

UNIVERSIDAD AUTÓNOMA METROPOLITANA



OPTIMIZACIÓN ROBUSTA DE
PORTAFOLIOS.

TESIS

QUE PARA OBTENER EL GRADO DE

MAESTRA EN CIENCIAS MATEMÁTICAS APLICADAS E INDUSTRIALES

PRESENTA

MARÍA DE LOS ÁNGELES PÉREZ ROJO

ASESOR: DRA. PATRICIA SAAVEDRA BARRERA

12 de junio de 2019

Índice general

Introducción	1
1. Antecedentes teóricos	5
1.1. Optimización de portafolios	5
1.1.1. Portafolio de inversión	5
1.1.2. Riesgo	6
1.1.3. Rendimiento de un portafolio	6
1.2. Modelos de portafolio	8
1.2.1. Modelo de Markowitz	8
1.2.2. Modelos factoriales estadísticos	12
2. Optimización convexa	15
2.1. Conjuntos convexos	15
2.1.1. Operaciones que preservan convexidad	17
2.2. Funciones convexas	18
2.3. Problemas de optimización convexa	23
2.3.1. Transformación de un QCP en un SOCP	28
2.4. Dualidad	29
3. Selección de portafolios robustos	41
3.1. Selección del portafolio robusto de mínima varianza	44
3.2. Selección del portafolio robusto de máximo rendimiento	47
3.3. Conjuntos de incertidumbre	48

3.3.1. Conjunto de incertidumbre para Σ	48
3.3.2. Conjunto de incertidumbre para $\vec{\mu}$	50
3.4. Modelo factorial	52
4. Resultados numéricos	61
4.1. Conjuntos de incertidumbre finitos	63
4.2. Conjuntos de incertidumbre no finitos	64
4.2.1. Portafolio robusto de mínima varianza	65
4.2.2. Portafolio robusto de máximo rendimiento	67
Conclusiones	71
Bibliografía	75

Introducción

El problema de selección del portafolio óptimo consiste en asignar cierto capital inicial en un conjunto de activos disponibles con el propósito de maximizar el "rendimiento" de la inversión mientras se minimiza el "riesgo". El primer modelo matemático para obtener el portafolio óptimo fue introducido por Markowitz (1952-1959). En este modelo el "rendimiento" de la inversión es medido mediante el valor esperado del rendimiento del portafolio y el "riesgo" asociado se calcula mediante la varianza del rendimiento del portafolio. Este modelo (media-varianza) ha tenido un gran impacto en la modelación de mercados financieros, sin embargo muchos investigadores lo han dejado de lado debido a que es bastante susceptible a perturbaciones en los parámetros del modelo ya que estos están sujetos a errores estadísticos, por tanto los resultados del problema de optimización no es muy confiable.

Para ejemplificar esto suponga que usted es un inversionistas y que le interesa constituir un portafolio de inversión para la semana del 12 de noviembre usando datos del periodo del 2 de enero al 31 de octubre del 2017. Con esa información determina el portafolio. Después de dos semanas cuando conoce los datos reales observa que el portafolio que seleccionó quedo fuera de la frontera eficiente que hubiera obtenido con los datos reales, ¿qué se puede hacer ante esta situación en la que no conocemos con certeza el comportamiento que tendran los datos a futuro? Simplemente no hay manera de predecir sin error dicho comportamiento. En la Figura 1 se ejemplifica esto, como podemos observar el portafolio obtenido con

los datos del primer periodo (W_1) queda totalmente fuera de la frontera eficiente obtenida con los datos reales al 12 de noviembre (frontera eficiente 2).

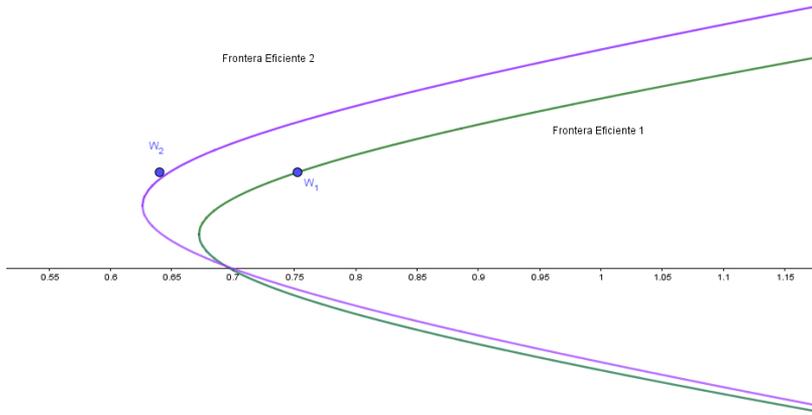


Figura 1

En esta tesis se propone una metodología alternativa que es robusta respecto a la incertidumbre en los parámetros, en particular se estudia a detalle la metodología desarrollada en el artículo *Robust selection problems* de Goldfarb e Iyengar [6], aunque en este artículo se hace énfasis en el modelo factorial sus ideas resultaran útiles para desarrollar nuestra propuesta.

Las perturbaciones en los parámetros son modelados como desconocidos pero acotados por ciertas *regiones de confianza* que resultan ser elipsoides en \mathbb{R}^n . Basados en este hecho los problemas de optimización se resuelven asumiendo el peor caso del comportamiento de estas perturbaciones. Este tipo de problemas son conocidos como problemas mini-max los cuales bajo ciertas condiciones pueden ser transformados en problemas de optimización cónica de segundo orden (SOCP), los cuales tienen un costo computacional similar al de un problema lineal con restricciones cuadráticas (QP).

Objetivo:

El objetivo principal de esta tesis es desarrollar la metodología necesaria para resolver problemas de optimización robusta aplicados a portafolios de inversión.

El contenido de esta tesis es el siguiente: En el Capítulo 1 se presentan ciertos conceptos importantes en la teoría de optimización de portafolios, tales como, portafolio de inversión, riesgo y rendimiento de un portafolio y modelo clásico de Markowitz. De igual manera se explica el modelo factorial. En el Capítulo 2 se desarrollan algunos temas de optimización convexa que son necesarios para el entendimiento del problema que se pretende resolver. En el capítulo 3 se desarrollan tres problemas de optimización robusta de portafolios: Problema robusto de mínima varianza, problema robusto de máximo rendimiento y el problema robusto desde el enfoque del modelo factorial. Además se plantean algunas técnicas para calcular los conjuntos de incertidumbre de los parámetros del modelo. En el Capítulo 4 se presentan los resultados obtenidos al resolver algunos problemas robustos numéricamente. Finalmente, se presentan las conclusiones del trabajo y la bibliografía empleada.

Capítulo 1

Antecedentes teóricos

1.1. Optimización de portafolios

1.1.1. Portafolio de inversión

En esta sección se definirá un portafolio de inversión, pero antes es necesario explicar el concepto de *acción*. Una acción es una proporción del capital económico de una institución y puede ser comprada o vendida a aquellos que desean obtener beneficios de las ganancias obtenidas por la empresa o institución que emite las acciones.

Un portafolio de inversión se compone de activos financieros que pueden ser acciones, bonos, etcetera. En particular en este trabajo nos ocuparemos de portafolios de acciones sin dividendos con los cuales se busca tener un mayor rendimiento y reducir el riesgo de una pérdida, en particular, de aquellas que tengan una magnitud grande. A continuación se explica el concepto de riesgo y las diversas formas en que éste se mide.

1.1.2. Riesgo

En el ámbito financiero existen diversos sucesos que pueden afectar los precios de las acciones que integran un portafolio de inversión, estos eventos conducen al concepto de riesgo. Un factor de riesgo es aquel que influye en la variación del precio de un activo. El problema principal al que se enfrenta un inversionista es el de conocer el peor escenario posible al que se expone cuando invierte en un instrumento de inversión con riesgo. Al invertir en un portafolio de activos el inversionista debe tomar en cuenta que hay gran cantidad de factores que afectan el precio de los activos y que además algunos activos son más riesgosos que otros. Un posible objetivo al invertir en un portafolio es obtener aquel que tenga menor riesgo con la mayor ganancia posible.

1.1.3. Rendimiento de un portafolio

Para abordar el problema de un portafolio de inversión es necesario estudiar el concepto de *rendimiento*. Supongamos que a un tiempo inicial t_0 deseamos invertir un monto inicial M en un portafolio con n activos. Los precios diarios de los activos en un periodo de tiempo previo a t_0 son conocidos. El conjunto de activos que se pretende comprar debe ser tal que minimice el riesgo del portafolio y produzca un rendimiento mayor o igual a una r^* dada. Para ejemplificar estos conceptos, estudiemos el problema de invertir en algún instrumento de inversión con una tasa de interés fija. Supongamos que invertimos un monto M_0 , al finalizar el periodo de inversión se tendrá un monto final M_1 , el cuál se puede calcular de la siguiente forma

$$M^1 = M^0 + rM^0 = (1 + r)M^0.$$

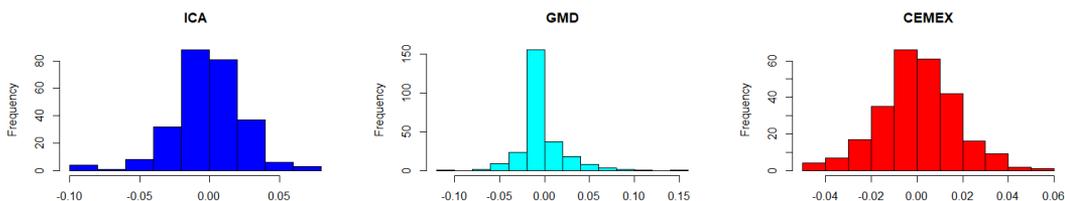
Al despejar r en la ecuación anterior obtenemos una expresión para el cálculo del rendimiento de la inversión

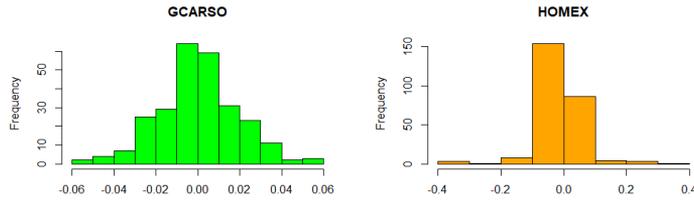
$$r = \frac{M^1 - M^0}{M^0}.$$

Cuando se trabaja con acciones, el rendimiento está relacionado con el cambio que hay en el precio de los activos y puede expresarse como

$$r = \frac{P^1 - P^0}{P^0},$$

donde P_0 y P_1 representan el precio del activo al inicio y al final del periodo, respectivamente. Es claro que el rendimiento que definimos para un instrumento de inversión con tasa fija coincide con el que se define para una acción. El rendimiento que se obtiene al invertir en un instrumento con tasa fija es una cantidad determinista ya que siempre sabremos la cantidad que vamos a obtener al finalizar el periodo. Al invertir en acciones el rendimiento obtenido es una cantidad aleatoria pues está sujeta a variaciones causadas por distintos factores. Por ser el rendimiento de una acción una variable aleatoria, es posible estimarlo mediante el valor esperado. Para el estudio de ciertos modelos es necesario asumir que los rendimientos de una acción son variables aleatorias con distribución normal. Para tener una idea de lo que esto significa analicemos el comportamiento histórico de cinco activos; por ejemplo, a través de su histograma de los rendimientos diarios obtenidos durante un año.





Observemos que en el caso de los histogramas de ICA, GMD y HOMEX los rendimientos no parecen tener una distribución normal a diferencia de CEMEX y GCARSO. En la mayoría de los casos no basta con un análisis gráfico para saber si los rendimientos de los activos tienen distribución normal, por lo que resulta conveniente realizar algún test de normalidad, por ejemplo, el test de Shapiro-Wilk. En el desarrollo de esta tesis la normalidad de los datos es un supuesto fuerte pues esto nos permitirá asociar la media y varianzas muestrales de los rendimientos al riesgo y rendimiento de un portafolio de inversión.

1.2. Modelos de portafolio

1.2.1. Modelo de Markowitz

Ahora estamos en posición de formular un problema matemático que conduce al portafolio de mínima varianza. Nuevamente vamos a suponer que tenemos n activos. Los valores esperados de los rendimientos son $\bar{r}_1, \bar{r}_2, \dots, \bar{r}_n$ y las covarianzas son σ_{ij} , para $i, j = 1, \dots, n$. Un portafolio está definido por n pesos w_1, w_2, \dots, w_n que representan el porcentaje del monto inicial invertido en el activo a_i . La suma de los w_i es 1 (se permiten valores negativos los cuales corresponden a las ventas en corto).

Para obtener el portafolio de mínima varianza fijamos algún valor arbitrario α y buscamos el portafolio de mínima varianza que tenga este rendimiento. Es importante aclarar que el modelo de Markowitz supone

que el rendimiento de cada activo tiene una distribución normal con media μ y varianza σ^2 . La varianza del portafolio se obtiene de la siguiente forma:

$$\begin{aligned}
 \sigma_p^2(\bar{w}) &= E[r_\pi - E(r_\pi)]^2 \\
 &= E\left[\sum_{i=1}^n w_i r_i - E\left(\sum_{i=1}^n w_i r_i\right)\right]^2 \\
 &= \sum_{i=1}^n \sum_{j=1}^n w_i w_j \text{Cov}(r_i, r_j) \\
 &\approx \sum_{i=1}^n \sum_{j=1}^n w_i w_j \overline{\text{Cov}}(r_i, r_j),
 \end{aligned}$$

donde $\overline{\text{Cov}}(r_i, r_j)$ representa la matriz de varianza-covarianza muestral, la cual a partir de ahora denotaremos como Σ .

Ahora que hemos definido la varianza de un portafolio estamos preparados para plantear el problema de obtener un portafolio con varianza mínima, es decir, el de menor riesgo. La forma matricial del problema del portafolio se formula a continuación

$$\begin{aligned}
 &\text{Min } \frac{1}{2} \bar{w}^t \Sigma \bar{w} \\
 &\text{s. a } \bar{w}^t \bar{r} \geq \alpha, \\
 &\quad \bar{w} \cdot \vec{1} = 1,
 \end{aligned}$$

donde \bar{r} es del vector de las medias muestrales de los rendimientos de los activos. Decimos que un portafolio es factible si es solución de este problema.

Solución del modelo de Markowitz

Se puede encontrar la solución de este problema (bajo el supuesto que se permiten las ventas en corto) con los multiplicadores de Lagrange dado que las restricciones son de igualdad. El Lagrangiano del problema es el siguiente:

$$L(\lambda, \mu) = \frac{1}{2} \vec{w}^t \Sigma \vec{w} + \lambda (\vec{w} \cdot \vec{1} - 1) + \mu (\vec{w}^t \cdot \vec{r} - \alpha).$$

Para obtener el portafolio óptimo se calculan las parciales del Lagrangiano respecto a \vec{w} , λ y μ . Al igualar las parciales a 0 se obtiene el siguiente sistema de ecuaciones.

$$\begin{aligned} \Sigma \vec{w} &= \lambda + \mu \vec{r}, \\ \vec{w}^t \cdot \vec{1} &= 1, \\ \vec{w}^t \cdot \vec{r} &= \alpha. \end{aligned}$$

Si suponemos que Σ es una matriz positiva definida, el sistema admite una solución única, así se obtiene los siguientes valores

$$\lambda(\alpha) = \frac{A\alpha - B}{\Delta}, \quad \mu(\alpha) = \frac{C - \alpha B}{\Delta},$$

donde $A = \mathbf{1}^t \Sigma^{-1} \mathbf{1}$, $B = \mathbf{1}^t \Sigma^{-1} \vec{r}$, $C = \vec{r}^t \Sigma^{-1} \vec{r}$ y $\Delta = AC - B^2$.

