problem with differentiable functionals for which strong duality obtains, any pair of primal and dual optimal points must satisfy the KKT conditions (8).

Remark on general operator constraints: If constraints are given by operators $G : \mathcal{X} \to \mathcal{Y}$ (with $G(u) \preceq_{K_{\mathcal{Y}}} 0$) and $E : \mathcal{X} \to \mathcal{Z}$ (with $E(u) = 0_{\mathcal{Z}}$), where $K_{\mathcal{Y}}$ is a cone in $\mathcal{Y}$, then Lagrange multipliers λ are in $\mathcal{Y}^*$ (specifically $\lambda^{\in} K_{\mathcal{Y}}$, the dual cone) and $\nu^{\in} \mathcal{Z}^*$. The stationarity condition becomes $\mathrm{D}J(u^{)}) + (\mathrm{D}G(u^{)}))^* \lambda^* + (\mathrm{D}E(u^{)}))^{\nu *} = 0_{\mathcal{X}^*}$, where $(\mathrm{D}G(u^{)}))^{:} \mathcal{Y}^* \to \mathcal{X}^*$ and $(\mathrm{D}E(u^{)}))^{:} \mathcal{Z}^* \to \mathcal{X}^*$ are the adjoint operators of the Fréchet derivatives $\mathrm{D}G(u^{)} : \mathcal{X} \to \mathcal{Y}$ and $\mathrm{D}E(u^{)} : \mathcal{X} \to \mathcal{Z}$.

KKT conditions for convex problems

When the primal problem is convex, the KKT conditions are also sufficient for the points to be primal and dual optimal. In other words, if J, G_i are convex functionals and E_j are affine functionals (i.e., $E_j(u) = \langle \ell_j^{;} u \rangle - c_j$ for some $\ell_j^{\in} \mathcal{X}^*, c_j \in \mathbf{R}$), and if $\tilde{u}, \tilde{\lambda}, \tilde{\nu}$ are any points that satisfy the KKT conditions:

$$\begin{aligned} G_i(\tilde{u}) &\leq 0, & i &= 1,\dots,m \\ E_j(\tilde{u}) &= 0, & j &= 1,\dots,p \\ \tilde{\lambda}_i &\geq 0, & i &= 1,\dots,m \\ \tilde{\lambda}_i G_i(\tilde{u}) &= 0, & i &= 1,\dots,m \end{aligned}$$

$$\mathrm{D}J(\tilde{u}) + \sum_{i=1}^{m} \tilde{\lambda} i \mathrm{D}G_i(\tilde{u}) + \sum j = 1^p \tilde{\nu} j \mathrm{D}E_j(\tilde{u}) = 0\mathcal{X}^*,$$

then $\tilde{u}$ and $(\tilde{\lambda}, \tilde{\nu})$ are primal and dual optimal, with zero duality gap.

To see this, note that the first two conditions state that $\tilde{u}$ is primal feasible. Since $\tilde{\lambda}_i \geq 0$ and G_i are convex, and E_j are affine, the Lagrangian $L(u, \tilde{\lambda}, \tilde{\nu})$ is convex in u. The last KKT condition states that its Fréchet derivative with respect to u vanishes at $u = \tilde{u}$. For a convex functional, a vanishing derivative implies that $\tilde{u}$ is a global minimizer. So, $\tilde{u}$ minimizes $L(u, \tilde{\lambda}, \tilde{\nu})$ over u. From this we conclude that

$$\begin{aligned} g(\tilde{\lambda}, \tilde{\nu}) &= L(\tilde{u}, \tilde{\lambda}, \tilde{\nu}) \\ &= J(\tilde{u}) + \sum_{i=1}^{m} \tilde{\lambda} i G_i(\tilde{u}) + \sum j = 1^p \tilde{\nu}_j E_j(\tilde{u}) \\ &= J(\tilde{u}), \end{aligned}$$

where in the last line we use $E_j(\tilde{u}) = 0$ (from primal feasibility) and $\tilde{\lambda}_i G_i(\tilde{u}) = 0$ (from complementary slackness). This shows that $J(\tilde{u}) = g(\tilde{\lambda}, \tilde{\nu})$. Since $\tilde{u}$ is primal feasible and $(\tilde{\lambda}, \tilde{\nu})$ is dual feasible (as $\tilde{\lambda} \geq 0$), and the duality gap $J(\tilde{u}) - g(\tilde{\lambda}, \tilde{\nu})$ is zero, they must be primal and dual optimal.

