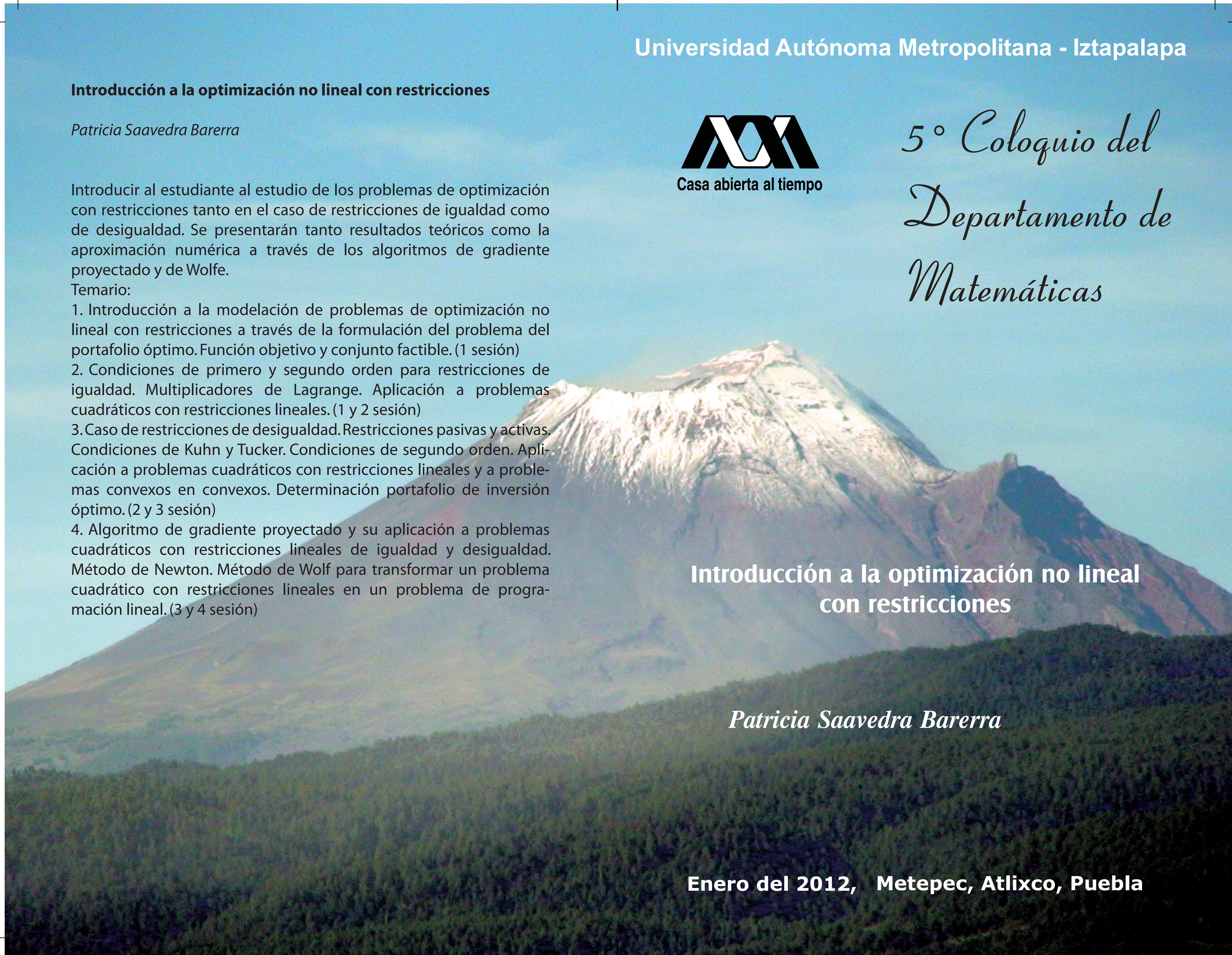

Introducción a la optimización no lineal con restricciones

Patricia Saavedra Barerra

Introducir al estudiante al estudio de los problemas de optimización con restricciones tanto en el caso de restricciones de igualdad como de desigualdad. Se presentarán tanto resultados teóricos como la aproximación numérica a través de los algoritmos de gradiente proyectado y de Wolfe.
Temario:
1. Introducción a la modelación de problemas de optimización no lineal con restricciones a través de la formulación del problema del portafolio óptimo. Función objetivo y conjunto factible. (1 sesión)
2. Condiciones de primero y segundo orden para restricciones de igualdad. Multiplicadores de Lagrange. Aplicación a problemas cuadráticos con restricciones lineales. (1 y 2 sesión)
3. Caso de restricciones de desigualdad. Restricciones pasivas y activas. Condiciones de Kuhn y Tucker. Condiciones de segundo orden. Aplicación a problemas cuadráticos con restricciones lineales y a problemas convexos en convexos. Determinación portafolio de inversión óptimo. (2 y 3 sesión)
4. Algoritmo de gradiente proyectado y su aplicación a problemas cuadráticos con restricciones lineales de igualdad y desigualdad. Método de Newton. Método de Wolf para transformar un problema cuadrático con restricciones lineales en un problema de programación lineal. (3 y 4 sesión)

Universidad Autónoma Metropolitana - Iztapalapa

Casa abierta al tiempo

5° Coloquio del Departamento de Matemáticas

Introducción a la optimización no lineal con restricciones

Patricia Saavedra Barerra

Enero del 2012, Metepec, Atlixco, Puebla

Introducción a la optimización no lineal con restricciones

Patricia Saavedra Barrera [1]

2 de diciembre de 2011

[1]Departamento de Matemáticas Universidad Autónoma Metropolitana 09340, Iztapalapa, México

Índice general

1. Modelos de optimización **7**
1.1. Introducción . . . 7
1.2. Algunos modelos de optimización . . . 13
1.3. Ejercicios . . . 21

2. Optimización con restricciones **23**
2.1. Introducción . . . 23
2.2. Restricciones de igualdad . . . 26
2.3. Caso de restricciones de desigualdad . . . 37
2.3.1. Funciones convexas . . . 47
2.4. Ejercicios . . . 49

3. Método de gradiente proyectado **55**
3.1. Método de gradiente proyectado . . . 55
3.1.1. Caso de restricciones lineales de igualdad . . . 56
3.1.2. Método de Newton . . . 57
3.1.3. Algoritmo de Newton . . . 58
3.2. Caso de restricciones de desigualdad . . . 62
3.3. Método de Frank-Wolfe . . . 66
3.4. Ejercicios . . . 68

Prólogo