Puesto que $\Sigma \vec{w}^* = \mu \vec{1} + \lambda \vec{r}$, entonces

$$\begin{aligned} \sigma_*^2 &= \vec{w}^{*t} \Sigma \vec{w}^* \\ &= \vec{w}^* (\mu \vec{1} + \lambda \vec{r}) \\ &= \mu \vec{w}^{*t} \vec{1} + \lambda \vec{w}^{*t} \vec{r} \\ &= \mu + \lambda \alpha, \end{aligned}$$

sustituyendo λ y μ se tiene que

$$\sigma_*^2 = \frac{A\alpha^2 - 2B\alpha + C}{\Delta}.$$

Si derivamos esta expresión implícitamente respecto a α e igualamos a 0 se obtiene que $\alpha = B/A$ y $\sigma(\alpha) = 1/A$. El portafolio \vec{w}_{mv} asociado a $(\sigma(\alpha), \alpha)$ se conoce como portafolio de mínima varianza. Observemos que para este portafolio los multiplicadores de Lagrange correspondientes son $\lambda = 0$ y $\mu = \frac{1}{A}$. Si tomamos $\mu = 0$ se tiene que $\alpha_d = C/B$ con $B \neq 0$ por tanto si llamamos \vec{w}_d al portafolio correspondiente se obtiene que $\sigma_d = \frac{\sqrt{C}}{B}$. El teorema de dos fondos nos dice que cualquier portafolio admisible se puede escribir como una combinación lineal del portafolio de mínima varianza y el portafolio \vec{w}_d ya que

$$\vec{w} = \lambda B \vec{w}_d + \mu A \vec{w}_{mv}$$

A cada \vec{w} dado se puede asociar una volatilidad y un rendimiento (σ_w, r_w) , respectivamente. El conjunto formado por cada uno de estos puntos es una hipérbola cuyo vértice es justamente el portafolio de mínima varianza y se le conoce como conjunto *riesgo-rendimiento*. A la parte de la hipérbola que va de w_{mv} hacia arriba se le conoce como *frontera eficiente*, y contiene a aquellos portafolios cuyo rendimiento asociado a cierta volatilidad es máximo.

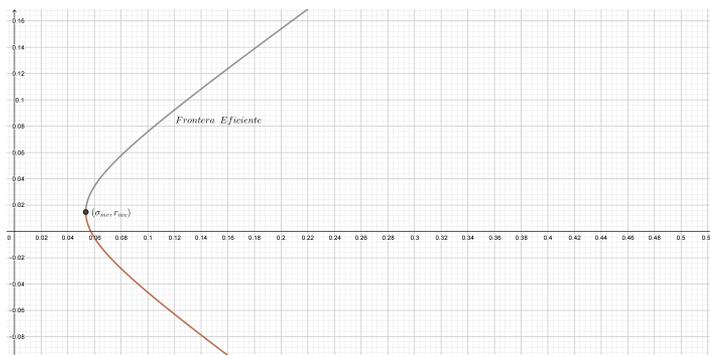


Figura 1.1: Conjunto *Riesgo-Rendimiento*

1.2.2. Modelos factoriales estadísticos

Un modelo factorial para rendimientos de activos tiene la siguiente forma general

$$\begin{aligned} R_{it} &= \alpha_i + \beta_{1i}f_{1t} + \beta_{2i}f_{2t} + \dots + \beta_{ki}f_{kt} + \varepsilon_{it} \\ &= \alpha_i\beta'_i f_t + \varepsilon_{it}. \end{aligned}$$

donde

- R_{it} es el rendimiento simple del activo i ($i = 1, 2, \dots, N$) en un periodo de tiempo t ($t = 1, 2, \dots, T$).
- f_{kt} es el k -ésimo factor ($k = 1, 2, \dots, K$).
- β_{ki} es la carga factorial o factor β para el activo i sobre el k -ésimo factor.
- ε_{it} es el error específico del activo i .

Supuestos del modelo factorial

1. Los términos de error específico de un activo ε_{it} , son no correlacionados con cada uno de los factores f_{kt} , es decir,

$$cov(f_{kt}, \varepsilon_{it}) = 0, \forall k, i, t.$$

2. Los términos de error de $R_{i,t}$ y $R_{j,s}$ son no correlacionados siempre que $i \neq j$, $t \neq s$, es decir,

$$cov(\varepsilon_{it}, \varepsilon_{js}) = 0.$$

En los modelos factoriales estadísticos los factores f_{kt} no son directamente observables por lo que deben extraerse de las observaciones de los rendimientos empleando técnicas estadísticas. Una de las principales técnicas para obtener los factores es el análisis factorial.

Análisis factorial

El análisis factorial tradicional asume una estructura ortogonal invariante en el tiempo

$$R_t = \mu + Bf_t + \varepsilon_t,$$
$$\text{var}(\varepsilon_t) = D,$$

donde D es una matriz diagonal con σ_{ij}^2 sobre la diagonal. Entonces la matriz de covarianza puede descomponerse como $\Omega = BB^t + D$.

Estimación

Realizar una estimación usando análisis factorial consiste de tres pasos:

1. Elegir los factores f_t
2. Estimar la matriz de carga factorial B y la matriz de residuos D .

Generalmente el problema del portafolio involucra grandes cantidades de activos por lo que una de las principales ventajas de utilizar esta técnica estadística es que nos permite hacer una reducción en la dimensión del problema. Es importante aclarar que los factores que se utilizan en el modelo factorial deben tener la capacidad de explicar gran parte del comportamiento de los activos financieros del mercado.

Capítulo 2

Optimización convexa

El objetivo de esta tesis es estudiar problemas de optimización aplicados a la teoría de portafolios, por lo cual es necesario definir ciertos conceptos que serán de gran utilidad para plantear problemas de optimización convexa los cuales se usarán para resolver problemas de selección de portafolios robustos.

2.1. Conjuntos convexos

Un conjunto es llamado convexo si este contiene a cualquier segmento de recta comprendido entre cualesquiera dos de sus puntos. A continuación se da una definición formal de conjunto convexo la cual resulta útil para probar la convexidad de un conjunto.

Definición 2.1.1 (Conjunto convexo). Un conjunto C es convexo si para todo $x_1, x_2 \in C$ y $0 \leq \theta \leq 1$ se cumple que

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

Un punto de la forma $\theta_1 x_1 + \dots + \theta_k x_k$ con $\theta_1 + \theta_2 + \dots + \theta_k = 1$ y $\theta_i \geq 0, i = 1, \dots, k$ es llamado, *combinación convexa* de puntos $x_1, x_2, \dots, x_k \in C$. Es posible probar que un conjunto es convexo si y sólo si contiene a cualquier combinación convexa de sus puntos.

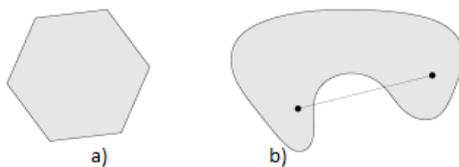


Figura 2.1: a) conjunto convexo, b) conjunto no convexo.

Definición 2.1.2 (Cono). Un conjunto C es llamado cono si para todo $x \in C$ y $\theta \geq 0$ se tiene $\theta x \in C$. Un conjunto C es cono convexo si es cono y es convexo, es decir, para cualquier $x_1, x_2 \in C$ y $\theta_1, \theta_2 \geq 0$ se tiene

$$\theta_1 x_1 + \theta_2 x_2 \in C.$$

Un punto de la forma $\theta_1 x_1 + \dots + \theta_k x_k$ con $\theta_1, \theta_2, \dots, \theta_k \geq 0$ es una combinación cónica de x_1, x_2, \dots, x_k . Si x_i está en un cono convexo C entonces toda combinación cónica de $x_i \in C$. Inversamente, un conjunto C es un cono convexo si y sólo si, este contiene a todas las combinaciones cónicas de sus elementos.

Ejemplos

A continuación se dan algunos ejemplos importantes de conjuntos convexos.

- Un segmento de línea es un conjunto convexo.
- Un rayo que tiene la forma $\{x_0 + \theta\nu \mid \theta \geq 0\}$ donde $\nu \neq 0$, es convexo y es cono convexo si su base $x_0 = 0$.
- El semiplano de un hiperplano es convexo.
- Una bola euclideana es un conjunto convexo.

- Un elipsoide, es decir, un conjunto de la forma

$$\mathcal{E} = \{x | (x - x_0)^t P^{-1} (x - x_0) \leq 1\}$$

con P simétrica y positiva definida, es un conjunto convexo.

- El cono "barquillo" asociado a la norma $\|\cdot\|$ es el conjunto

$$C = \{(x, t) | \|x\| \leq t\} \subseteq \mathbb{R}^{n+1}.$$

Este conjunto es convexo.

- Un poliedro se define como el conjunto solución de un número finito de igualdades y desigualdades lineales,

$$\mathcal{P} = \{x | a_j^T x \leq b_j, j = 1, 2, \dots, m, c_j^T x = d_j, j = 1, 2, \dots, p.\}$$

Este conjunto resulta de la intersección de un número finito de semiplanos e hiperplanos. Es fácil probar que un poliedro es un conjunto convexo.

2.1.1. Operaciones que preservan convexidad

Una operación que juega un papel importante en la teoría de convexidad es la intersección, debido a que preserva la convexidad de conjuntos convexos.

Si S_1, S_2 son conjuntos convexos entonces $S_1 \cap S_2$ también lo es. Esta propiedad se puede extender a un número infinito de conjuntos convexos, es decir, si S_α es convexo $\forall \alpha \in \mathcal{A}$, entonces $\bigcap_{\alpha \in \mathcal{A}} S_\alpha$ es convexo, por lo que los conos convexos también son cerrados bajo intersección.

Ejemplo 2.1.3. Un poliedro es intersección de hiperplanos y semiplanos los cuales son conjuntos convexos, por tanto es un conjunto convexo.

En el ejemplo anterior se estableció la convexidad de un conjunto expresando a este como la intersección de semiplano. Es fácil mostrar que todo conjunto convexo cerrado es una intersección (usualmente infinita) de semiplanos. De hecho, un conjunto convexo cerrado S es la intersección de todos los semiplanos que lo contienen, es decir,

$$S = \bigcap \{ \mathcal{H} \mid \mathcal{H} \text{ es un semiplano, } S \subseteq \mathcal{H} \}.$$

2.2. Funciones convexas

Definición 2.2.1. Una función $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es convexa si A es un conjunto convexo y si para cada $x, y \in A$ y $0 \leq \theta \leq 1$, se tiene que

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

Decimos que una función es estrictamente convexa si la desigualdad es estricta, siempre que $x \neq y$ y $0 < \theta < 1$. Decimos que f es concava si $-f$ es convexa, y estrictamente concava si $-f$ es estrictamente convexa.

Geométricamente una función convexa es aquella en la que el segmento de línea entre los puntos $(x, f(x))$ y $(y, f(y))$ está por encima de la gráfica de f . En la figura 2.3 se da un ejemplo de una función convexa en \mathbb{R} .



Figura 2.2: Gráfica de una función convexa en \mathbb{R} .

Ejemplos

Los siguientes son ejemplos de funciones convexas en \mathbb{R} .

1. Función exponencial: e^{ax} es convexa en \mathbb{R} para cualquier $a \in \mathbb{R}$.
2. Función potencia: x^a es convexa en \mathbb{R}_{++} cuando $a \geq 1$ o $a \leq 0$.
3. Potencia de valor absoluto: $|x|^p$ con $p \geq 1$, es convexa en \mathbb{R} .

Es fácil probar la convexidad de estas funciones usando la definición 2.2.1 o verificando que la segunda derivada es no negativa.

Lo siguientes son ejemplos de funciones convexas en \mathbb{R}^n

1. Toda norma en \mathbb{R}^n es convexa.
2. La función $f(x) = \max\{x_1, x_2, \dots, x_n\}$ es convexa en \mathbb{R}^n .
3. La función $f(x, y) = \log(e^{x_1} + e^{x_2} + \dots, e^{x_n})$ es convexa en \mathbb{R}^n y puede interpretarse como una aproximación diferenciable de la función \max , ya que

$$\max\{x_1, x_2, \dots, x_n\} \leq f(x) \leq \max\{x_1, x_2, \dots, x_n\} + \log n, \quad \forall x \in \mathbb{R}^n.$$

La convexidad de estos ejemplos puede ser probada verificando la desigualdad de la definición 2.2.1 o probando que el Hessiano es semidefinido positivo. Utilizando la definición de función convexa probaremos la convexidad de los ejemplos 1 y 2.

Convexidad de las normas:

Si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una norma y $0 \leq \theta \leq 1$, entonces

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

$$= \theta f(x) + (1 - \theta)f(y),$$

esto se sigue de la desigualdad del triángulo y de la homogeneidad de una norma.

Convexidad de la función $\text{máx}(\mathbf{x})$:

Para todo $0 \leq \theta \leq 1$, la función $f(x) = \text{máx}\{x_1, x_2, \dots, x_n\}$ satisface

$$\begin{aligned} f(\theta x + (1 - \theta)y) &= \text{máx}_i(\theta x_i + (1 - \theta)y_i) \\ &\leq \theta \text{máx}_i x_i + (1 - \theta)\text{máx}_i y_i \\ &= \theta f(x) + (1 - \theta)f(y). \end{aligned}$$

Funciones afines

Una función $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ es llamada afín si es resultado de una suma de funciones lineales y una constante, es decir, tiene la forma $f(x) = Ax + b$, donde $A \in \mathbb{R}^{m \times n}$ y $b \in \mathbb{R}^m$. Supongamos que $S \subset \mathbb{R}^n$ es convexo y $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ es una función afín. Entonces la imagen de S bajo f

$$f(S) = \{f(x) | x \in S\}$$

es convexa. Similarmente, si $f : \mathbb{R}^k \rightarrow \mathbb{R}^n$ es una función afín, la imagen inversa de S , bajo f

$$f^{-1}(S) = \{x | f(x) \in S\}$$

es convexa.

Epígrafe de una función

La gráfica de una función $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ se define como el conjunto

$$\{(x, f(x)) | x \in A\},$$

que es un subconjunto de \mathbb{R}^{n+1} . El epígrafe de una función se define como

$$\mathbf{epi} \mathbf{f} = \{(x, t) | x \in A, f(x) \leq t\},$$

que es también un conjunto de \mathbb{R}^{n+1} . “Epi” significa “sobre”, por tanto epígrafe significa, “sobre la gráfica”. Esta definición se ilustra en la figura 2.4. El concepto que liga a los conjuntos convexos con la funciones convexas es precisamente el epígrafe de una función.

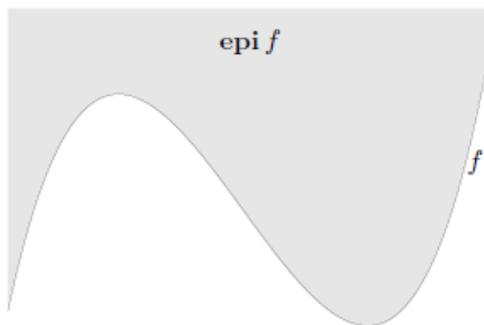


Figura 2.3: Epígrafe de la función f .

Proposición 1. *Una función es convexa si y sólo si su epígrafe es un conjunto convexo.*

Muchos resultados sobre funciones convexas son probados geométricamente usando el epígrafe y aplicando resultados para conjuntos convexos. Como un ejemplo, consideremos la condición de primer orden para convexidad:

$$f(y) \geq f(x) + \nabla f(x)^t (y - x),$$

donde f es convexa y $x, y \in \mathbf{dom} \mathbf{f}$. Es posible interpretar esta desigualdad en términos del epígrafe de f . Si $(y, t) \in \mathbf{epi} \mathbf{f}$, entonces

$$t \geq f(y) \geq f(x) + \nabla f(x)^t (y - x).$$

Condiciones de primer orden

Supongamos que f es diferenciable (∇f existe en cada punto del dominio de f que es un conjunto abierto). Entonces f es convexa si y sólo si el dominio de f es convexo y además

$$f(y) \geq f(x) + \nabla f(x)^t(y - x)$$

se satisface para todo x, y en el dominio de f . En la figura 2.5 se ilustra este hecho.