In summary, for any convex optimization problem with differentiable objective and constraint functionals (where equality constraints are affine), any points that satisfy the KKT conditions are primal and dual optimal, and have zero duality gap.

If a convex optimization problem with differentiable functionals satisfies Slater's condition, then the KKT conditions provide necessary and sufficient conditions for optimality: Slater's condition implies that the optimal duality gap is zero and the dual optimum is attained, so u is optimal if and only if there exist (λ, ν) that, together with u, satisfy the KKT conditions.

The KKT conditions play an important role in optimization, particularly in infinite-dimensional settings such as optimal control and calculus of variations. In a few special cases, it is possible to solve the KKT conditions analytically. More generally, many algorithms for convex optimization are conceived as, or can be interpreted as, methods for solving the KKT conditions.

Infinite-Dimensional Example: Minimizing a Quadratic Functional on l^2

Let $\mathcal{X} = l^2$, the Hilbert space of square-summable real sequences. An element $u \in l^2$ is a sequence $u = (u_1, u_2, u_3, \dots)$ such that its l^2-norm $\|u\|l^2 = \left(\sum k = 1^\infty u_k^2\right)^{1/2}$ is finite. The inner product in l^2 is $(u, v)l^2 = \sum k = 1^\infty u_k v_k$.

Primal Problem: Let $a = (a_k)_{k=1}^\infty$ be a fixed sequence of positive real numbers such that $a_k \geq \delta > 0$ for some δ and for all k. Let $b = (b_k)_{k=1}^\infty$ be another fixed sequence in l^2. Let C be a real constant.

We aim to:

$$
\begin{aligned}
\text{minimize} \quad & J(u) = \frac{1}{2} \sum_{k=1}^{\infty} a_k u_k^2 \\
\text{subject to} \quad & E(u) = \sum_{k=1}^{\infty} b_k u_k - C = 0.
\end{aligned}
$$

Here:

- $J : l^2 \to \mathbf{R}$ is the objective **functional**. It maps an infinite sequence $u \in l^2$ to a real number. The condition $a_k \geq \delta > 0$ ensures $J(u)$ is well-defined and strictly convex on l^2.
- $E : l^2 \to \mathbf{R}$ is the equality constraint **functional**. By the Cauchy-Schwarz inequality, $|\sum b_k u_k| \leq \|b\|l^2 \|u\|l^2 < \infty$, so this functional is well-defined and continuous (linear) on l^2.

Lagrangian Functional: The Lagrangian $L : l^2 \times \mathbf{R} \to \mathbf{R}$ is given by:

$$
L(u, \nu) = J(u) + \nu E(u) = \frac{1}{2} \sum_{k=1}^{\infty} a_k u_k^2 + \nu \left(\sum_{k=1}^{\infty} b_k u_k - C \right),
$$

where $\nu \in \mathbf{R}$ is the Lagrange multiplier associated with the equality constraint $E(u) = 0$.

KKT Conditions: We seek a primal optimal $u^* \in l^2$ and a dual optimal $\nu^* \in \mathbf{R}$. Since the problem is convex and the constraint is affine, and assuming a feasible point exists, strong duality is expected. The KKT conditions are:

1. **Primal Feasibility:** $E(u^{)}) = \sum_{k=1}^{\infty} b_k u_k^{-} C = 0$.