Esta obra está diseñada para ser una introducción a la optimización no lineal con restricciones, presentando tanto los aspectos teóricos como numéricos, a través de la modelación matemática de algunos problemas sencillos que se presentan en el mundo real como el problema del portafolio de acciones. Los ejemplos se han seleccionado no sólo con un afan ilustrativo sino para motivar al alumno al estudio de algunos temas particulares como la programación convexa.

Los antecedentes que se requieren son cálculo diferencial de varias variables, un buen curso de álgebra lineal que incluya formas cuadráticas y vectores y valores propios de matrices simétricas, y un primer curso de análisis numérico. En el primer capítulo se presentan algunos problemas de optimización y su modelación matemática; en el capítulo 2 se dan condiciones necesarias y suficientes para que un problema de optimización no lineal con restricciones de igualdad admita una solución y las correspondientes para el caso de restricciones de desigualdad, las llamadas condiciones de Kuhn-Tucker. Por último, en el capítulo 3 se presentan el método de gradiente proyectado para problemas con restricciones lineales de igualdad y desigualdad y el método de Wolfe para transformar un problema de programación cuadrática con restricciones lineales en un problema de programación lineal. Cada capítulo cuenta al final con una lista de ejercicios.

Capítulo 1

Modelos de optimización

1.1. Introducción

Los adelantos en la computación permiten actualmente a los científicos estudiar sistemas físicos, biológicos y sociales cada vez más complejos. La modelación matemática es una herramienta sencilla, sistemática y poderosa para manejar la numerosa información que se requiere para entender dichos sistemas. A partir de la segunda mitad de este siglo, se han multiplicado las ramas del conocimiento que usan la modelación matemática como parte de su metodología. Las aplicaciones de esta ciencia son numerosísimas: desde el estudio de las proteínas hasta el tránsito aéreo; desde el manejo de acciones en una casa de bolsa hasta la predicción de resultados en una elección popular.

¿Qué es un modelo?

¿Por qué los anillos de Saturno no caen sobre este planeta? Piense un momento; ahora intente reconstruir los pasos que usted siguió para responder a la pregunta. Posiblemente, lo primero que hizo fue imaginar a Saturno con sus anillos. Imaginar es una forma de ver con la mente. Después, a lo mejor, pensó que algo en común tienen la luna y la tierra y Saturno y sus anillos; por último, concluyó que la fuerza gravitacional debe jugar un papel importante en la explicación.