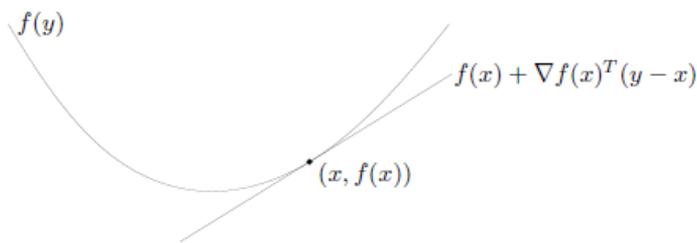


Figura 2.4

La desigualdad de la definición muestra que a partir de la información local de una función convexa es posible derivar información global. Esta es posiblemente la propiedad más importante de las funciones convexas y permite explicar algunas propiedades notables de funciones convexas y problemas de optimización convexa.

Condiciones de segundo orden

Supongamos que $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es dos veces diferenciable, es decir, su *Hessiano* o segunda derivada $\nabla^2 f$ existe en cada punto de A , que es un conjunto abierto. Entonces f es convexa si y sólo si A es convexo y su Hessiano es semidefinido positivo para cualquier $x \in A$, es decir

$$\nabla^2 f(x) \succeq 0 \quad \forall x \in A.$$

Ejemplo 2.2.2. Consideremos la función cuadrática $f : \mathbb{R}^n \rightarrow \mathbb{R}$, definida por

$$f(x) = (1/2)x^t P x + q^t x + r,$$

con $P \in \mathbb{R}^{n \times n}$, $q \in \mathbb{R}^n$ y $r \in \mathbb{R}$. Ya que $\nabla^2 f(x) = P$, $\forall x \in \mathbb{R}^n$, f es convexa si y sólo si P es una matriz definida positiva.

2.3. Problemas de optimización convexa

Utilizamos la notación

$$\begin{aligned} \min f_0(x), & \tag{2.1} \\ \text{s. a } f_i(x) \leq 0, & \quad i = 1, 2, \dots, m \\ h_i(x) = 0, & \quad i = 1, 2, \dots, p. \end{aligned}$$

para describir el problema de encontrar un x que minimice la función $f_0(x)$ entre todos aquellos que satisfacen las condiciones

$$f_i(x) \leq 0 \text{ y } h_i(x) = 0.$$

La función $f_0(x)$ es llamada *función objetivo* o *función de costo*. Las desigualdades $f_i(x) \leq 0$ son llamadas *restricciones de desigualdad* y las ecuaciones $h_i(x) = 0$ son llamadas *restricciones de igualdad*. El conjunto de puntos

$$\mathcal{D} = \bigcap_{i=0}^m \text{dom } \mathbf{f}_i \cap \bigcap_{i=1}^p \text{dom } \mathbf{h}_i,$$

es llamado dominio del problema de optimización.

Un punto x es factible si satisface las restricciones de igualdad y desigualdad. El problema 2.1 es factible si existe al menos un punto x factible. Al conjunto de todos los puntos factibles se le conoce como *conjunto factible*.

El valor óptimo del problema 2.1 se define como

$$p^* = \min\{f_0(x) \mid f_i(x) \leq 0, i = 1, 2, \dots, m, h_i(x) = 0, i = 1, 2, \dots, p.\}$$

Decimos que x^* es un punto óptimo si es factible y $f_0(x^*) = p^*$. El conjunto de todos los puntos óptimos se denota por,

$$X_{opt} = \{x \mid f_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, p, f_0(x) = p^*\}$$

Si existe un punto óptimo para el problema 2.1 decimos que el valor óptimo es alcanzado, por tanto es soluble.

Si x es factible y $f_i(x) = 0$ decimos que la i -ésima restricción de desigualdad $f_i(x) \leq 0$ es activa en x . Si $f_i(x) < 0$ decimos que es pasiva. Las restricciones de igualdad son activas en todos los puntos factibles. Decimos que una restricción es redundante si al omitirla el conjunto factible no cambia.

Una propiedad importante de los problemas de optimización convexa es el hecho de que cualquier óptimo local es también un óptimo global.

Problemas de optimización convexa en forma estándar

Un problema de optimización convexa tiene la siguiente forma

$$\begin{aligned} \min f_0(x), & \quad (2.2) \\ \text{s. a } f_i(x) \leq 0, & \quad i = 1, 2, \dots, m, \\ a_i^t x = b_i, & \quad i = 1, 2, \dots, p. \end{aligned}$$

donde f_i es convexa para $i = 0, 1, \dots, m$. Comparando con el problema 2.1, el problema convexo tiene tres condiciones adicionales:

1. La función objetivo debe ser convexa.

2. Las restricciones de desigualdad deben ser funciones convexas.
3. Las restricciones de igualdad deben ser funciones afines.

El conjunto factible de un problema de optimización convexa es un conjunto convexo ya que resulta de la intersección del dominio del problema

$$\mathcal{D} = \bigcap_{i=0}^m \text{dom } \mathbf{f}_i$$

que es un conjunto convexo. Por tanto en un problema de optimización convexa, se minimiza una función objetivo convexa sobre un conjunto convexo. Si la función objetivo es estrictamente convexa, entonces el conjunto óptimo contiene a lo más un punto.

Con un pequeño abuso de notación podemos referirnos al problema (2.2) como un problema de optimización concava si f_0 es concava y las restricciones de desigualdad son convexas. El problema de optimización concava es fácil de resolver minimizando la función objetivo $-f_0(x)$.

Problema epígrafe

El problema epígrafe es una variación del problema estandar (2.1) el cual se puede expresar de la siguiente forma

$$\begin{aligned} & \text{minimizar } t & (2.3) \\ & \text{s. a } f_0(x) - t \leq 0, \quad i = 1, 2, \dots, m, \\ & \quad \quad h_i(x) = 0, \quad i = 1, 2, \dots, p. \end{aligned}$$

con $x \in \mathbb{R}^n$ y $t \in \mathbb{R}$.

Es fácil ver que este problema es equivalente al problema original. (x, t) es un óptimo para (2.3) si y sólo si x es un óptimo del problema (2.1) y $t = f_0(x)$. La función objetivo del problema epígrafe es una función

linear en las variables x, t . El problema epígrafe puede ser interpretado geoméricamente como un problema en el *espacio gráfica* (x, t) en el que se minimiza t sobre el epígrafe de f_0 sujeto a las restricciones en x . Esto puede observarse en la figura 2.6.

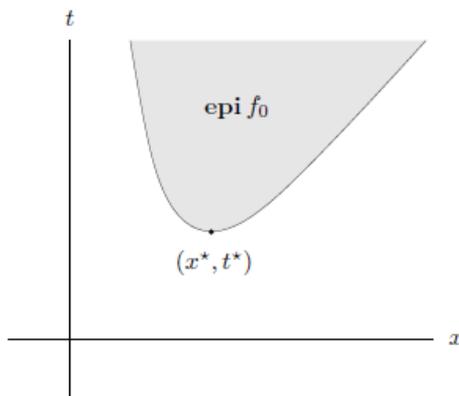


Figura 2.5: Interpretación geométrica del problema (2.3)

Problemas de optimización lineal (LP)

Cuando la función objetivo y las restricciones son funciones afines el problema de optimización se conoce como *problema de programación lineal*. este tipo de problemas tienen la siguiente forma

$$\begin{aligned} &\text{minimizar } c^t x + d && (2.4) \\ &\text{sujeto a } Gx \preceq h \\ &Ax = B. \end{aligned}$$

donde $G \in \mathbb{R}^{m \times n}$ y $A \in \mathbb{R}^{p \times n}$. Este es un ejemplo de problema de optimización convexa. La interpretación geométrica de un problema de programación lineal se muestra en la figura 2.7.



Figura 2.6: \mathcal{P} representa el conjunto factible. Las curvas de nivel de la función objetivo son planos ortogonales a c

Problemas de optimización cuadrática(QP)

El problema de optimización convexa 2.2 es llamado *problema de programación cuadrática* cuando la función objetivo (que es convexa) es una función cuadrática y las restricciones son funciones afines. Un problema de programación cuadrática se expresa como sigue

$$\begin{aligned} &\text{minimizar } \frac{1}{2}x^t P x + x^t x + r, \\ &\text{sujeto a } Gx \preceq h, \\ &\quad Ax = b, \end{aligned} \tag{2.5}$$

donde $P \in \mathbb{R}_+^{n \times n}$, $G \in \mathbb{R}^{m \times n}$ y $A \in \mathbb{R}^{p \times n}$. En un problema de programación cuadrática se minimiza una función cuadrática sobre un poliedro como se muestra en la figura 2.7.

Programación cónica de segundo orden (SOCP)

Un problema relacionado al problema QP es el problema de programación cónica de segundo orden conocido como SOCP por sus siglas en inglés,

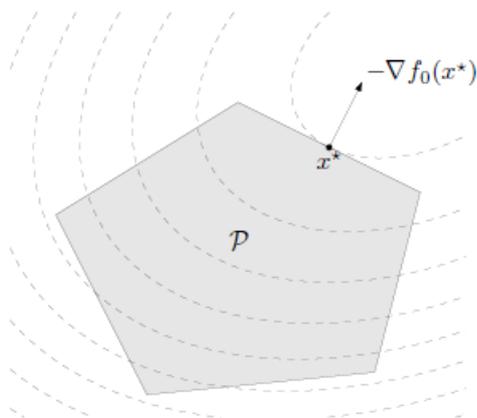


Figura 2.7

el cual se expresa de la siguiente forma

$$\begin{aligned} & \text{minimizar } f^t x, \\ & \text{sujeto a } \|A_i x + b_i\|_2 \leq c_i^t x + d_i, \quad i = 1, 2, \dots, m, \\ & \quad \quad \quad Fx = g. \end{aligned} \tag{2.6}$$

donde $x \in \mathbb{R}^n$ es la variable de optimización, $A_i \in \mathbb{R}^{n_i \times n}$ y $F \in \mathbb{R}^{p \times n}$.

Una restricción de la forma

$$\|Ax + b\|_2 \leq c^t x + d$$

donde $A \in \mathbb{R}^{k \times n}$, es conocida como *restricción cónica de segundo orden* ya que es similar a pedir que la función afín $(Ax + b, c^t x + d)$ se encuentre en el cono de segundo orden \mathbb{R}^{k+1} .

2.3.1. Transformación de un QCP en un SOCP

Un QCP (problema de optimización con restricciones cuadráticas) se define como sigue

$$\begin{aligned} & \text{minimizar } c^t x & (2.7) \\ & \text{sujeto a } x^t Q_i x + 2q_i^t x + \gamma_i \leq 0, \quad i = 1, \dots, p, \end{aligned}$$

donde la variable de optimización es $x \in \mathbb{R}^n$ y los parámetros $c \in \mathbb{R}^n$, $\gamma_i \in \mathbb{R}$, $q_i \in \mathbb{R}^n$ y $Q_i \in \mathbb{R}^{n \times n}$ definida positiva $\forall i = 1, \dots, p$. Sin pérdida de generalidad vamos a asumir que la función objetivo es lineal. Como Q_i es positiva definida $\forall i = 1, \dots, p$. entonces podemos escribir $Q_i = V_i^t V_i$ para algún $V_i \in \mathbb{R}^{m \times n}$.

Observemos que

$$\begin{aligned} x^t Q_i x + 2q_i^t x + \gamma_i \leq 0 & \Rightarrow 4x^t Q_i x \leq -4(\gamma_i + 2q_i^t x) \\ & \Rightarrow 4x^t V_i^t V_i x \leq (1 - (\gamma_i + 2q_i^t x))^2 - (1 + (\gamma_i + 2q_i^t x))^2 \\ & \Rightarrow \begin{bmatrix} 2V_i x \\ 1 + \gamma_i + 2q_i^t x \end{bmatrix}^t \begin{bmatrix} 2V_i x \\ 1 + \gamma_i + 2q_i^t x \end{bmatrix} \leq (1 - \gamma_i - 2q_i^t x)^2 \\ & \Rightarrow \left\| \begin{bmatrix} 2V_i x \\ 1 + \gamma_i + 2q_i^t x \end{bmatrix} \right\| \leq 1 - \gamma_i - 2q_i^t x. \end{aligned}$$

Por tanto el problema (2.7) es equivalente al siguiente SOCP

$$\begin{aligned} & \text{minimizar } c^t x & (2.8) \\ & \text{sujeto a } \left\| \begin{bmatrix} 2V_i x \\ 1 + \gamma_i + 2q_i^t x \end{bmatrix} \right\| \leq 1 - \gamma_i - 2q_i^t x, \quad i = 1, \dots, p. \end{aligned}$$

2.4. Dualidad

Consideremos el problema de optimización estándar

$$\begin{aligned}
 &\text{minimizar } f_0(x) && (2.9) \\
 &\text{sujeto a } f_i(x) \leq 0, \quad i = 1, 2, \dots, m, \\
 &\quad \quad \quad h_i(x) = 0, \quad i = 1, \dots, p.
 \end{aligned}$$

Supongamos que el dominio $\mathcal{D} \neq \emptyset$ y denotemos al valor óptimo de (2.7) por p^* . Definimos el Lagrangiano $L : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$ asociado al problema (2.7) como

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$$

con $\text{dom } L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$. Denotemos por λ_i al multiplicador de Lagrange asociado a la i -ésima restricción de desigualdad $f_i(x) \leq 0$, similarmente vamos a denotar por ν_i al multiplicador asociado a la i -ésima restricción de igualdad $h_i(x) = 0$. Los vectores λ, ν son llamados vectores duales o multiplicadores vectoriales de Lagrange asociados al problema (2.7).

Función dual de Lagrange

Definimos la función dual de Lagrange, $g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ como el valor mínimo del Lagrangiano sobre x . Para $\lambda \in \mathbb{R}^m$ y $\nu \in \mathbb{R}^p$

$$\begin{aligned}
 g(\lambda, \nu) &= \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \\
 &= \inf_{x \in \mathcal{D}} (f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)).
 \end{aligned}$$

Cuando el Lagrangiano es no acotado inferiormente la función dual toma el valor $-\infty$. La función dual produce cotas inferiores para el valor óptimo p^* de (2.7). Para cualquier $\lambda \succeq 0$ y ν se cumple que

$$g(\lambda, \nu) \leq p^*. \quad (2.10)$$

Ejemplo 2.4.1. Consideremos el problema de optimización

$$\begin{aligned} &\text{minimizar } x^t x, \\ &\text{sujeto a } Ax = b. \end{aligned}$$

donde $A \in \mathbb{R}^{p \times n}$. Observemos que este problema no tiene restricciones de desigualdad y tiene p restricciones de igualdad. El Lagrangiano del problema es

$$L(x, \nu) = x^t x + \nu^t (Ax - b),$$

con dominio $\mathbb{R}^{n \times p}$. La función dual está dada por $g(\nu) = \inf_x L(x, \nu)$, como $L(x, \nu)$ es una función cuadrática convexa de x podemos encontrar el mínimo de la condición de optimalidad haciendo

$$\nabla_x L(x, \nu) = 2x + A^t \nu = 0,$$

entonces $x = -\frac{1}{2}A^t \nu$. La función dual esta dada por

$$\begin{aligned} g(\nu) &= L\left(-\frac{1}{2}A^t \nu, \nu\right) \\ &= -\frac{1}{4}\nu^t A A^t \nu - b^t \nu. \end{aligned}$$

que es una función concava cuadrática con dominio \mathbb{R}^p . La propiedad (2.8) establece que para cualquier $\nu \in \mathbb{R}^p$ se tiene

$$-\frac{1}{4}\nu^t A A^t \nu - b^t \nu \leq \inf\{x x^t \mid Ax = b\},$$

Ejemplo 2.4.2 (Programación lineal estándar). Consideremos un LP en forma estándar

$$\begin{aligned} &\text{minimizar } c^t x && (2.11) \\ &\text{sujeto a } Ax = b, \\ & && x \succeq 0. \end{aligned}$$

que tiene restricciones de desigualdad $f_i(x) = -x_i$, $i = 1, 2, \dots, n$.