2. **Stationarity of the Lagrangian:** The Fréchet derivative of L with respect to u, $\mathrm{D}_u L(u^{\cdot}\nu^{)})$, must be the zero functional in $(l^2)^* \cong l^2$. For sequence spaces, this implies that the partial derivative of L with respect to each component u_k must be zero at $(u^{\cdot}\nu^{)})$:

$$\frac{\partial L}{\partial u_k}(u^{\cdot}\nu^{)}) = a_k u_k^* + \nu^* b_k = 0 \quad \text{for each } k = 1, 2, \dots$$

From the stationarity condition (2), we obtain an expression for each u_k^*:

$$a_k u_k^* + \nu^* b_k = 0$$

Since $a_k > 0$, we can solve for u_k^*:

$$u_k^* = -\frac{\nu^* b_k}{a_k}. \tag{9}$$

Next, we substitute this expression for u_k^* into the primal feasibility constraint (1):

$$\sum_{k=1}^{\infty} b_k u_k^* = C$$

$$\sum_{k=1}^{\infty} b_k \left(-\frac{\nu^* b_k}{a_k}\right) = C$$

$$-\nu^* \sum_{k=1}^{\infty} \frac{b_k^2}{a_k} = C.$$

Let $S = \sum_{k=1}^{\infty} \frac{b_k^2}{a_k}$. Since $b \in l^2$ (so $\sum b_k^2 < \infty$) and $a_k \geq \delta > 0$, $S \leq \frac{1}{\delta} \sum b_k^2 < \infty$, so the sum S converges. Assuming $S \neq 0$ (which is true if at least one $b_k \neq 0$), we can solve for ν^*:

$$\nu^* = -\frac{C}{S}.$$

Finally, substituting ν back into equation (9), we find the components of the optimal sequence u:

$$u_k^* = -\frac{b_k}{a_k}\left(-\frac{C}{S}\right) = \frac{C b_k}{a_k S}.$$

The optimal solution is the sequence $u^* = \left(\frac{C b_k}{a_k S}\right)_{k=1}^{\infty}$. We must verify that $u^* \in l^2$:

$$\|u^{\|} l^{2^2} = \sum k = 1^{\infty} (u_k^{)2} = \sum_{k=1}^{\infty} \left(\frac{C b_k}{a_k S}\right)^2 = \frac{C^2}{S^2} \sum_{k=1}^{\infty} \frac{b_k^2}{a_k^2}.$$

Since $a_k \geq \delta > 0$, $a_k^2 \geq \delta^2 > 0$, so $\frac{1}{a_k^2} \leq \frac{1}{\delta^2}$. Thus, $\sum \frac{b_k^2}{a_k^2} \leq \frac{1}{\delta^2} \sum b_k^2 = \frac{1}{\delta^2} \|b\|_{l^2}^2 < \infty$. Therefore, $u^* \in l^2$.

Classical Portfolio Optimization Problem

We consider a classical portfolio problem with n assets or stocks held over a period of time. Let x_i denote the amount of asset i held throughout the period, with x_i in dollars, at the price at the beginning of the period. A normal long position in asset i corresponds to $x_i > 0$; a short position in asset i (i.e., the obligation to buy the asset at the end of the period) corresponds to $x_i < 0$. Let p_i denote the relative price change of asset i over the period, i.e., its change in price over the period divided by its price at the beginning of the period.

The overall return on the portfolio is

$$r = p^T x,$$

given in dollars. The optimization variable is the portfolio vector $x \in \mathbb{R}^n$.

A wide variety of constraints on the portfolio can be considered. The simplest set of constraints is that $x_i \geq 0$ (i.e., no short positions) and $1^T x = B$ (i.e., the total budget to be invested is B, which is often taken to be one).

We take a stochastic model for price changes: $p \in \mathbb{R}^n$ is a random vector, with known mean $\mathbb{E}[p] = p$ and covariance Σ. Therefore with portfolio $x \in \mathbb{R}^n$, the return r is a (scalar) random variable with mean $p^T x$ and variance $x^T \Sigma x$. The choice of portfolio x involves a trade-off between the mean of the return, and its variance.

The classical portfolio optimization problem, introduced by Markowitz, is the quadratic program:

$$\begin{array}{ll} \text{minimize} & x^T \Sigma x \\ \text{subject to} & p^T x \geq r_{\min}, \\ & 1^T x = 1, \\ & x \succeq 0, \end{array}$$

where x, the portfolio, is the variable. Here we find the portfolio that minimizes the return variance (associated with the risk of the portfolio), subject to achieving a minimum acceptable mean return $r_{\min}$, and satisfying the portfolio budget and no-shorting constraints.

4 Applications in Portfolio & Risk Management

Convex optimization and functional analysis provide the mathematical backbone for many techniques used in finance, particularly in portfolio optimization, asset allocation, and risk management. These tools allow us to frame and solve problems where the goal is to make the best financial decision under constraints and uncertainty.