¿Imaginó un anillo, dos o tres? ¿Eran sólidos, con espesor, o densas nubes de polvo? Cada lector se representará a Saturno de una manera diferente. La imagen que nos venga a la mente es producto de los conocimientos que se hayan acumulado desde la primaria y de la imaginación que se tenga; de ella dependerá su explicación sobre al hecho de que los anillos no caigan sobre Saturno. Cada representación aproximada de Saturno y sus anillos es un modelo más de este sistema.

La palabra modelo será usada en este artículo en un sentido más amplio que la definición del pequeño Larousse: Objeto que se reproduce imitando a otro objeto o representación a escala de un objeto. Entenderemos aquí por modelo a una representación, por medio de un objeto, imagen, símbolo o concepto, de otro objeto o de un proceso físico, biológico, económico, etcétera. Establecer modelos forma parte del método científico que se ha usado desde el Renacimiento para generar conocimiento en Occidente y se debe entre otros a Bacon, Galileo y Descartes. A continuación presentamos a grandes rasgos y en forma esquemática por medio de la Figura 1.1 en que consiste este método. Este diagrama fue tomado de un artículo que escribió Diego Bricio Hernández, ver [5].

El primer paso es observar el fenómeno que nos interesa. Con esta información y los conocimientos previos que tengamos se propone alguna explicación o conjetura. Esta para convertirse en conclusión debe ser comprobada por medio de experimentos o probada por medio de un razonamiento lógico. Si se confirma la conjetura se sigue la flecha que dice *sí* y ésta se incorpora al resto del conocimiento que se tenga sobre el objeto de estudio. En caso negativo, se sigue la flecha *no*, y se modifica la conjetura o se revisa la validez de los conocimientos aplicados. Los conocimientos previos que haya en el tema forman el marco teórico en el que se inscribe el problema. Modelar es el vehículo que nos permite pasar de la etapa de la formulación de la conjetura al establecimiento de la conclusión. Es muy importante antes de proponer un modelo, entender bien el problema con el fin de seleccionar las variables que intervienen y las relaciones esenciales entre éstas. De esta forma se propone un modelo lo más sencillo posible, sin que la simplificación trivialice el problema. En ocasiones hay que considerar casos particulares para obtener soluciones analíticas o asintóticas que den lugar a resultados cualitativos para entender mejor en que consiste el problema original. No sólo sirve un modelo para establecer conclusiones, también es indispensable para predecir el comportamiento del sistema que se observa o para optimizar su comportamiento. Ilustremos estas ideas por medio de Saturno y sus anillos.