El Lagrangiano de (2.9) es

$$\begin{aligned} L(x, \lambda, \nu) &= c^t x + \nu^t (Ax - b) - \sum_{i=1}^n \lambda_i x_i \\ &= -b^t \nu + (c + A + \lambda)^t x, \end{aligned}$$

por tanto la función dual de este problema es

$$\begin{aligned} g(\lambda, \nu) &= \inf_x L(x, \lambda, \nu) \\ &= -b^t \nu + \inf_x (c + A + \lambda)^t x. \end{aligned}$$

Una función lineal es acotada inferiormente cuando esta es idénticamente cero, entonces $g(\lambda, \nu) = -\infty$ excepto cuando

$$g(\lambda, \nu) = \begin{cases} -b^t \nu & \text{si } A^t \nu - \lambda + C = 0 \\ -\infty & \text{otro caso} \end{cases}$$

notemos que g es finita solo sobre un conjunto afín propio de $\mathbb{R}^m \times \mathbb{R}^p$. La propiedad de cota inferior es no trivial sólo cuando $\lambda \succeq 0$ y $A^t \nu - \lambda + c = 0$, cuando esto ocurre $-b^t \nu$ es una cota inferior del valor óptimo del problema (2.9).

Problema dual de Lagrange

Para cada par (λ, ν) con $\lambda \succeq 0$, la función dual de Lagrange nos proporciona una cota inferior para el óptimo p^* del problema (2.7). De esta manera se tiene una cota inferior que depende de los parámetros λ y ν . Una pregunta natural es ¿Cuál es la mejor cota inferior que podemos obtener de la función dual de Lagrange? Esto nos lleva al siguiente problema de optimización

$$\text{maximizar } g(\lambda, \nu) \tag{2.12}$$

sujeto a $\lambda \succeq 0$.

Este problema se conoce como *Problema dual de Lagrange* asociado al problema (2.7). En este contexto el problema original (2.7) es llamado *problema primal*.

El término *dual factible* para describir a la pareja (λ, ν) con $\lambda \succeq 0$ y $g(\lambda, \nu) > \infty$ ahora toma sentido. Esto implica como el nombre lo sugiere que (λ, ν) es factible para el problema dual. A partir de ahora nos referiremos a (λ^*, ν^*) como el dual óptimo o multiplicadores de Lagrange óptimos (si estos existen) para el problema (2.10).

El problema dual de Lagrange es un problema de optimización convexa ya que la función objetivo a maximizar es concava y las restricciones son funciones convexas. Esto sucede a pesar de que el problema primal no sea convexo.

Ejemplo 2.4.3 (Lagrangiano dual de un LP). En el ejemplo (2.4.2) se mostró que la función dual de Lagrange para el problema de programación lineal estándar

$$\begin{aligned} &\text{minimizar } c^t x, & (2.13) \\ &\text{sujeto a } Ax = b \\ & & x \succeq 0. \end{aligned}$$

está dado por

$$g(\lambda, \nu) = \begin{cases} -b^t \nu & \text{si } A^t \nu - \lambda + C = 0 \\ -\infty & \text{otro caso.} \end{cases}$$

Estrictamente hablando, el problema dual de Lagrange de un LP estándar es maximizar esta función dual g sujeta a $\lambda \succeq 0$, i.e.,

$$\begin{aligned} &\max g(\lambda, \nu), & (2.14) \\ &\text{sujeto a } \lambda \succeq 0. \end{aligned}$$

aquí g es finita sólo cuando $A^t\nu - \lambda + c = 0$. Es posible formular un problema equivalente haciendo estas restricciones de igualdad explícitas.

$$\begin{aligned} & \text{maximizar} && -b^t\nu && (2.15) \\ & \text{sujeto a} && A^t\nu - \lambda + c = 0 \\ & && \lambda \succeq 0. \end{aligned}$$

Este problema a su vez puede ser expresado como

$$\begin{aligned} & \text{maximizar} && -b^t\nu && (2.16) \\ & \text{sujeto a} && A^t\nu - \lambda + c \succeq 0. \end{aligned}$$

que es un LP con restricciones de desigualdad. Con un abuso de terminología nos referimos al problema (2.13) o (2.14) como el *Lagrangiano dual de un LP estándar*.

Ejemplo 2.4.4 (Lagrangiano dual de un LP con restricciones de desigualdad.). De forma similar podemos hallar el problema dual de Lagrange de un LP con restricciones de desigualdad.

$$\begin{aligned} & \text{minimizar} && c^t x && (2.17) \\ & \text{sujeto a} && Ax \succeq b. \end{aligned}$$

El Lagrangiano de este problema es

$$\begin{aligned} L(\lambda, \nu) &= c^t x + \lambda^t (Ax - b) \\ &= -b^t \lambda + (A^t x + c)^t x, \end{aligned}$$

por lo que

$$g(\lambda) = \inf_x L(x, \lambda)$$

$$= -b^t \lambda + \inf_x (A^t x + c)^t x.$$

Observemos que el ínfimo de una función lineal es $-\infty$, excepto cuando es idénticamente cero, entonces la función dual es

$$g(\lambda, \nu) = \begin{cases} -b^t \nu & \text{si } A^t \nu - \lambda + C = 0 \\ -\infty & \text{otro caso.} \end{cases}$$

La variable dual $\lambda \succeq 0$ es factible si $\lambda \succeq 0$ y $A^t \lambda + c = 0$.

El dual de Lagrange del problema (2.15) es maximizar g sobre todos los $\lambda \succeq 0$. De nuevo podemos reformular esto explícitamente incluyendo las condiciones de factibilidad duales como restricciones

$$\begin{aligned} &\text{maximizar } -b^t \lambda, & (2.18) \\ &\text{sujeto a } A^t \lambda + c = 0 \\ & & \lambda \succeq 0, \end{aligned}$$

que es un LP estándar.

Notemos la interesante simetría entre los LP estándar y estándar con restricciones y sus respectivos duales. El dual de un LP estándar es un LP con restricciones de desigualdad y viceversa. Es fácil verificar que el dual de Lagrange de (2.16) es equivalente al primal (2.15).

Dualidad débil

El valor óptimo del problema dual de Lagrange es por definición la mejor cota inferior que podemos encontrar para la solución p^* del problema primal de la función dual de Lagrange, i.e.,

$$d^* \leq p^* \tag{2.19}$$

que se mantiene incluso cuando la función objetivo del problema primal es no convexa, esta propiedad es llamada *Dualidad débil*.

La desigualdad de dualidad débil (2.17) se preserva aun cuando d^* y p^* son infinitos. Por ejemplo si el problema primal es no acotado inferiormente, es decir, $p^* = -\infty$, se tendrá que $d^* = -\infty$, i.e., el problema dual no es factible. Inversamente si el problema dual es no acotado superiormente, es decir, $d^* = \infty$ entonces, $p^* = \infty$, es decir el problema primal no es factible.

La diferencia $p^* - d^*$ es conocida como brecha de dualidad óptima (optimal duality gap en inglés), ya que esta es la diferencia entre el óptimo primal y la mejor cota inferior que se puede obtener de la función dual de Lagrange. La brecha de dualidad óptima siempre es no negativa.

La desigualdad (2.17) en ocasiones es utilizada para encontrar una cota inferior del valor óptimo cuando el problema primal es difícil de resolver, ya que el problema dual es siempre convexo y en algunos casos puede ser resuelto eficientemente para encontrar d^* .

Dualidad fuerte y calificación de restricción

Si la igualdad $p^* = d^*$ se satisface, es decir, la brecha de dualidad es cero, entonces decimos que la dualidad fuerte se cumple. Esto significa que la mejor cota que podemos obtener de la función dual de Lagrange es estrecha. La dualidad fuerte en general no siempre se cumple, pero si el problema original es convexo, i.e., tiene la forma

$$\begin{aligned} &\text{minimizar } f_0(x) \\ &\text{sujeto a } f_i(x) \leq 0, i = 1, 2, \dots, m, \\ &\quad Ax = b, \end{aligned}$$

con f_0, \dots, f_m convexas, usualmente (pero no siempre) se tendrá dualidad fuerte.

Hay algunos resultados que establecen condiciones en el problema, más allá de la convexidad, bajo las cuales la dualidad fuerte se cumple. Estas condiciones son llamadas *calificaciones de restricción*.

Una calificación de restricción simple es, por ejemplo, la desigualdad de Slater: Existe un $x \in \mathbf{Relint D}$ tal que

$$f_i(x) < 0 \text{ y } Ax = b.$$

Tal punto es en ocasiones llamado *estrictamente factible* ya que las restricciones de desigualdad se satisfacen con desigualdad estricta. El teorema de Slater establece que la dualidad fuerte se mantiene si la condición de Slater se cumple y el problema es convexo.

La condición de Slater puede ser refinada cuando algunas de las restricciones de desigualdad f_i son afines, es decir, si las primeras k restricciones de desigualdad f_1, f_2, \dots, f_k son afines, entonces, la dualidad fuerte se cumple siempre que se cumplan las siguientes condiciones que más débiles:

Existe un $x \in \mathbf{Relint D}$ con

$$\begin{aligned} f_i(x) &\leq 0, \quad i = 0, 1, \dots, k, \quad f_i(x) < 0, \quad i = 1, 2, \dots, m. \\ Ax &= b. \end{aligned}$$

En otras palabras las desigualdades afines no necesitan cumplirse con desigualdad estricta.

Ejemplo 2.4.5. Recordemos el problema (2.4.1)

$$\begin{aligned} &\text{minimizar } x^t x, \\ &\text{sujeto a } Ax = b. \end{aligned}$$

Su dual asociado es el problema de

$$\text{maximizar } -\frac{1}{4}\nu^t AA^t \nu - b^t \nu,$$

que es un problema concavo cuadrático sin restricciones.

La condición de Slater simplemente establece que el problema primal es factible, i.e., $p^* = d^*$ siempre que $b \in \mathcal{R}(A)$, es decir $p^* < \infty$. De hecho para este problema siempre tendremos dualidad fuerte, incluso cuando $p^* = \infty$. Este es el caso cuando $b \notin \mathcal{R}(A)$, entonces existe z , con $A^t z = 0$, $b^t z \neq 0$. Se sigue que la función dual es no acotada superiormente a lo largo de la recta $\{tz | t \in \mathbb{R}\}$ entonces $d^* = \infty$ también.

Condiciones de óptimalidad KKT

Supongamos que $f_0, f_1, \dots, f_m, h_1, h_2, \dots, h_p$ son diferenciables (y por lo tanto tienen dominios abiertos). Sea x^* y (λ^*, ν^*) los óptimos primales y duales respectivamente con brecha igual a cero. Como x^* minimiza $L(x, \lambda^*, \nu^*)$ respecto a x se sigue que su gradiente debe anularse en x^* , i.e.,

$$\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i \nabla f_i(x^*) + \sum_{i=1}^p \nu_i \nabla h_i(x^*) = 0 \quad (2.20)$$

Entonces las siguientes

$$\begin{aligned} f_i(x^*) &\leq 0 \quad i = 1, 2, \dots, m, \\ h_i(x^*) &= 0 \quad i = 1, 2, \dots, p, \\ \lambda_i^* &\geq 0 \quad i = 1, 2, \dots, m, \\ \lambda_i^* f_i(x^*) &= 0, \end{aligned}$$

son llamadas condiciones de Karush-Kuhn-Tucker (KKT).

Para un problema convexo se tiene que las condiciones KKT son suficientes para los puntos que sean óptimos primal y dual. En otras palabras,

si las f_i son convexas, las h_i son afines, y además \bar{x} , $\bar{\lambda}$, $\bar{\nu}$ son puntos que satisfacen las condiciones de KKT, entonces \bar{x} y $(\bar{\lambda}, \bar{\nu})$ son óptimos primal y dual respectivamente con brecha de dualidad igual a cero.

En resumen para cualquier problema de optimización convexa con función objetivo y restricciones diferenciables, cualquier punto que satisfaga las condiciones de KKT son primal y dual óptimos y tienen brecha de dualidad igual a cero.

Si un problema de óptimalidad con restricciones y función objetivo diferenciables satisface la condición de Slater, entonces las condiciones de KKT proporcionan condiciones necesarias y suficientes para la óptimalidad: La condición de Slater implica que si la brecha es cero y el óptimo dual es alcanzado, entonces, x es óptimo si y sólo si hay (λ, ν) que junto con x satisfacen las condiciones de KKT. Las condiciones de KKT juegan un papel importante en optimización. En algunos (muy pocos) casos especiales es posible resolver las condiciones de KKT analíticamente.

Capítulo 3

Selección de portafolios robustos

Para comenzar este capítulo vamos a suponer que se tienen n activos, el vector de rendimientos de los activos sobre un periodo simple es denotado por $\vec{r} \in \mathbb{R}^n$ donde el activo i retorna $(1 + r_i)$ pesos por cada peso invertido. Asumiremos que los rendimientos de los activos en periodos de tiempo distintos son variables aleatorias que se distribuyen normales y que son independientes. Vamos a suponer también que no están permitidas las ventas en corto, es decir, $\phi_i \geq 0 \forall i = 1, 2, \dots, n$.

Para poder iniciar con el planteamiento de los problemas de selección de portafolios robustos es necesario definir los siguientes conjuntos.

$$S_{\Sigma} = \{\Sigma : \Sigma = \Sigma_0 + \Delta \succeq \mathbf{0}, \Delta = \Delta^t, \|\mathbf{N}^{-1/2} \Delta \mathbf{N}^{-1/2}\| \leq \eta\} \quad (3.1)$$

con $\Sigma_0 \succeq \mathbf{0}$ y $\mathbf{N} \succ \mathbf{0}$, es decir, Σ_0 es semidefinida positiva y \mathbf{N} es definida positiva, notación que utilizaremos de aquí en adelante.

$$S_m = \{\vec{\mu} : \vec{\mu} = \vec{\mu}_0 + \vec{\zeta} \in \mathbb{R}^n, |\zeta_i| \leq \gamma_i, i = 1, \dots, n\} \quad (3.2)$$

Llamamos a S_Σ y S_m conjuntos de incertidumbre pues adquieren la estructura de regiones de confianza en torno a los estimadores de los parámetros en cuestión.

Como se explicó en el primer capítulo, la posición de un inversionista en el mercado es descrita por un portafolio $\vec{\phi} \in \mathbb{R}^n$ donde la i -ésima componente ϕ_i representa el porcentaje de su riqueza inicial invertida en el activo i , así el rendimiento \vec{r}_ϕ del portafolio $\vec{\phi}$ está dado por

$$\vec{r}_\phi \sim \mathcal{N}(\mu^t \vec{\phi}, \vec{\phi}^t \Sigma \vec{\phi})$$

donde $\vec{\mu}$ denota al vector de medias y Σ denota a la matriz de varianza-covarianza muestral de los rendimientos de los activos.