4.1 Convex Optimization in Portfolio Theory

One of the classical applications is the **Mean-Variance Portfolio Optimization**, introduced by Harry Markowitz. The objective is to minimize the portfolio variance (a measure of risk) for a given expected return:

$$
\begin{aligned}
\min_{w\in\mathbb{R}^n} \quad & w^\top \Sigma w \\
\text{subject to} \quad & \mu^\top w \geq r \\
& \mathbf{1}^\top w = 1 \\
& w \geq 0
\end{aligned}
\tag{10}
$$

Here, w is the vector of portfolio weights, Σ is the covariance matrix of asset returns, μ is the expected return vector, r is the desired return, and the constraint $\mathbf{1}^\top w = 1$ ensures full capital allocation.

This is a quadratic convex optimization problem because the objective function is convex (positive semi-definite Σ) and the constraints are linear.

4.2 Risk Management using Convex Risk Measures

Risk measures such as **Conditional Value at Risk (CVaR)** can be formulated using convex optimization. CVaR is defined as the expected loss in the worst α-fraction of cases:

$$
\mathrm{CVaR}_\alpha(L) = \min_{\eta\in\mathbb{R}} \left\{ \eta + \frac{1}{1-\alpha}\mathbb{E}[(L-\eta)^+] \right\}
\tag{11}
$$

where L is the loss random variable and $(x)^+ = \max(x, 0)$. This representation is convex and can be minimized to manage downside risk more effectively than traditional Value-at-Risk (VaR).

4.3 Utility Maximization

Another major application is the maximization of expected utility of terminal wealth. Assuming a concave utility function $U(\cdot)$ and initial capital x_0, the problem is:

$$
\max_{w} \quad \mathbb{E}[U(w^\top R)]
\tag{12}
$$

Here, R is the random return vector. Under certain assumptions, this optimization problem is convex and solvable using tools from functional analysis.

4.4 Robust Portfolio Optimization

Market parameters like expected returns and covariances are often uncertain. Robust optimization handles this by optimizing the worst-case outcome over an uncertainty set:

$$
\min_{w} \max_{\mu\in\mathcal{U}_\mu, \Sigma\in\mathcal{U}_\Sigma} w^\top \Sigma w - \lambda \mu^\top w
\tag{13}
$$

This nested optimization ensures that the solution is stable even when data is noisy or misspecified. Convex optimization tools such as ellipsoidal uncertainty sets and duality theory are employed here.

4.5 Support Vector Machines in Portfolio Optimization

Support Vector Machines (SVM) are widely used in portfolio management for classification, return prediction, and risk assessment. SVM is a supervised learning algorithm that finds an optimal hyperplane to separate data into distinct classes. In the context of portfolio optimization, it has several applications:

- **Asset Classification:** SVM can be used to classify assets into different categories (e.g., high-risk vs. low-risk) based on their historical returns and volatility. Mathematically, this is achieved by solving:

$$\min_{w,b} \|w\|^2 \tag{14}$$

subject to the constraints:

$$y_i(w^\top x_i + b) \geq 1, \quad i = 1, \ldots, n \tag{15}$$

where x_i is the feature vector of asset i, $y_i \in \{-1, +1\}$ indicates the class (e.g., high-risk or low-risk), w is the weight vector, and b is the bias term.

- **Return Prediction:** SVM can be applied as Support Vector Regression (SVR) to predict asset returns:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{16}$$

subject to:

$$r_i - (w^\top x_i + b) \leq \epsilon + \xi_i, \quad (w^\top x_i + b) - r_i \leq \epsilon + \xi_i^* \tag{17}$$

where r_i is the actual return, ϵ is the tolerance, ξ_i, ξ_i^* are slack variables, and C is the regularization parameter.

- **Risk Assessment:** By using kernel SVM, we can cluster assets with similar risk characteristics, allowing for more effective diversification. The kernel SVM is formulated as:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j) \tag{18}$$

subject to:

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C \tag{19}$$

where $\kappa(x_i, x_j)$ is the kernel function, enabling non-linear classification.