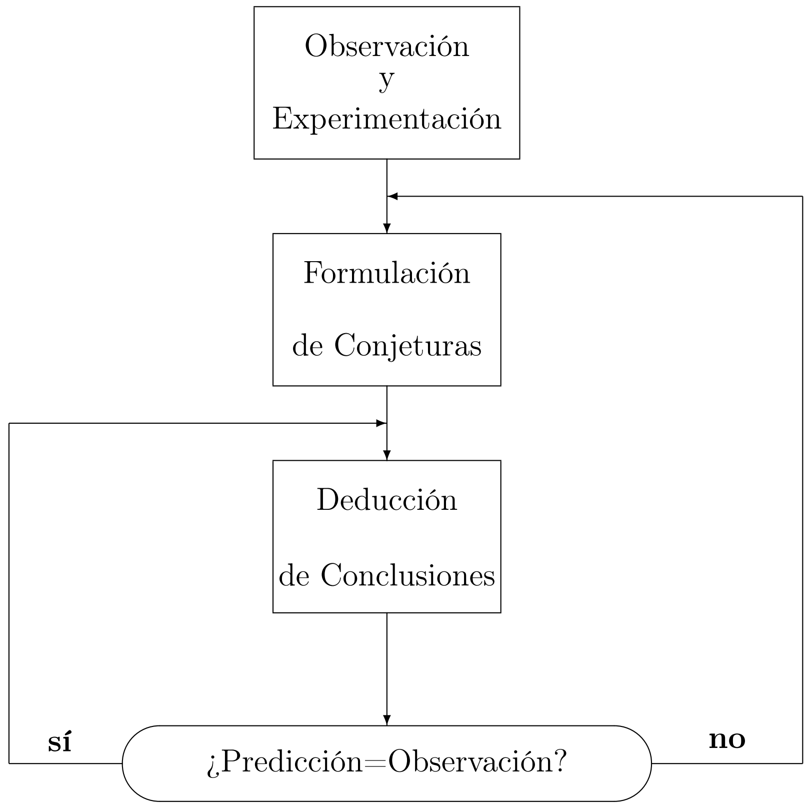

Figura 1.1. Diagrama estructural del método científico

Desde la época de Galileo se había observado que el comportamiento de Saturno era distinto al de otros planetas. En esos años, el alcance de los telescopios era demasiado corto para distinguir con nitidez a los anillos por lo que las observaciones dejaban mucho que desear. Con base en sus observaciones, Galileo concluyó que la posición de Saturno estaba ocupada por tres planetas: el mayor colocado en medio con dos apéndices pequeños a sus lados. Durante los siguientes 50 años, los astrónomos no encontraron una explicación adecuada de lo que pasaba; ¡hasta llegaron a concebir a Saturno como una taza con dos asas! La búsqueda de una explicación plausible se veía obstaculizada por el hecho de que la visibilidad de Saturno y sus anillos depende de la posición que tenga la órbita de la tierra respecto a la de este planeta. En ocasiones los anillos no son visibles mientras que en otras se ven totalmente abiertos. En 1655, las leyes de Kepler y el incremento en la calidad de las observaciones permitió a Christian Huygens concluir la existencia de un anillo delgado y plano que sin tocar a Saturno, lo rodeaba. En 1675, las observaciones de Cassini lo obligaron a rechazar esta idea y proponer otro modelo que consistía de dos anillos: uno externo y poco brillante y otro interior muy brillante, divididos ambos por una línea obscura. Hasta 1850, usando el telescopio del observatorio de Harvard, Bond descubrió que los anillos eran tres y no dos.

La explicación de la naturaleza de los anillos y su comportamiento fue siempre a la par con la representación de Saturno. Fue el descubrimiento del tercer anillo diáfano, semitransparente y polvoriento que sugirió a Maxwell en 1857 que los anillos consistían de miles de partículas orbitando alrededor de Saturno. Esta idea es muy cercana al modelo actual y ha sido corroborada por los datos que últimamente han enviado las sondas norteamericanas.

¿Qué tan bueno es un modelo? Su bondad depende de qué tan bien cumpla con los objetivos que se buscaban al plantearlo. Por ejemplo, si proponemos un modelo que considere a los anillos como sólidos, tendremos problemas; pues Laplace demostró, en 1785, que en ese caso los anillos caerían irremediablemente sobre Saturno, por lo que nuestro modelo no describe bien el comportamiento de este planeta. El modelo y las conclusiones que respecto a él se infieran, están estrechamente ligadas. Por ello establecer modelos es un proceso dinámico; se les modifica a medida que se tienen mejores observaciones.

Distintas clases de modelos

¿Qué clase de modelos podemos tener? Muy diversos: los hay análogos que simplemente imitan al objeto de estudio modificando su escala como la maqueta de una casa o del sistema solar; hay modelos diagrámaticos que a través de una imagen, un dibujo o un diagrama describen al objeto de estudio, como la Figura 1 de este artículo, y modelos conceptuales, que recurren a ideas para representar, como los modelos matemáticos. Varias clases de modelos pueden intervenir en la generación de un conocimiento.

Intentar definir lo que es un modelo matemático es una empresa difícil. Además de las conocidas trampas de lenguaje a las que se enfrenta uno, siempre se corre el riesgo de ser poco preciso y muy ambicioso. Una propuesta sería la siguiente: un modelo matemático es una representación abstracta expresada en lenguaje matemático de un proceso, fenómeno o sistema físico, biológico, económico, social, etcétera. ¿Cómo se plantea un modelo matemático? Ilustrémoslo obteniendo la trayectoria que sigue una bala al ser disparada por un cañón.

Supongamos que un cañón forma un ángulo de 30° respecto al suelo y que una bala con una masa igual a uno es lanzada, en el tiempo $t = 0$, desde el origen con una rapidez que denotaremos como v_0 de 1000 m/seg. Haremos algunas suposiciones antes de establecer el modelo con objeto de simplificar el planteamiento del mismo: asumamos que es un día claro y sin viento, lo que nos permite suponer que la bala se moverá en un plano y supongamos también que la fricción del aire no es significativa.