El objetivo de un inversionista es maximizar el rendimiento de la inversión sujeta a ciertas restricciones sobre el riesgo. Un modelo matemático para la selección de este portafolio fue propuesto por Harry Markowitz en 1952 y asumía que el valor esperado de los rendimientos $E[\vec{r}_\phi]$, así como la covarianza $Var[\vec{r}_\phi]$ son valores conocidos. En este modelo el rendimiento de la inversión es el valor esperado del rendimiento del portafolio $E[\vec{r}_\phi]$ y el riesgo asociado es la varianza $Var[\vec{r}_\phi]$. El objetivo del inversionista es elegir un portafolio $\vec{\phi}^*$ que tenga mínima varianza entre todos aquellos que tienen un valor esperado mayor o igual que cierto valor α , es decir $\vec{\phi}^*$ es la solución del problema de optimización cuadrático convexo

$$\begin{aligned} &\text{minimizar } Var[\vec{r}_\phi], \\ &\text{sujeto a } E[\vec{r}_\phi] \geq \alpha, \\ &\quad \mathbf{1}^t \vec{\phi} = 1. \end{aligned} \tag{3.3}$$

un inconveniente del modelo de Markowitz es el hecho de que el portafolio óptimo $\vec{\phi}^*$ es demasiado sensible a los parámetros del mercado ($E[\vec{r}_\phi], Var[\vec{r}_\phi]$) puesto que usualmente son obtenidos de datos ruidosos, frecuentemente $\vec{\phi}^*$ amplifica el ruido. Al introducir “medidas de incertidumbre” en los modelos de mercado, intentaremos corregir dicha sensibilidad a

perturbaciones. Con el objetivo de obtener portafolios que se desempeñen mejor, es necesario representar la incertidumbre de la información de los parámetros del mercado mediante los conjuntos S_m y S_Σ . Estos portafolios son soluciones de problemas de optimización min-max apropiadamente definidos, llamados *problemas de selección de portafolios robustos*.

El problema robusto análogo al modelo de Markowitz es el siguiente:

$$\begin{aligned} \min \quad & \max_{\{\Sigma \in S_\Sigma\}} \text{Var}[\vec{r}_\phi], \\ \text{s. a} \quad & \min_{\{\vec{\mu} \in S_m\}} E[\vec{r}_\phi] \geq \alpha, \\ & \vec{1}^t \vec{\phi} = 1. \end{aligned} \tag{3.4}$$

El objetivo del problema robusto de selección del portafolio de mínima varianza (3.4) es minimizar el peor caso de la varianza del portafolio sujeto a la restricción de que el peor caso del rendimiento esperado del portafolio es al menos α . Lo que se espera de esta nueva formulación es que la sensibilidad de la solución óptima del problema sea significativamente más pequeña que la del problema clásico (3.3).

Un problema estrechamente relacionado es el *problema robusto de rendimiento máximo*. En este problema el objetivo es maximizar el peor caso del valor esperado del rendimiento sujeto a una restricción sobre el peor caso de la varianza, es decir, se requiere resolver el siguiente problema de optimización

$$\begin{aligned}
& \max \min_{\{\vec{\mu} \in S_m\}} E[\vec{r}_\phi], & (3.5) \\
& \text{s. a } \max_{\{\Sigma \in S_\Sigma\}} \text{Var}[\vec{r}_\phi] \leq \lambda, \\
& \vec{1}^t \vec{\phi} = 1.
\end{aligned}$$

3.1. Selección del portafolio robusto de mínima varianza

En esta sección se realizará un análisis detallado del problema robusto de mínima varianza (3.4). Se demostrará que para los conjuntos de incertidumbre definidos en (3.1)-(3.2) el problema se reduce a un problema de programación cónica de segundo orden (SOCP). En consecuencia, el problema dual podrá ser resuelto de la misma forma.

El problema robusto de mínima varianza se define de la siguiente manera

$$\begin{aligned}
& \min \max_{\{\Sigma \in S_\Sigma\}} \{\vec{\phi}^t \Sigma \vec{\phi}\} & (3.6) \\
& \text{s.a } \min_{\{\vec{\mu} \in S_m\}} \vec{\mu}^t \vec{\phi} \geq \alpha, \\
& \vec{1}^t \vec{\phi} = 1.
\end{aligned}$$

Haciendo un cambio de variable podemos plantear (3.6) como

$$\begin{aligned}
& \text{minimizar } \nu & (3.7) \\
& \text{s.a } \max_{\{\Sigma \in S_\Sigma\}} \{\vec{\phi}^t \Sigma \vec{\phi}\} \leq \nu, \\
& \min_{\{\vec{\mu} \in S_m\}} \vec{\mu}^t \vec{\phi} \geq \alpha, \\
& \vec{1}^t \vec{\phi} = 1.
\end{aligned}$$

3.1. SELECCIÓN DEL PORTAFOLIO ROBUSTO DE MÍNIMA VARIANZA45

Un caso particular de este problema es aquel en el que los conjuntos de incertidumbre son finitos, es decir, $S_\Sigma = \{\Sigma_1, \Sigma_2, \dots, \Sigma_p\}$ y $S_m = \{\mu_1, \mu_2, \dots, \mu_s\}$ entonces el problema se transforma en el siguiente

$$\begin{aligned}
 & \min \nu & (3.8) \\
 \text{s.a. } & \vec{\phi}^t \Sigma_k \vec{\phi} \leq \nu, \quad \forall k = 1, 2, \dots, p, \\
 & \vec{\mu}_k^t \vec{\phi} \geq \alpha, \quad \forall k = 1, 2, \dots, s, \\
 & \vec{1}^t \vec{\phi} = 1.
 \end{aligned}$$

Si los conjuntos S_Σ y S_m no son finitos, definimos $\tilde{\Delta} = \mathbf{N}^{-1/2} \Delta \mathbf{N}^{-1/2}$, de esta manera

$$\begin{aligned}
 & \text{máx}\{\vec{\phi}^t \Sigma \vec{\phi} : \Sigma \in S_\Sigma\} \\
 & = \text{máx}\{\vec{\phi}^t (\Sigma_0 + \Delta) \vec{\phi} : \Delta = \Delta^t, \|\mathbf{N}^{-1/2} \Delta \mathbf{N}^{-1/2}\| \leq \eta, \Sigma_0 + \Delta \succeq \mathbf{0}\} \\
 & = \text{máx}\{\vec{\phi}^t \Sigma_0 \vec{\phi} + (\mathbf{N}^{1/2} \vec{\phi})^t \tilde{\Delta} (\mathbf{N}^{1/2} \vec{\phi}) : \|\tilde{\Delta}\| \leq \eta, \Sigma_0 + \mathbf{N}^{1/2} \tilde{\Delta} \mathbf{N}^{1/2} \succeq \mathbf{0}\} \\
 & \leq \vec{\phi}^t \Sigma_0 \vec{\phi} + \eta (\mathbf{N}^{1/2} \vec{\phi})^t (\mathbf{N}^{1/2} \vec{\phi}). & (3.9)
 \end{aligned}$$

donde (3.9) se sigue de las propiedades de la norma de una matriz.

Como $\|\tilde{\Delta}\| = \max_{1 \leq i \leq m} \{|\lambda_i(\tilde{\Delta})|\}$ y $\mathbf{N} \succ \mathbf{0}$, la cota (3.9) es alcanzada cuando

$$\tilde{\Delta}^* = \frac{\eta (\mathbf{N} \vec{\phi}) (\mathbf{N} \vec{\phi})^t}{\|\mathbf{N}^{1/2} \vec{\phi}\|^2}$$

a no ser que $\vec{\phi} = \mathbf{0}$. Por lo tanto (3.9) está dado por

$$\vec{\phi}^t (\Sigma_0 + \eta \mathbf{N}) \vec{\phi} \tag{3.10}$$

y es alcanzado cuando

$$\tilde{\Delta}^* = \frac{\eta \mathbf{N} \vec{\phi} \vec{\phi}^t \mathbf{N}}{\vec{\phi}^t \mathbf{N} \vec{\phi}^t},$$

a no ser que $\vec{\phi} = \mathbf{0}$. Ya que $\Sigma_0 + \mathbf{N}^{1/2} \tilde{\Delta} \mathbf{N}^{1/2} \succeq \mathbf{0}$, se sigue que la desigualdad en (3.9) es de hecho una igualdad, por tanto

$$\max\{\vec{\phi}^t \Sigma \vec{\phi} : \Sigma \in S_\Sigma\} = \vec{\phi}^t (\Sigma_0 + \eta \mathbf{N}) \vec{\phi}. \quad (3.11)$$

Cuando los conjuntos de incertidumbre S_m y S_Σ son definidos como en (3.1) y (3.2) el peor caso de los rendimientos de un portafolio fijo está dado por

$$\min_{\{\bar{\mu} \in S_m\}} \bar{\mu}^t \vec{\phi} = \mu_0^t \vec{\phi} - \bar{\gamma}^t |\vec{\phi}|; \quad (3.12)$$

por tanto de (3.9) y (3.10) el problema (3.7) puede reformularse como

$$\begin{aligned} & \min \nu \\ \text{s.a. } & \vec{\phi}^t (\Sigma_0 + \eta \mathbf{N}) \vec{\phi} \leq \nu \\ & \mu_0^t \vec{\phi} - \bar{\gamma}^t |\vec{\phi}| \geq \alpha, \\ & \bar{\mathbf{1}}^t \vec{\phi} = 1. \end{aligned} \quad (3.13)$$

claramente este no es un SOCP, sin embargo, al no permitir las ventas en corto, el problema se plantea de la siguiente manera

$$\begin{aligned} & \min \nu \\ \text{s.a. } & \left\| \begin{bmatrix} 2\mathbf{S}\vec{\phi} \\ 1 - \nu \end{bmatrix} \right\| \leq 1 + \nu \\ & \mu_0^t \vec{\phi} - \bar{\gamma}^t \vec{\phi} \geq \alpha, \\ & \bar{\mathbf{1}}^t \vec{\phi} = 1, \\ & \phi_i \geq 0 \quad \forall i = 1, \dots, n. \end{aligned} \quad (3.14)$$

donde $\Sigma_0 + \eta\mathbf{N} = \mathbf{S}^t\mathbf{S}$.

Claramente este problema es un SOCP y puede ser resuelto de manera eficiente usando la paquetería computacional necesaria, además el costo computacional es similar al de resolver un problema de programación cuadrática.

3.2. Selección del portafolio robusto de máximo rendimiento

El problema robusto de máximo rendimiento está dado por

$$\begin{aligned} & \text{maximizar} \quad \min_{\{\vec{\mu} \in S_m\}} \vec{\mu}^t \vec{\phi} & (3.15) \\ & \text{s.a.} \quad \max_{\{\Sigma \in S_\Sigma\}} \{\vec{\phi}^t \Sigma \vec{\phi}\} \leq \lambda, \\ & & \vec{\mathbf{1}}^t \vec{\phi} = 1, \\ & & \phi_i \geq 0. \end{aligned}$$

De (3.2) y (3.10) el problema anterior es equivalente al siguiente problema

$$\begin{aligned} & \text{maximizar} \quad (\mu - \gamma)^t \vec{\phi} & (3.16) \\ & \text{s.a.} \quad \vec{\phi}^t (\Sigma_0 + \eta\mathbf{N}) \vec{\phi} \leq \lambda, \\ & & \vec{\mathbf{1}}^t \vec{\phi} = 1, \\ & & \vec{\phi} \geq 0. \end{aligned}$$

Al igual que el problema robusto de mínima varianza, el problema (3.16) puede ser planteado como un SOCP de la siguiente manera

$$\begin{aligned}
& \text{maximizar } (\mu - \gamma)^t \vec{\phi} & (3.17) \\
& \text{s.a. } \left\| \begin{bmatrix} 2\mathbf{S}\vec{\phi} \\ 1 - \lambda \end{bmatrix} \right\| \leq 1 + \lambda, \\
& \quad \vec{1}^t \vec{\phi} = 1, \\
& \quad \vec{\phi} \geq 0.
\end{aligned}$$

con $\Sigma_0 + \eta \mathbf{N} = \mathbf{S}^t \mathbf{S}$.

3.3. Conjuntos de incertidumbre

Como se explicó en secciones anteriores, dar un tratamiento especial a los errores estadísticos que provienen de los datos que se utilizan para calcular el portafolio óptimo es de gran importancia para obtener mejores resultados por lo que en esta sección se construirán teóricamente los conjuntos de incertidumbre S_Σ y S_μ . Estos resultados serán utilizados en secciones posteriores para realizar los experimentos numéricos del portafolio robusto.

3.3.1. Conjunto de incertidumbre para Σ

Cómo se explicó en capítulos anteriores el conjunto de incertidumbre se define como

$$S_\Sigma = \{ \Sigma : \Sigma = \Sigma_0 + \Delta \succeq \mathbf{0}, \Delta = \Delta^t, \|\mathbf{N}^{-1/2} \Delta \mathbf{N}^{-1/2}\| \leq \eta \} \quad (3.18)$$

con $\Sigma_0 \succeq \mathbf{0}$ y $\mathbf{N} \succ \mathbf{0}$ y además la norma $\|\mathbf{A}\|$ está dada por el máximo de los eigenvalores de \mathbf{A} .

Nuestro nuevo propósito es "hallar" un método para parametrizar este conjunto de incertidumbre. Sea Σ la matriz real de varianza-covarianza (desconocida) de los rendimientos y sea Σ_{ml} el estimador de máxima verosimilitud de Σ obtenido de los rendimientos. Se sabe que si $p < m$, es

decir, el número de observaciones es mayor a la dimensión de la matriz, entonces $\Sigma_{\text{ml}} \succ 0$ con probabilidad 1 y entonces

$$\Sigma_{\text{ml}} \sim W_{p-1}((p-1)\Sigma)$$

donde $W_q(\mathbf{A})$ denota una distribución Wishart con q grados de libertad centrada en la matriz \mathbf{A} ; es decir la función de densidad de $f(\Sigma_{\text{ml}}|\Sigma)$ está dada por

$$f(\Sigma_{\text{ml}}|\Sigma) = \begin{cases} c|(\Sigma)^{-1}|^{(p-1)/2}|\Sigma_{\text{ml}}|^{(p-m-2)/2}e^{-[(p-1)/2]\text{Tr}(\Sigma^{-1}\Sigma_{\text{ml}})}, & \Sigma_{\text{ml}} \succ 0, \\ 0, & \text{c.o.c.} \end{cases}$$

Un argumento muy similar establece que Σ_{ml} pertenece al conjunto

$$\mathcal{J} = \{\tilde{\Sigma} : \tilde{\Sigma} = \Sigma + \Delta \succeq 0, \Delta = \Delta^t, \|\Sigma^{-1/2}\Delta\Sigma^{-1/2}\| \leq \beta\} \quad (3.19)$$

con probabilidad ω^m si la cota β es la solución de la ecuación

$$\mathcal{F}_{\Gamma_p}(1 + \beta) - \mathcal{F}_{\Gamma_p}(1 - \beta) = \omega \quad (3.20)$$

donde \mathcal{F}_{Γ_p} denota la función de distribución acumulada de una variable aleatoria $\Gamma((p-1)/2, (p+1)/2)$. Supongamos que $\beta < 1$, o equivalentemente, $\omega^m < \omega_{\max}(p) = (\mathcal{F}_{\Gamma_p}(2))^m$. Entonces se tiene

$$\begin{aligned} & \|\Sigma_{\text{ml}}^{-1/2}(\Sigma - \Sigma_{\text{ml}})\Sigma_{\text{ml}}^{-1/2}\| & (3.21) \\ &= \|\Sigma_{\text{ml}}^{-1/2}\Sigma^{1/2}\Sigma^{1/2}(\Sigma - \Sigma_{\text{ml}})\Sigma^{-1/2}\Sigma^{-1/2}\Sigma_{\text{ml}}^{-1/2}\| \\ &\leq \|\Sigma_{\text{ml}}^{-1/2}\Sigma^{1/2}\| \|\Sigma^{-1/2}(\Sigma - \Sigma_{\text{ml}})\Sigma^{-1/2}\| \|\Sigma^{1/2}\Sigma_{\text{ml}}^{-1/2}\| \\ &\leq \frac{\beta}{1 - \beta}, \end{aligned}$$

donde (3.21) se sigue del hecho de que $\Sigma_{\text{ml}} \in \mathcal{J}$, por tanto de (3.21) se tiene que

$$\mathcal{S} = \left\{ \Sigma : \Sigma_{\text{ml}} + \Delta \succeq \mathbf{0}, \Delta = \Delta^t, \|\Sigma_{\text{ml}}^{-1/2} \Delta \Sigma_{\text{ml}}^{-1/2}\| \leq \frac{\beta}{1-\beta} \right\}$$

contienen un ω^n -región de confianza para la matriz de covarianza real Σ ; es decir, una parametrización natural del conjunto de incertidumbre en (3.18) es $\Sigma_0 = \mathbf{N} = \Sigma_{\text{ml}}$ y $\eta = \frac{\beta}{1-\beta}$, así el conjunto de incertidumbre en (3.18) queda completamente parametrizado considerando la región de confianza alrededor del estimador de máxima verosimilitud.