In portfolio management, SVM provides a powerful tool for identifying asset categories, predicting returns, and managing risk through efficient classification and regression techniques.

4.6 Functional Analysis in Finance

Functional analysis becomes important when modeling financial systems over time, for example, in continuous-time finance (e.g., Black-Scholes theory). Asset prices and wealth trajectories are often modeled as functions in infinite-dimensional Banach or Hilbert spaces.

Linear operators on these function spaces, norms, and dual spaces help define and solve optimal control problems in dynamic asset allocation.

4.7 Summary of Applications

- **Convex Optimization**: Used in asset allocation, mean-variance optimization, CVaR minimization, robust portfolio models.
- **Functional Analysis**: Applied in modeling financial processes over time, dynamic portfolio theory, and pricing in continuous-time models.
- **Duality**: Offers insight into pricing mechanisms, shadow prices, and optimal trade-offs in resource-constrained settings.

These mathematical tools are crucial not only for theoretical financial research but also for real-world applications such as robo-advisory platforms, high-frequency trading systems, and regulatory stress testing.

Extensions

Many extensions are possible. One standard extension is to allow short positions, i.e., $x_i < 0$. To do this, we introduce variables x_{long} and x_{short}, with

$$x_{\text{long}} \succeq 0, \quad x_{\text{short}} \succeq 0, \quad x = x_{\text{long}} - x_{\text{short}}, \quad 1^T x_{\text{short}} \leq \eta 1^T x_{\text{long}}.$$

The last constraint limits the total short position at the beginning of the period to some fraction η of the total long position at the beginning of the period.

As another extension, we can include linear transaction costs in the portfolio optimization problem. Starting from a given initial portfolio x_{init}, we buy and sell assets to achieve the portfolio x, which we then hold over the period as described above. We are charged a transaction fee for buying and selling assets, proportional to the amount bought or sold. To handle this, we introduce variables u_{buy} and u_{sell}, which determine the amount of each asset we buy and sell before the holding period. We have the constraints:

$$x = x_{\text{init}} + u_{\text{buy}} - u_{\text{sell}}, \quad u_{\text{buy}} \succeq 0, \quad u_{\text{sell}} \succeq 0.$$

We replace the simple budget constraint $1^T x = 1$ with the condition that the initial buying and selling, including transaction fees, involve zero net cash:

$$(1 - f_{\text{sell}}) 1^T u_{\text{sell}} = (1 + f_{\text{buy}}) 1^T u_{\text{buy}}.$$

Here the left-hand side is the total proceeds from selling assets, less the selling transaction fee, and the right-hand side is the total cost, including transaction fee, of buying assets. The constants $f_{\text{buy}} \geq 0$ and $f_{\text{sell}} \geq 0$ are the transaction fee rates for buying and selling (assumed the same across assets, for simplicity).

The problem of minimizing return variance, subject to a minimum mean return, and the budget and trading constraints, is a quadratic program with variables x, u_{buy}, and u_{sell}.

5 Results and Discussion

We explored how convexity provides mathematical tractability and how duality offers both theoretical insights and algorithmic benefits. In economics and finance, dual variables carry meaningful interpretations, and in machine learning, dual formulations reduce complexity and improve performance.

A key insight is that many real-world optimization problems can be formulated in a convex framework, enabling us to use powerful tools like the Karush-Kuhn-Tucker (KKT) conditions and strong duality.

6 Conclusion

This report demonstrated the relevance of Functional Analysis concepts — particularly convex sets, convex functions, and Lagrange duality — in a variety of applied contexts. These mathematical tools are indispensable in modern optimization, with broad applications in economic modeling, financial decision-making, and data-driven machine learning algorithms.

Future work could explore operator theory applications in reinforcement learning and the use of Banach space methods in financial derivatives pricing.

References

[1] Boyd, S., & Vandenberghe, L. (2004). *Convex Optimization.* Cambridge University Press.

[2] Luenberger, D. G. (1997). *Optimization by Vector Space Methods.* Wiley.

[3] Bertsekas, D. P. (1999). *Nonlinear Programming.* Athena Scientific.

[4] Kreyszig, E. (1989). *Introductory Functional Analysis with Applications.* Wiley.

[5] Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of Machine Learning.* MIT Press.