Nos interesa determinar el tipo de curva que describe la trayectoria del misil a lo largo de todo tiempo t que dure su movimiento, por lo que las incógnitas del problema son los puntos del plano $(x(t), y(t))$. Debemos encontrar una relación que nos permita ligar la información que tenemos como el ángulo de tiro y la rapidez inicial, que son los datos del problema, con $x(t)$ y $y(t)$. Por medio del ángulo de tiro y de la rapidez inicial podemos obtener para el tiempo $t = 0$, una velocidad en la dirección horizontal y una velocidad en la dirección vertical que denotaremos como $v_x(0)$ y $v_y(0)$, respectivamente. Esto se hace usando las siguientes expresiones que se obtienen con trigonometría

$$v_x(0) = v_0 \cos 30^\circ = 866.025 \qquad \text{y} \qquad v_y(0) = v_0 \operatorname{sen} 30^\circ = 500.$$

Para establecer el modelo matemático apliquemos la física que aprendimos en la preparatoria: por hipótesis la fuerza gravitacional es la única fuerza que afecta a la velocidad inicial; esta fuerza también se puede descomponer en una componente horizontal y otra vertical. La horizontal es cero mientras que la vertical es de -9.8 porque empuja a la bala hacia el suelo. Por lo tanto, la velocidad horizontal es la misma a lo largo del movimiento de la bala, así que la distancia recorrida en la dirección x al tiempo t es

$$x(t) = v_x(0)\, t = 866.025\, t. \tag{1.1}$$

En el caso del movimiento vertical, ésta se ve afectada por la componente vertical de la fuerza gravitacional, por lo que $v_y(t) = v_y(0) - 9.8\, t$ y

$$y(t) = 500t - \frac{9.8}{2}t^2. \tag{1.2}$$

De esta forma hemos determinado a $x(t)$ y $y(t)$ pero, ¿qué trayectoria sigue la bala? Para ello, despejemos de (1.1) la variable t, $t = x/866.025$ y substituyamos en la ecuación (1.2)

$$y(x) = \frac{500}{866.025}x - 4.9(\frac{x}{866.025})^2. \tag{1.3}$$

Esta es la ecuación de una parábola con vértice en $(44187.203, 12755.74)$. Para determinar el alcance del cañón, se calcula la abscisa x para la cual la altura es cero, o sea $y = 0$; igualando (1.3) a cero se tiene que

$$\frac{x}{866.025}(500 - \frac{4.9}{866.025}x) = 0.$$

La altura es cero en la posición inicial y cuando $x = 88,367.34$ m. Observemos que este mismo análisis se puede hacer para cualquier velocidad inicial y cualquier ángulo de tiro. Las expresiones (1.1) y (1.2) sintetizan el modelo mátematico que describe la trayectoria de una bala.

Distintos tipos de modelos matemáticos

A pesar de que cualquier intento de clasificación tiene el inconveniente de ser esquemático y reduccionista, con objeto de que la presentación de lo que es un modelo matemático sea lo más sencilla posible, adoptaremos la clasificación que sugiere Mark Meerschaert en su libro Mathematical Modeeling, véase [7]. Según él, la gran mayoría de los modelos matemáticos pertenecen a una de las siguientes categorías: modelos de optimización, modelos dinámicos y modelos probabilísticos. Un modelo dinámico es aquel que depende del tiempo, como el ejemplo anterior; el probabilístico es aquel en el que hay incertidumbre y, por último, un modelo de optimización consiste en determinar el valor óptimo de un grupo de variables. La realidad es sumamente compleja por lo que al tratar de modelarla se requiere combinar distintos tipos de modelos a la vez.

Problemas que involucren el determinar el óptimo de una función aparecen muy frecuentemente cuando se hace un modelo matemático. No importa qué tipo de problema se esté estudiando, siempre se desea maximizar los beneficios y minimizar los riesgos: empresarios tratan de controlar las variables con el fin de maximizar sus ganancias y de reducir los costos. Las personas que trabajan en la explotación de los recursos renovables como pesquerías o bosques tratan de encontrar un equilibrio entre obtener la máxima ganancia y la conservación de recursos. Los bioquímicos buscan reducir los efectos colaterales de nuevos medicamentos. Todos estos problemas tienen en común que se busca controlar ciertas variables para obtener el mejor resultado.

1.2. Algunos modelos de optimización

Los modelos de optimización buscan determinar el valor de las variables independientes, sujetas éstas en muchos casos a restricciones, que maximizan o minimizan el valor de una función. A continuación se presentan varios modelos de optimización.