3.3.2. Conjunto de incertidumbre para $\vec{\mu}$

Recordemos que el conjunto de incertidumbre para el vector de medias de los rendimientos $\vec{\mu}$ se define como

$$S_m = \{ \vec{\mu} : \vec{\mu} = \vec{\mu}_0 + \vec{\zeta} \in \mathbb{R}^n, |\zeta_i| \leq \gamma_i, i = 1, \dots, n \} \quad (3.22)$$

El método de estimadores de máxima verosimilitud establece que a partir del estadístico de Hotelling se puede obtener una región de confianza al $(1 - \alpha)100\%$ para el vector de medias de una normal multivariada con matriz de varianza-covarianza desconocida (ver Análisis Multivariante de César Sánchez [?]) de la siguiente forma:

$$\{ \vec{\mu} \in \mathbb{R}^p : \frac{n-p}{p} (\vec{\mu}_0 - \vec{\mu})^t \Sigma_0^{-1} (\vec{\mu}_0 - \vec{\mu}) \leq f_{p,n-p,\alpha} \} \quad (3.23)$$

donde Σ_0 es la matriz de varianza-covarianza muestral y $\vec{\mu}_0$ es el vector de medias calculado de los datos. Esta región constituye un elipsoide en \mathbb{R}^p centrado en $\vec{\mu}_0$ cuyos ejes van en la dirección de los eigenvalores de Σ_0 y la longitud de los radios es

$$\sqrt{\lambda_i} \sqrt{\frac{n-p}{p} f_{p,n-p,\alpha}}.$$

Partiendo de lo anterior existen tres métodos diferentes para obtener el conjunto de incertidumbre S_m .

T de Student

El método de la T de Student es un método muy eficiente para calcular intervalos de confianza para las componentes del vector $\vec{\mu}$ el cuál asegura que los intervalos de confianza para la componentes de $\vec{\mu}$ están definidos por

$$\left(\vec{\mu}_0 - t_{n-1, \alpha/2} \frac{\sqrt{\text{diag}(\Sigma_0)}}{\sqrt{n}}, \vec{\mu}_0 + t_{n-1, \alpha/2} \frac{\sqrt{\text{diag}(\Sigma_0)}}{\sqrt{n}} \right).$$

con $t_{n-1, \alpha/2}$ el cuantil $1 - \alpha/2$ de la distribución T de Student con $n - 1$ grados de libertad. De este modo el conjunto de intervalos contienen a $\vec{\mu}$ con una probabilidad de $1 - \alpha$. El nivel de confianza se refiere a la probabilidad individual de cada intervalo por lo que la probabilidad de que todos los intervalos simultaneamente contengas a la componente correspondiente de $\vec{\mu}$ será en general menor al nivel de confianza fijado.

Método de Scheffé

Para poder satisfacer el nivel de confianza simultaneo es preciso modificar la construcción de los intervalos haciendolos mas amplios. Una forma eficiente de ampliarlos es cambiando el valor $t_{n-1, \alpha/2}$ por una constante adecuada. El método para obtener estos intervalos es conocido como *Método de Scheffé* el cual define los nuevos intervalos de confianza simultaneos como

$$\left(\vec{\mu}_0 - \sqrt{\frac{p}{n-p} f_{p, n-p, \alpha}} \frac{\sqrt{\text{diag}(\Sigma_0)}}{\sqrt{n}}, \vec{\mu}_0 + \sqrt{\frac{p}{n-p} f_{p, n-p, \alpha}} \frac{\sqrt{\text{diag}(\Sigma_0)}}{\sqrt{n}} \right).$$

Método de Bonferroni

Otro método para obtener intervalos de confianza simultaneos es el *Método de Bonferroni*. Es una alterativa válida cuando se requieren una cantidad finita de intervalos simultaneos ya que no está basado en la naturaleza probabilistica del problema en cuestión (a diferencia del mpetodo

de Scheffé). El fundamento de este método radica en la subaditividad de la probabilidad.

En este caso los intervalos de confianza se determinan de la siguiente manera

$$\left(\vec{\mu}_0 - t_{n-1, \alpha/2p} \frac{\sqrt{\text{diag}(\boldsymbol{\Sigma}_0)}}{\sqrt{n}}, \vec{\mu}_0 + t_{n-1, \alpha/2p} \frac{\sqrt{\text{diag}(\boldsymbol{\Sigma}_0)}}{\sqrt{n}} \right).$$

3.4. Modelo factorial

En esta sección se presentará un procedimiento para obtener el portafolio robusto de mínima varianza desde un enfoque estadístico, planteando los rendimientos de los activos como un modelo de factores. En este caso asumiremos nuevamente que el rendimiento \vec{r} en un periodo simple es una variable aleatoria dada por

$$\vec{r} = \vec{\mu} + \mathbf{B}^t \vec{f} + \vec{\epsilon} \quad (3.24)$$

donde $\vec{\mu} \in \mathbb{R}^n$ es el vector de medias de los rendimientos, $\vec{f} \sim \mathcal{N}(\mathbf{0}, \mathbf{F}) \in \mathbb{R}^n$ es el vector de factores que explican los rendimientos, $\mathbf{B} \in \mathbb{R}^{m \times n}$ es la matriz de carga de los factores y $\vec{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$ es el vector de residuos de los rendimientos. Otro supuesto importante que haremos es que el vector de residuos $\vec{\epsilon}$ es independiente del vector de factores \vec{f} . La matriz de covarianza \mathbf{F} es positiva definida y la matriz de covarianza $\mathbf{D} = \text{diag}(\vec{d})$ es semidefinida positiva. Con lo anterior definido, vamos a suponer que el vector de rendimientos es una variable aleatoria normal multivariada con media $\vec{\mu}$ y matriz de covarianza $\mathbf{B}^t \mathbf{F} \mathbf{B} + \mathbf{D}$ lo cual se denota como sigue

$$\vec{r} \sim \mathcal{N}(\vec{\mu}, \mathbf{B}^t \mathbf{F} \mathbf{B} + \mathbf{D}).$$

Supongamos que los elementos diagonales d_i de la matriz de covarianzas \mathbf{D} yacen en un intervalo $[d_i, \bar{d}_i]$ es decir, el conjunto de incertidumbre S_d para la matriz \mathbf{D} se define como sigue

$$S_d = \{\mathbf{D} : \mathbf{D} = \text{diag}(\vec{d}), d_i \in [d_i, \bar{d}_i], i = 1, \dots, n\}. \quad (3.25)$$

Supongamos también que las columnas de la matriz \mathbf{B} son conocidas y que \mathbf{B} pertenece al conjunto elíptico de incertidumbre S_b dado por

$$S_b = \{\mathbf{B} : \mathbf{B} = \mathbf{B}_0 + \mathbf{W}, \|\mathbf{W}_i\|_g \leq \rho_i, i = 1, \dots, n\}, \quad (3.26)$$

donde \mathbf{W}_i es la i -ésima columna de la matriz \mathbf{W} y $\|\mathbf{W}\|_g = \sqrt{\mathbf{W}^t \mathbf{G} \mathbf{W}}$ denota la norma elíptica de \mathbf{W} con respecto a una matriz simétrica y positiva definida \mathbf{G} .

Asumiremos que el vector de medias de los rendimientos pertenece al conjunto de incertidumbre S_m dado por

$$S_m = \{\vec{\mu} : \vec{\mu} = \vec{\mu}_0 + \vec{\xi}, |\xi_i| \leq \gamma_i, i = 1, \dots, n\}, \quad (3.27)$$

así cada componente de $\vec{\mu}$ yace dentro de cierto intervalo. La elección de los conjuntos de incertidumbre es motivada por el hecho de que la matriz de carga factorial y el vector de medias de los rendimientos $\vec{\mu}$ son estimados por el método de regresión lineal. La justificación de la estructura de los conjuntos de incertidumbre así como la elección adecuada de la matriz \mathbf{G} y las cotas $\gamma_i, \rho_i, \bar{d}_i, d_i, i = 1, \dots, n$ serán discutidos más adelante.

Puesto que

$$\vec{r}_\phi \sim \mathcal{N}(\vec{\phi}^t \vec{\mu}, \vec{\phi}^t (\mathbf{B}^t \mathbf{F} \mathbf{B} + \mathbf{D}) \vec{\phi}),$$

el problema de mínima varianza está dado por

$$\begin{aligned} \min_{\{\mathbf{B} \in S_b\}} \max_{\{\mathbf{D} \in S_d\}} \{ \vec{\phi}^t \mathbf{B}^t \mathbf{F} \mathbf{B} \} + \max_{\{\mathbf{D} \in S_d\}} \{ \vec{\phi}^t \mathbf{D} \vec{\phi} \}, \\ \text{sujeto a } \min_{\{\vec{\mu} \in S_m\}} \vec{\mu}^t \vec{\phi} \geq \alpha, \\ \vec{1}^t \vec{\phi} = 1. \end{aligned} \quad (3.28)$$

en este caso las cotas $\underline{d}_i \leq d_i \leq \bar{d}_i$ implican que $\vec{\phi}^t \mathbf{D} \vec{\phi} \leq \vec{\phi}^t \bar{\mathbf{D}} \vec{\phi}$, donde $\bar{\mathbf{D}} = \text{diag}(\bar{d})$. Como hemos asumido que la matriz de covarianza de \vec{f} es estrictamente positiva definida, la función $\|x\|_f : x \mapsto \sqrt{x^t \mathbf{F} x}$ define una norma en \mathbb{R}^m . Así (3.22) es equivalente al siguiente problema robusto de mínimos cuadrados aumentado

$$\begin{aligned} \min \quad & \max_{\{\mathbf{B} \in S_b\}} \|\mathbf{B} \vec{\phi}\|_f^2 + \vec{\phi}^t \bar{\mathbf{D}} \vec{\phi}, & (3.29) \\ \text{sujeto a} \quad & \min_{\{\vec{\mu} \in S_m\}} \vec{\mu}^t \vec{\phi} \geq \alpha, \\ & \vec{\mathbf{1}}^t \vec{\phi} = 1. \end{aligned}$$

al introducir variables auxiliares ν y δ , podemos reformular (3.23) como sigue

$$\begin{aligned} \text{minimizar} \quad & \nu + \delta & (3.30) \\ \text{s.a} \quad & \max_{\{\mathbf{B} \in S_b\}} \|\mathbf{B} \vec{\phi}\|_f^2 \leq \nu, \\ & \vec{\phi}^t \bar{\mathbf{D}} \vec{\phi} \leq \delta, \\ & \min_{\{\vec{\mu} \in S_m\}} \vec{\mu}^t \vec{\phi} \geq \alpha, \\ & \vec{\mathbf{1}}^t \vec{\phi} = 1. \end{aligned}$$

Si suponemos que S_m y S_b son conjuntos finitos, entonces el problema se transforma en el siguiente

$$\begin{aligned} \min \quad & \lambda + \delta & (3.31) \\ \text{s. a} \quad & \max_{\{\mathbf{B} \in S_b\}} \|\mathbf{B}_k \vec{\phi}\|_f^2 \leq \lambda, \quad \forall k = 1, \dots, s, \\ & \vec{\phi}^t \bar{\mathbf{D}} \vec{\phi} \leq \delta, \\ & \vec{\mu}_k^t \vec{\phi} \geq \alpha, \quad \forall k = 1, \dots, r, \end{aligned}$$

$$\vec{\mathbf{1}}^t \vec{\phi} = 1.$$

al igual que los problemas anteriores este problema puede transformarse fácilmente en un SOCP.

En el artículo de El Ghaoui and Le Bret [4] se demostró que para el conjunto

$$S_b = \{\mathbf{B} : \mathbf{B} = \mathbf{B}_0 + \mathbf{W}, \|\mathbf{W}\| = \sqrt{\text{Tr}(\mathbf{W}^t \mathbf{W})} \leq \rho\}$$

el problema puede ser aun transformado en un SOCP.

Cuando los conjuntos de incertidumbre S_m y S_b son definidos como en (3.20) y (3.21) el peor caso de los rendimientos de un portafolio fijo está dado por

$$\min_{\vec{\mu} \in S_m} \vec{\mu}^t \vec{\phi} = \mu_0^t \vec{\phi} - \vec{\gamma}^t |\vec{\phi}|; \quad (3.32)$$

y el peor caso de la varianza es el siguiente

$$\begin{aligned} &\text{maximizar } \|(\mathbf{B}_0 + \mathbf{W})\vec{\phi}\|_f^2, \\ &\text{sujeto } \|\mathbf{W}_i\|_g \leq \rho_i, \quad i = 1, \dots, n. \end{aligned} \quad (3.33)$$

Ya que las restricciones $\|\mathbf{W}_i\|_g \leq \rho_i, \quad i = 1, \dots, N$ implican que

$$\|\mathbf{W}\vec{\phi}\|_g = \left\| \sum_{i=1}^N \phi_i \mathbf{W}_i \right\|_g \leq \sum_{i=1}^N |\phi_i| \|\mathbf{W}_i\|_g \leq \sum_{i=1}^N \rho_i |\phi_i|, \quad (3.34)$$

el problema de optimización se transforma en el siguiente.

$$\begin{aligned} &\text{maximizar } \|\mathbf{B}_0 \vec{\phi} + \mathbf{w}\|_f^2, \\ &\text{sujeto a } \|\mathbf{w}\|_g \leq r, \end{aligned} \quad (3.35)$$

donde $r = \bar{\rho}^t |\vec{\phi}| = \sum_{i=1}^N \rho_i |\phi_i|$, lo cual implica que el valor óptimo de (3.26) satisface (3.24).

La función objetivo de (3.29) es convexa, por tanto, la solución óptima \mathbf{w}^* yace en la frontera del conjunto admisible, i.e., $\|\mathbf{w}^*\|_g = r$.

Para $i = 1, 2, \dots, n$, definimos

$$\mathbf{W}_i^* = \begin{cases} \frac{\phi_i}{|\phi_i|} \frac{\rho_i}{r} \mathbf{w}^*, & \phi_i \neq 0 \\ \frac{\phi_i}{r} \mathbf{w}^*, & \text{en otro caso.} \end{cases} \quad (3.36)$$

Es claro que $\|\mathbf{W}_i^*\|_g = \rho_i$, $\forall i = 1, \dots, N$, entonces, \mathbf{W}^* es una solución de (3.24), además $\mathbf{W}^* \vec{\phi} = \sum_{i=1}^N \phi_i \mathbf{W}_i^* = \mathbf{w}^*$. Por tanto el valor óptimo de (3.27) y (3.29) es el mismo. Así, para un portafolio fijo $\vec{\phi}$, el peor caso de la varianza es menor que ν si y solo si,

$$\max_{\{\mathbf{y}: \|\mathbf{y}\|_g \leq r\}} \|\mathbf{y}_0 + \mathbf{y}\|_f^2 \leq \nu \quad (3.37)$$

donde $\mathbf{y}_0 = \mathbf{B}_0 \vec{\phi}$ y $r = \bar{\rho}^t |\vec{\phi}|$.

Los siguientes lemas resultan de gran utilidad para transformar la restricción (3.31) en una colección de restricciones cónicas de segundo orden, lineales de igualdad y desigualdad.