Optimización lineal con restricciones

Una compañía empacadora de fruta busca maximizar la ganancia que obtiene de la venta de latas de piña, mango y guayaba. Supongamos, para simplificar el problema, que la compañía vende todo lo que produce por lo que busca optimizar su producción en lo que se refiere a la utilización de la maquinaría. Cuenta con tres máquinas: la máquina A que limpia la fruta, la máquina B que cuece la fruta y la máquina C que la enlata. Las máquinas no pueden trabajar 24 horas en forma continua, cada día varias horas deben consagrarse a su mantenimiento. Supongamos que la máquina A trabaja 8 horas al día, la B 10 horas y la C 12 horas. Para producir un lote de mango, que consiste de 100 latas, de se requiere tres horas de la máquina A, 3 horas de la máquina B y 4 horas de la máquina C; para producir un lote de piña se requiere 4 horas de la A, 2 horas de la B y 4 horas de la C y, por último, para un lote de guayaba se requiere 2 horas de la A, 2.5 horas de la B y 4 horas de la C. El costo de un lote de mango es de \$1000.00, de piña \$900.00 y de guayaba \$850.00. ¿Cuántos lotes de cada una de las frutas deben producirse para obtener el máximo de ganancia si los lotes se venden al doble del costo?

Para construir un modelo matemático observemos primero que las incógnitas de nuestro problema son el número de lotes de cada fruta que deben producirse. Como la unidad de producción es el lote denotemos con x el número de lotes de mango, con y el de piña y con z el de guayaba. A continuación notemos que la ganancia, que denotaremos con la letra G, depende de la venta total y ésta está dada por la suma de las ventas de cada fruta que, a su vez, se calcula, multiplicando el número de lotes por el precio de venta. Asi que G depende de x, y y z de la siguiente forma:

$$G(x, y, z) = 2000x + 1800y + 1700z.$$

La solución de nuestro problema es un punto (x, y, z) en el espacio $\Re^3$. G es una función que va de $\Re^3 \to \Re$. Pero la solución que buscamos no es cualquier punto de $\Re^3$ ya que las variables x, y y z deben satisfacer ciertas condiciones. Por ejemplo, el número de lotes debe ser positivo, no tiene sentido obtener valores negativos y esta condición se expresa matemáticamente por

$$x, y, z \geq 0.$$

Segundo, cada máquina tiene restricciones en su uso y se conoce el número de horas de cada máquina que se requieren para producir cada fruta. Por ejemplo, para la máquina A el número de horas que se utiliza al día no debe rebasar las 8 horas asi que, la suma de horas que se usa en cada fruta debe ser menor o igual a 8; por otro lado, el número de horas que se usa en cada fruta se calcula multiplicando el número de lotes por las horas que se requieren para producir cada lote, es decir 3 por x para el mango, 4 por y para la piña y 3 por z para la guayaba. Asi que

$$3x + 4y + 2z \leq 8.$$

Aplicando un razonamiento similar para las máquinas B y C se tiene $3x + 2y + 2.5z \leq 10$ y $4x + 4y + 4z \leq 12$, respectivamente.

Resumiendo, el problema que hay que maximizar es el siguiente: Determinar el máximo de una función G que denotaremos como Max G

$$\begin{array}{rcl}
\text{Max } G(x, y, z) & = & 2000x + 1800y + 1700z, \\
\text{sujeto a :} \qquad x, y, z & \geq & 0, \\
3x + 4y + 3z & \leq & 8, \\
3x + 2y + 2.5z & \leq & 10, \\
4x + 4y + 4z & \leq & 12.
\end{array}$$

Este es un problema de optimización con restricciones. Como la función G y las restricciones son funciones lineales respecto a sus variables independientes, este es un problema de optimización lineal con restricciones lineales que se conoce con en el nombre de programación lineal. El problema de programación lineal general, escrito en forma vectorial, es de la forma:

$$\begin{array}{rcl} \text{Max} \quad F(x) & = \vec{c}^{\,t}\vec{x}, & \\ \text{sujeto a} : A\vec{x} & \leq & \vec{d}, \\ \vec{x} & \geq & 0. \end{array}$$

Ajuste polinomial por mínimos cuadrados

Dados (x_i, y_i) observaciones con $i = 0, \ldots, m$ determinar el polinomio $p(x)$ de grado n que mejor aproxima a los datos en el sentido de mínimos cuadrados, es decir que satisface

$$Min \sum_{i=0}^{m} [p(x_i) - y_i]^2.$$

Un polinomio de grado n tiene la siguiente forma

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

basta con determinar los $n+1$ coeficientes para determinar el polinomio por lo que el problema anterior se reduce a determinar el vector $(a_0, a_1, \ldots, a_n)$ de $\Re^{n+1}$ tal que

$$Min \sum_{i=0}^{m} [a_n x_i^n + a_{n-1} x_i^{n-1} + \cdots + a_1 x_i + a_0 - y_i]^2.$$

Este es un problema de minimización cuadrática sin restricciones.