Lema 3.4.1. Sean $r, \nu > 0$, $\mathbf{y}_0, \mathbf{y} \in \mathbb{R}^m$ y $\mathbf{F}, \mathbf{G} \in \mathbb{R}^{m \times m}$ matrices positivas definidas. Entonces la restricción

$$\max_{\{\mathbf{y}: \|\mathbf{y}\|_g \leq r\}} \|\mathbf{y}_0 + \mathbf{y}\|_f^2 \leq \nu \quad (3.38)$$

es equivalente a la siguiente colección de restricciones:

Existe $\tau, \sigma \geq 0$ y $\mathbf{t} \in \mathbb{R}_+^m$ que satisface

$$\begin{aligned}
\nu &\geq \tau + \vec{\mathbf{1}}^t \mathbf{t}, \\
\sigma &\leq \frac{1}{\lambda_{\max}(\mathbf{H})}, \\
r^2 &\leq \sigma \tau, \\
w_i^2 &\leq (1 - \sigma \lambda_i) t_i, \quad i = 1, \dots, m.
\end{aligned} \tag{3.39}$$

donde $\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^t$ es la descomposición espectral de $\mathbf{H} = \mathbf{G}^{-1/2}\mathbf{F}\mathbf{G}^{-1/2}$, $\mathbf{\Lambda} = \text{diag}(\lambda_i)$ y $\mathbf{w} = \mathbf{Q}^t\mathbf{H}^{1/2}\mathbf{G}^{1/2}\mathbf{y}_0$.

Lema 3.4.2 (Procedimiento \mathcal{S}). Sea $F_i(x) = \vec{x}^t \mathbf{A}_i \vec{x} + 2\vec{b}_i^t \vec{x} + c_i$, $i = 1, \dots, p$, funciones cuadráticas de $\vec{x} \in \mathbb{R}^n$. Entonces $\mathbf{F}_0(\vec{x}) \geq 0 \forall \vec{x}$ tal que $\mathbf{F}_i(\vec{x}) \geq 0$, $i = 1, \dots, p$, si existe $\tau_i \geq 0$ tal que

$$\begin{bmatrix} c_0 & \mathbf{b}_0^t \\ \mathbf{b}_0 & \mathbf{A}_0 \end{bmatrix} - \sum_{i=1}^p \tau_i \begin{bmatrix} c_i & \mathbf{b}_i^t \\ \mathbf{b}_i & \mathbf{A}_i \end{bmatrix} \succeq 0.$$

Además, si $p = 1$ se satisface la contrario si existe \vec{x}_0 tal que $\mathbf{F}_1(\vec{x}_0) > 0$.

La prueba de este lema puede encontrarse en [2] (Convex Optimization pg. 657 y 658).

Demostración. (Lema 3.3.1.)

Haciendo $\vec{y} = r\bar{y}$ y usando el hecho de que $\|\vec{x}\|^2 = \vec{x} \cdot \vec{x}$ se tiene que

$$\max_{\{\vec{y}: \|\vec{y}\|_g \leq r\}} \|\vec{y}_0 + \vec{y}\|_f^2 \leq \nu$$

es equivalente a

$$\vec{y}_0^t \mathbf{F} \vec{y}_0 - 2r\vec{y}_0^t \mathbf{F} \bar{y} - r^2 \bar{y}^t \mathbf{F} \bar{y} + \nu \geq 0 \tag{3.40}$$

para toda \bar{y} tal que $1 - \bar{y}^t \mathbf{G} \bar{y} \geq 0$.

Como $\bar{y} = 0$ es estrictamente factible para $1 - \bar{y}^t \mathbf{G} \bar{y} \geq 0$, entonces el procedimiento \mathcal{S} implica que (3.34) se satisface para todo $1 - \bar{y}^t \mathbf{G} \bar{y} \geq 0$ si y solo si existe $\tau \geq 0$ tal que

$$\mathbf{M} = \begin{bmatrix} \nu - \tau - \bar{y}_0^t \mathbf{F} \bar{y}_0 & -r \bar{y}_0^t \mathbf{F} \\ -r \mathbf{F} \bar{y}_0 & \tau \mathbf{G} - r^2 \mathbf{F} \end{bmatrix} \succeq 0. \quad (3.41)$$

Sea $\mathbf{H} = \mathbf{G}^{-1/2} \mathbf{F} \mathbf{G}^{-1/2}$ y su descomposición espectral $\mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^t$, con $\mathbf{\Lambda} = \text{diag}(\lambda)$ y definimos $\vec{w} = \mathbf{Q}^t \mathbf{H}^{1/2} \mathbf{G}^{1/2} \bar{y}_0 = \mathbf{\Lambda}^{1/2} \mathbf{Q}^t \mathbf{G}^{1/2} \bar{y}_0$. Observemos que $\bar{y}_0^t \mathbf{F} \bar{y}_0 = \vec{w}^t \vec{w}$ entonces la matriz $\mathbf{M} \succeq 0$ si y sólo si

$$\begin{aligned} \bar{\mathbf{M}} &= \begin{bmatrix} 1 & 0^t \\ 0 & \mathbf{Q}^t \mathbf{G}^{-1/2} \end{bmatrix} \mathbf{M} \begin{bmatrix} 1 & 0^t \\ 0 & \mathbf{G}^{-1/2} \mathbf{Q} \end{bmatrix} \\ &= \begin{bmatrix} \nu - \tau - \vec{w}^t \vec{w} & -r \vec{w}^t \mathbf{\Lambda}^{1/2} \\ -r \mathbf{\Lambda} \vec{w} & \tau \mathbf{I} - r^2 \mathbf{\Lambda} \end{bmatrix} \\ &\succeq 0. \end{aligned}$$

El complemento de Schur de los renglones y columnas distintas de cero de $\tau \mathbf{I} - r^2 \mathbf{\Lambda}$ está dado por

$$\nu - \tau - \vec{w}^t \vec{w} - r^2 \left[\sum_{i: \tau \neq r^2 \lambda_i} \frac{w_i^2 \lambda_i}{\tau - r^2 \lambda_i} \right] = \nu - \tau - \sum_{i: \sigma \lambda_i \neq 1} \frac{w_i^2}{1 - \sigma \lambda_i} \quad (3.42)$$

donde $\sigma = r^2 / \tau$.

Por tanto la matriz $\bar{\mathbf{M}}$ es semidefinida positiva si $\tau \geq r^2 \lambda_i, \forall i = 1, \dots, m$, $w_i = 0$ para toda i tal que $\tau = r^2 \lambda_i$ y

$$\nu - \tau - \sum_{i: \sigma \lambda_i \neq 1} \frac{w_i^2}{1 - \sigma \lambda_i} \geq 0.$$

Se sigue que

$$\max_{\{y: \|y\|_g \leq r\}} \|\mathbf{y}_0 + \mathbf{y}\|_f^2 \leq \nu$$

si y sólo si existen σ , $\tau \geq 0$ y $\vec{t} \in \mathbb{R}_+^m$ tal que

$$\begin{aligned} \nu &\geq \tau + \vec{\Gamma}^t \vec{t}, \\ r^2 &= \sigma\tau, \\ w_i &= (1 - \sigma\lambda_i)t_i, \quad i = 1, 2, \dots, m, \\ \sigma &\leq \frac{1}{\lambda_{\max}(\mathbf{H})}. \end{aligned}$$

□

Definición 1. Dado $\mathbf{B}_0 \in \mathbb{R}^{m \times n}$ y $\mathbf{F}, \mathbf{G} \in \mathbb{R}^{m \times m}$ positivas definidas, se define $\mathcal{H}(\mathbf{B}_0, \mathbf{F}, \mathbf{G})$ como el conjunto de todos los vectores $(r; \nu; \vec{\phi}) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^n$ tal que $(r, \nu, \vec{y}_0 = \mathbf{B}_0 \vec{\phi})$ satisfacen

$$\begin{aligned} \nu &\geq \tau + \vec{\Gamma}^t \mathbf{t}, \\ \sigma &\leq \frac{1}{\lambda_{\max}(\mathbf{H})}, \\ \left\| \begin{bmatrix} 2r \\ \sigma - \tau \end{bmatrix} \right\| &\leq \sigma + \tau, \\ \left\| \begin{bmatrix} 2w_i \\ 1 - \sigma\lambda_i - t_i \end{bmatrix} \right\| &\leq 1 - \sigma\lambda_i + t_i, \quad i = 1, \dots, m. \end{aligned} \tag{3.43}$$

Donde $\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^t$ es la descomposición espectral de $\mathbf{H} = \mathbf{G}^{-1/2}\mathbf{F}\mathbf{G}^{-1/2}$, $\mathbf{\Lambda} = \text{diag}(\lambda_i)$ y $\mathbf{w} = \mathbf{Q}^t \mathbf{H}^{1/2} \mathbf{G}^{1/2} \mathbf{y}_0$.

Puesto que no se permiten ventas en corto, de (3.26), (3.32) y la definición (1) se sigue que (3.23) puede ser reformulado como sigue

$$\min \nu + \delta, \tag{3.44}$$

$$\begin{aligned}
\text{s. a } & \left\| \begin{bmatrix} 2\bar{\mathbf{D}}^{1/2}\vec{\phi} \\ 1 - \delta \end{bmatrix} \right\| \leq 1 + \delta, \\
& (\vec{\mu}_0^t - \vec{\gamma}^t)\vec{\phi} \geq \alpha, \\
& \vec{\mathbf{1}}^t\vec{\phi} = 1, \\
& (r; \nu; \vec{\phi}) \in \mathcal{H}(\mathbf{B}_0, \mathbf{F}, \mathbf{G}).
\end{aligned}$$

este problema es también un SOCP.

La construcción de los conjuntos de incertidumbre en el caso del modelo factorial puede ser consultada en [6] (Robust Selection Problems).

Capítulo 4

Resultados numéricos

En esta sección se presentarán los resultados obtenidos al resolver numéricamente alguno de los problemas teóricos planteados en el capítulo anterior. El objetivo principal de estas pruebas computacionales es el de hacer una comparación entre el resultado de calcular el portafolio óptimo desde un enfoque clásico y el de realizarlo desde el enfoque robusto. En la primer parte se presentarán los resultados de calcular un portafolio robusto en el que se trabaja con conjuntos de incertidumbre finitos, el caso continuo se presentará en la segunda parte de este capítulo. En el caso del Modelo Factorial no fue posible obtener resultados numéricos consistentes debido a que la tarea de encontrar los factores adecuados no es tan simple, por tanto se deja este problema planteado para trabajos futuros. Para obtener los resultados numéricos se utilizó el paquete CVX de Matlab, desarrollado por Stephen P. Boyd y Michael C. Grant el cuál es un paquete con un gran alcance diseñado para resolver problemas de optimización convexa.

Se tomaron las series financieras correspondientes a los precios diarios de las acciones de seis empresas que operan en México (Liverpool, Elektra, Sports World, Walmart, Chedraui y Samborns). Ya que los datos de las series se tomaron diariamente se decidió construir la matriz de rendimientos y multiplicarla por 100 para no trabajar con números muy pequeños.

Con el objeto de tener un portafolio que sirviera de referencia para comparar los resultados obtenidos a través de los distintos métodos existentes para obtener un portafolio óptimo se tomó el periodo del 18 de enero al 17 de noviembre del 2017. La matriz de varianza-covarianza así como el vector de medias de los rendimientos obtenidos son los siguientes:

$$\Sigma = \begin{bmatrix} 2.2018 & -0.2952 & -0.2541 & 0.1852 & 0.3172 & 0.0037 \\ -0.2952 & 12.5512 & 0.0125 & 0.3053 & -0.2131 & 0.0758 \\ -0.2541 & 0.0125 & 3.1067 & -0.0978 & -0.0235 & 0.0375 \\ 0.1852 & 0.30523 & -0.0978 & 1.2780 & 0.1102 & 0.0657 \\ 0.3172 & -0.2131 & -0.0235 & 0.1102 & 2.1764 & -0.1835 \\ 0.0037 & 0.0758 & 0.0375 & 0.0657 & -0.1835 & 1.6405 \end{bmatrix}, \vec{\mu} = \begin{bmatrix} -0.0372 \\ 0.5222 \\ 0.0924 \\ 0.1176 \\ 0.0068 \\ -0.0124 \end{bmatrix}.$$

Con estos datos se calculó el portafolio óptimo desde el enfoque clásico, el cual se muestra a continuación. Se incluye también su volatilidad y rendimiento asociados.

$$\vec{\phi}_R = (0.1021, 0.0782, 0.1614, 0.3342, 0.1442, 0.1799), \sigma = 0.67, r = 0.09.$$

Los datos utilizados para probar los modelos propuestos en esta tesis fueron los de las acciones de las empresas ya mencionadas. El periodo que se tomó en este caso fue del 2 de enero al 31 octubre (dos semanas antes del periodo de referencia). La matriz de varianza-covarianza y el vector de medias de los rendimientos para este periodo son los siguientes:

$$\Sigma_0 = \begin{bmatrix} 2.3565 & -0.2080 & -0.2279 & 0.2377 & 0.4238 & 0.0618 \\ -0.2080 & 12.4701 & 0.0568 & 0.3041 & -0.1741 & 0.0649 \\ -0.2279 & 0.0568 & 3.4988 & -0.0854 & -0.0334 & 0.0480 \\ 0.2377 & 0.3041 & -0.0854 & 1.3310 & 0.2462 & 0.0819 \\ 0.4238 & -0.1741 & -0.0334 & 0.2462 & 2.5592 & -0.1647 \\ 0.0618 & 0.0649 & 0.0481 & 0.0820 & -0.1647 & 1.6820 \end{bmatrix}, \vec{\mu}_0 = \begin{bmatrix} -0.0629 \\ 0.5020 \\ 0.0486 \\ 0.0649 \\ 0.0034 \\ -0.0126 \end{bmatrix}.$$

4.1. Conjuntos de incertidumbre finitos

En esta sección se realizaron los experimentos numéricos necesarios para probar la metodología propuesta en el caso en que los conjuntos de incertidumbre se asumen finitos, se resolvió el problema (3.8) planteado en el capítulo 3. En este caso se calcularon cuatro diferentes portafolios, en cada uno se plantearon cuatro distintos escenarios para la matriz de varianza-covarianza Σ y cuatro escenarios para el vector de medias $\vec{\mu}$, el valor de $\alpha = 0.09$ fue elegido tomando la media de las entradas de $\vec{\mu}_0$. En las siguientes tablas se muestran los resultados obtenidos.

$\vec{\phi}_1$	(0.0431, 0.1378, 0.1535, 0.3532, 0.1302, 0.1822)
Σ	$\Sigma_{1,2} = \Sigma_0 \pm 5 \times 10^{-2}$, $\Sigma_{4,5} = \Sigma_0 \pm 1 \times 10^{-2}$
$\vec{\mu}$	$\vec{\mu}_{1,2} = \vec{\mu}_0 \pm 5 \times 10^{-3}$, $\vec{\mu}_{3,4} = \vec{\mu}_0 \pm 1 \times 10^{-3}$
σ	0.8324
r	0.0950

Tabla 4.1

$\vec{\phi}_2$	(0.0502, 0.1307, 0.1572, 0.3471, 0.1313, 0.1880)
Σ	$\Sigma_{1,2} = \Sigma_0 \pm 5 \times 10^{-2}$, $\Sigma_{4,5} = \Sigma_0 \pm 1 \times 10^{-2}$
$\vec{\mu}$	$\vec{\mu}_{1,2} = \vec{\mu}_0 \pm 5 \times 10^{-4}$, $\vec{\mu}_{3,4} = \vec{\mu}_0 \pm 1 \times 10^{-4}$
σ	0.8172
r	0.0905

Tabla 4.2

$\vec{\phi}_3$	(0.0509, 0.1299, 0.1572, 0.3471, 0.1313 0.1880)
Σ	$\Sigma_{1,2} = \Sigma_0 \pm 5 \times 10^{-4}$, $\Sigma_{4,5} = \Sigma_0 \pm 1 \times 10^{-4}$
$\vec{\mu}$	$\vec{\mu}_{1,2} = \vec{\mu}_0 \pm 5 \times 10^{-5}$, $\vec{\mu}_{3,4} = \vec{\mu}_0 \pm 1 \times 10^{-5}$
σ	0.7848
r	0.0900

Tabla 4.3

$\vec{\phi}_4$	(0, 0.2140, 0.1597, 0.3994, 0.1086, 0.1181)
Σ	$\Sigma_{1,2} = \Sigma_0 \pm 5 \times 10^{-6}$, $\Sigma_{4,5} = \Sigma_0 \pm 1 \times 10^{-6}$
$\vec{\mu}$	$\vec{\mu}_{1,2} = \vec{\mu}_0 \pm 5 \times 10^{-2}$, $\vec{\mu}_{3,4} = \vec{\mu}_0 \pm 1 \times 10^{-2}$
σ	0.9961
r	0.14

Tabla 4.4

En las Tablas 4.1 y 4.2 podemos notar que los portafolios obtenidos tienen volatilidades similares ya que las perturbaciones realizadas en estos casos son cercanas. El portafolio 4 tiene el riesgo más alto, esto es resultado de que la perturbación realizada en el vector de medias fue muy grande en relación a los otros casos. Como podemos notar el portafolio con el mejor compartamiento fue el portafolio 3 ya que tiene el menor riesgo de los 4 con un rendimiento de 0.09 que coincide con el que se obtuvo en el portafolio de referencia.