Optimización de portafolios

Determinar la composición de un portafolio de inversión, integrado por acciones de empresas que se negocian en la Bolsa Mexicana de Valores (BMV), cuyo riesgo sea el menor posible y que obtenga un rendimiento más alto que una inversión a plazo fijo.

Al tiempo $t = 0$ se tiene un monto M que se desea invertir a una semana en un portafolio de inversión, integrado con acciones de n empresas. Se tiene como datos los precios diarios de cada una de las acciones en los tres meses previos a $t = 0$. El número de acciones de cada empresa se debe determinar de tal forma que el riesgo del portafolio sea mínimo y su rendimiento semanal sea igual o mayor a una r^* dada.

Para tener una mejor idea del problema, revisemos algunos conceptos de finanzas. Un monto M^0 que se invierte en el banco a un interés r anual, al término de un año se convierte en un monto M^1 igual a

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

Observemos que $r = \frac{M^1 - M^0}{M^0}$ es la ganancia relativa, se le conoce como el rendimiento de la inversión, y en el caso de los depósitos a plazo fijo coincide con la tasa de interés.

En el caso de las acciones, como de otros activos financieros, el rendimiento durante un periodo, se define por las variaciones relativas del precio del activo y está dado por

$$r = \frac{P^1 - P^0}{P^0}, \tag{1.4}$$

con P^0 el precio al tiempo inicial y P^1 al tiempo final. Observemos que $P^1 = (1 + r)P^0$, por lo que el concepto de rendimiento coincide con el que definimos para depósitos bancarios.

Los rendimientos de un depósito bancario son deterministas porque al depositar el dinero sabemos de antemano el rendimiento exacto que se recibirá a la fecha de vencimiento; en el caso de las acciones, las variaciones del precio dependen de muchos factores: del desempeño de la empresa, de la situación económica del país, del tipo de cambio, de las tasas de interés e inclusive de qué tan optimistas o pesimistas sean los participantes en el mercado accionario. En suma, son tantos los factores que intervienen, que es difícil prever de antemano si se incrementarán o se reducirán y, más difícil aún, en cuánto lo harán. Dado que no podemos determinar con certeza el rendimiento a futuro de cada acción, ésta se comporta como una variable aleatoria. En consecuencia, al tiempo $t = 0$, a lo más a lo que podemos aspirar es a calcular el valor esperado del rendimiento de una acción.

Una forma de calcular el valor esperado de una variable aleatoria es a través del cálculo del primer momento de la distribución. ¿Qué tipo de distribución tienen los rendimientos de los activos con riesgo? Para tener una idea analicemos el comportamiento histórico de éstos; por ejemplo, a través de un histograma de los rendimientos diarios de cada acción.

Supongamos que los rendimientos son normales entonces basta con determinar su esperanza y su varianza para determinar su distribución. Cuando no se conoce esta información, se puede estimar a través de la media y varianza muestral. El rendimiento diario esperado $E(r_i)$ se puede estimar por medio de los datos a través de la media muestral

$$E(r_i) \approx \overline{r}_i = \frac{1}{M}\sum_{j=1}^{M}\frac{P_i^{j+1}-P_i^j}{P_i^j} \approx \frac{1}{M}\sum_{j=1}^{M}\ln\left(\frac{P_i^{j+1}}{P_i^j}\right).$$

La varianza σ_i^2 mide qué tanto se alejan los rendimientos reales del valor promedio, por lo que es una forma adecuada de evaluar el riesgo de una acción. La varianza muestral $\overline{\sigma}_i^2$ es un buen estimador de la varianza y se calcula por

$$\overline{\sigma}_i^2 = \frac{1}{M-1}\sum_{j=1}^{M}\left[\ln\left(\frac{P_i^{j+1}}{P_i^j}\right)-\overline{r}_i\right]^2.$$