4.2. Conjuntos de incertidumbre no finitos

En esta sección se presentan los resultados obtenidos al suponer que los conjuntos de incertidumbre no son finitos. Los datos utilizados fueron

los mismos que se presentaron al inicio de este capítulo. Se utilizó el mismo portafolio $\vec{\phi}_R$ de referencia obtenido anteriormente para comparar los resultados.

4.2.1. Portafolio robusto de mínima varianza

En lo sucesivo se presentarán los resultados obtenidos al resolver numéricamente el problema robusto de mínima varianza (3.14) que se desarrolló en el capítulo anterior. Para resolver este problema es necesario definir una matriz \mathbf{N} , un vector $\vec{\gamma}$ y un escalar η . Recordemos que en este caso la matriz \mathbf{N} puede ser tomada como la matriz Σ_0 .

En la sección 3.3 se desarrolló un método para obtener una región de confianza para la matriz de varianza-covarianza Σ lo cual se utilizó para obtener el valor de $\eta = 0.3647$. En el capítulo 3 se desarrollaron tres métodos diferentes para calcular el vector $\vec{\gamma}$ (Bonferroni, T de Student y Scheffé), en la Tabla 4.5 se muestran los resultados obtenidos en cada método

Método	$\vec{\gamma}$
Bonferroni (1)	(0.1085, 0.2496, 0.1322, 0.0815, 0.1131, 0.09168)
T de Student (2)	(0.2059, 0.4735, 0.2509, 0.1547, 0.2145, 0.1739)
Scheffé (3)	(0.3789, 0.8715, 0.4616, 0.2847, 0.3948, 0.3200)

Tabla 4.5

Usando estos resultados para los conjuntos de incertidumbre del vector de medias se calculó el portafolio robusto en cada caso. En la siguiente tabla se muestran los portafolios obtenidos así como el riesgo y el rendimiento asociado a cada uno.

	σ	r	$\vec{\phi}$
\mathcal{C}	0.7844	0.0900	(0.0510 0.1299, 0.1527, 0.3471, 0.131 0.1880)
B	1.88	0.23	(0, 0.3965, 0, 0.6034, 0, 0)
T	∞	0	(0, 0, 0, 0, 0, 0)
S	∞	0	(0, 0, 0, 0, 0, 0)

Tabla 4.6

En la Tabla 4.5 podemos observar que tan grande son los conjuntos de incertidumbre para el parámetro $\vec{\mu}$. El método de Bonferroni aporta regiones más pequeñas. De la tabla 4.6 podemos ver que solo en el caso de Bonferroni se obtuvo un portafolio diferente del trivial lo cual se debe a que la perturbación máxima que se hace al parámetro es menor que en los otros dos casos. Como podemos notar el tamaño del conjunto de incertidumbre si es de relevancia en este problema.

En la Figura 4.1 se graficó la frontera eficiente de los datos reales, es decir, el periodo del 18 de enero al 17 de noviembre del 2017. Se graficó el riesgo y el rendimiento asociados al portafolio obtenido mediante el método clásico de Markowitz y el obtenido mediante el método de Bonferroni.

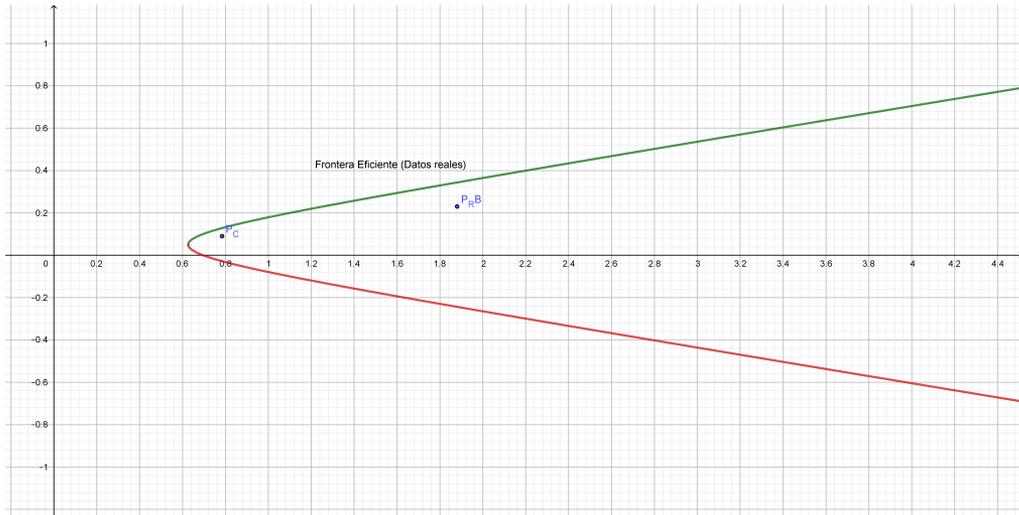


Figura 4.1: Problema robusto de mínima varianza.

4.2.2. Portafolio robusto de máximo rendimiento

En esta sección se presentarán los resultados obtenidos al resolver el problema robusto de máximo rendimiento (3.17) desarrollado en el capítulo anterior. Los datos utilizados para obtener el portafolio de referencia y el portafolio robusto son los mismos que en el problema robusto de mínima varianza. Recordemos que en este problema es necesario fijar una cota λ para la varianza de el portafolio, en este caso el valor de $\lambda = 3.9830$ fue elegido tomando la media de las varianzas de cada activo. El portafolio de referencia obtenido en este caso es el siguiente:

$$\vec{\phi}_r = (0, 0.5403, 0.0874, 0.3723, 0, 0), \quad \sigma = 1.9957, \quad r = 0.3340.$$

Se calculó el problema clásico de máximo rendimiento así como tres portafolios robustos distintos correspondientes a los distintos métodos desarrollados para obtener el conjunto de incertidumbre para $\vec{\mu}$. En la Tabla 4.7 se muestran los resultados obtenidos.

	σ	r	$\vec{\phi}$
\mathcal{C}	1.9957	0.3000	(0, 0.5421, 0.1101, 0.3478, 0, 0)
B	1.7083	0.0999	(0, 0.4333, 0, 0.5666, 0, 0)
T	1.7083	-0.0386	(0, 0.4333, 0, 0.5666, 0, 0)
S	1.1537	-0.2199	(0, 0, 0, 0, 1, 0)

Tabla 4.7

En la tabla 4.7 podemos observar que en este caso el tamaño del conjunto de incertidumbre del parámetro $\vec{\mu}$ no trajo problemas ya que en los tres casos fue posible obtener una solución diferente de la trivial. A igual que en el problema anterior el método de Bonferroni dió mejores resultados, en este caso el riesgo del portafolio obtenido es menor que en el caso del portafolio de referencia. Nótese que los portafolios obtenidos vía T de Student y Bonferroni tienen el mismo riesgo sin embargo tienen distinto rendimiento. Como se explicó en los problemas anteriores, los resultados obtenidos dependen de que tanto se perturbe el vector de medias o la matriz de varianza-covarianza. En este caso el resultado nos dice que si somos más pesimistas respecto al comportamiento del vector de medias de los rendimientos, el resultado será un portafolio con mayor pérdida.

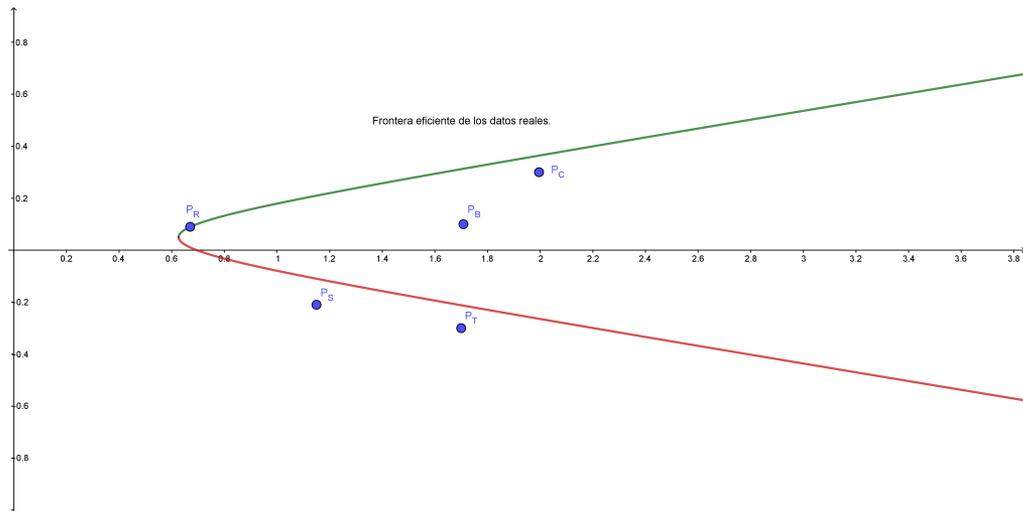


Figura 4.2: Problema robusto de máximo rendimiento.

En la Figura 4.2 se graficó la frontera eficiente de los datos reales y se muestra el riesgo y el rendimiento asociado a cada portafolio obtenido. Como podemos ver los portafolios obtenidos usando los conjuntos de incertidumbre construidos mediante la T de Student y Scheffé no resultaron satisfactorios pues no se ubican en la frontera eficiente de los datos. El portafolio obtenido mediante el modelo clásico se encuentra cerca de la frontera eficiente sin embargo su riesgo es mayor que el obtenido resolviendo el problema robusto de máximo rendimiento.

Conclusiones

El problema de selección de un portafolio óptimo es bastante susceptible a la incertidumbre que proviene de los datos que se utilizan para resolverlo por tanto es importante darle un tratamiento especial tomando en cuenta los conjuntos de incertidumbre de los parámetros que se utilizan. Al intentar trabajar este tipo de problemas es necesario replantearlos como problemas mini-max los cuales bajo ciertas condiciones pueden ser transformados en problemas lineales con restricciones cuadráticas o cónicas de segundo orden. Al realizar los cálculos para obtener el portafolio óptimo robusto se constató que no hay diferencia entre resolverlos como problemas con restricciones cuadráticas o como SOCP ya que tienen el mismo costo computacional.

En el caso en el que se asume que los conjuntos de incertidumbre son finitos se escogieron los distintos escenarios de manera arbitraria pero cabe mencionar que en los resultados obtenidos se pudo observar que el portafolio que aportó mejores resultados fue aquel en el que se definieron escenarios con variaciones muy pequeñas respecto al escenario original. Es importante mencionar esto ya que como se explicará más adelante el tamaño de la perturbación si influye en la solución del problema.

La construcción de los conjuntos de incertidumbre es una parte importante en este trabajo pues como se explicó en la sección 3.3 existen distintas maneras de calcular el conjunto de incertidumbre para el vector de medias

$\bar{\mu}$, sin embargo como se vió en la sección de resultados el tamaño de estos conjuntos sí resultó relevante en la resolución del problema robusto de mínima varianza. Los conjuntos obtenidos a partir de los métodos de la T de Student y de Scheffé no fueron del todo útiles ya que aportaron conjuntos de incertidumbre muy grandes de tal suerte que el problema de optimización no pudo ser resuelto. A pesar de lo antes mencionado en el caso del problema robusto de máximo rendimiento el tamaño de los intervalos no fue problema para poder obtener el portafolio. Cabe mencionar que los mejores resultados se obtuvieron utilizando el método de Bonferroni.

Resolver el problema del portafolio robusto desde el enfoque del modelo factorial es un problema complejo debido a que conseguir los factores adecuados es una tarea complicada. Uno de los principales obstáculos es elegir factores útiles, es decir, factores que expliquen satisfactoriamente el comportamiento de los rendimientos de las acciones. Otro de los problemas que se presentaron fue la falta de información ya que una vez definidos los factores nos encontramos con que la información con la que se cuenta en las redes es bastante limitada. Por los anteriores motivos el desarrollo de este problema de selección del portafolio robusto solo se estudió teóricamente dejando la aplicación para trabajos futuros.

Como pudo observarse en la sección 4.2 el portafolio de referencia tuvo una volatilidad de 0.67 y un rendimiento de 0.09. Sin embargo el portafolio robusto en el que se calculó el conjunto de incertidumbre de $\bar{\mu}$ mediante el método de Bonferroni arrojó una volatilidad de 1.88 y un rendimiento de 0.23. Esto significa que este es el peor escenario en la inversión al que se enfrentará el inversionista tomando en cuenta el peor comportamiento de los parámetros del modelo.

El problema de máximo rendimiento no presentó los problemas que presentó el problema de mínima varianza respecto al tamaño de los conjuntos de incertidumbre obtenidos mediante los diferentes métodos ya mencionados. Como se explicó en el párrafo anterior, los resultados obtenidos hablan

del resultado al que un inversionista puede exponerse si decide tomar en cuenta un comportamiento demasiado pesimista en los parámetros. Esto se ejemplifica en los resultados del modelo robusto de máximo rendimiento pues como puede verse en la tabla 4.7 el portafolio obtenido vía Bonferroni tiene la misma volatilidad que el obtenido vía T de Student. El rendimiento en el segundo caso corresponde a una pérdida lo cuál es resultado de haber tomado un conjunto de incertidumbre mas amplio para el vector de medias $\vec{\mu}$.

Bibliografía

- [1] ANDERSON, T.W. An introduction to multivariate statistical analysis. Wiley, Julio 2003.
- [2] BOYD, S. AND VANDENBERGHE, L. Convex Optimization. Cambridge University, 2004.
- [3] CHI, X. AND LIU, S. A one-step smoothing Newton method for second-order cone programming. Science Direct, Diciembre 2007.
- [4] EL GHAOUI, L., H, LEBRET. Robust solutions to least-squares problems with uncertain data. Society for Industrial and Applied Mathematics. J. Matrix, 1997. Anal. Appl. 18(4) 1035-1064.
- [5] GOLDFARB, D. AND IYENGAR, G. Robust portfolio selection problems. Mathematics of operations research. Vol. 28, No. 1, Febrero 2003, pag. 138.
- [6] GOLDFARB, D. AND IYENGAR, G. Robust convex quadratically constrained programs. Springer-Verlag , Mayo 2003.
- [7] LUENBERGER, D.G. Investment Science. New York, Oxford, 1998.
- [8] MCNEIL, A.J., FREY, R. AND EMBRECHTS, P. Quantitative Risk Management Concepts, Techniques and Tools. Ed. Princeton University, 2005.

- [9] SOUSA, M., VANDENBERGHE, L., BOYD, S., LEBRET, H. Applications of second order cone programming. Elsevier Science Inc, 1997.