Es importante también determinar la dependencia entre los rendimientos de las acciones. La covarianza mide esta dependencia. Se estima la covarianza a través de la covarianza muestral $\overline{Cov}(r_i, r_j)$ que se calcula por

$$\overline{Cov}(r_i, r_j) = \frac{1}{M-1}\sum_{k=1}^{M}\left(\ln(\frac{P_i^{k+1}}{P_i^k})-\overline{r}_i\right)\left(\ln(\frac{P_j^{k+1}}{P_j^k})-\overline{r}_j\right).$$

Formulación matemática del problema

El rendimiento relativo de un activo A_i se denotará por r_i y se define por la expresión (1.4). Si el precio al tiempo final P_i^1 es una variable aleatoria, también lo es r_i. Sea m_i el número de acciones que se compran del activo i. Entonces

$$\begin{aligned} M &= m_1P_1^0 + \ldots m_nP_n^0, \\ 1 &= \frac{m_1P_1^0}{M} + \cdots + \frac{m_nP_n^0}{M}. \end{aligned}$$

Sean $w_i = \frac{m_iP_i^0}{M}$ la variable que representa el porcentaje del capital M invertido en el activo A_i. Las variables w_i son las variables del problema de optimización. La ventaja de definir a las variables como w_i es que éstas no dependen del monto a invertir, por lo que podemos plantear el problema para cualquier monto M.

Las restricciones que deben satisfacer las w_i son las siguientes:

1. Para que se cumpla el requisito de que el costo del portafolio sea igual a M se debe satisfacer que

$$\sum_{i=1}^{n} w_i = 1.$$

2. La segunda restricción es que el rendimiento del portafolio sea mayor al de un depósito a plazo fijo, supongamos que el rendimiento de éste es igual a r^*.

   Para formular esta restricción en términos de las w_i, se hace lo siguiente: denotemos por V^0 el valor del portafolio al tiempo cero, V^1 el valor del portafolio al tiempo t_1 y como r_p al rendimiento del portafolio al tiempo $t = 1$. El rendimiento del portafolio es igual a

$$r_p = \frac{V^1 - V^0}{V^0},$$

   como $V^0 = M$ entonces

$$\begin{aligned} r_p &= \frac{1}{M}\sum_{i=1}^{n} m_i[P_i^1 - P_i^0] = \sum_{i=1}^{n} \frac{m_i P_i^0}{M}\frac{[P_i^1 - P_i^0]}{P_i^0}, \\ &= \sum_{i=1}^{n} w_i r_i. \end{aligned}$$

   La segunda restricción se formula matemáticamente de la siguiente forma:

$$E(r_p) = \sum_{i=1}^{n} w_i E(r_i) \approx \sum_{i=1}^{n} w_i \overline{r_i} = r^*.$$

La función a minimizar se llama la función objetivo. La función objetivo es el riesgo del portafolio. El riesgo de un portafolio puede medirse de muchas formas. En el caso que se suponga que los rendimientos son normales, la varianza del portafolio es una buena medida de su riesgo ya que cualquier otra medida de riesgo depende de la varianza, por ejemplo el *VaR*. La varianza de un portafolio se calcula de la forma siguiente:

$$\begin{aligned}
\sigma_p^2 &= E[(r_p - E(r_p))^2],\\
&= E[(\sum_{i=1}^{n} w_i r_i - E(r_p))^2],\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} w_i\ w_j\ E([r_i - E(r_i)]\ [r_j - E(r_j)]),\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(r_i, r_j) w_i w_j \approx \sum_{i=1}^{n}\sum_{j=1}^{n} \overline{Cov}(r_i, r_j) w_i w_j.
\end{aligned}$$

En suma la formulación matemática del problema del portafolio óptimo es

$$\begin{array}{ll}
 & Min\ \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\overline{Cov}(r_i, r_j)\ w_i\ w_j\\
\text{sujeto a} & \sum_{i=1}^{n} w_i\overline{r_i} = r^*,\\
 & \sum_{i=1}^{n} w_i = 1.
\end{array}$$

La función objetivo se divide por un medio por comodidad. Si se denota como $[\Sigma]$ la matriz con componentes $[\Sigma]_{ij} = \overline{Cov}(r_i, r_j)$, a $\vec{w}$ como el vector con componentes w_i, $\overrightarrow{r}$ el vector con componentes $\overline{r}_i$, y $\overrightarrow{1}$ al vector con todos sus componentes igual a uno, la forma vectorial del problema anterior es

$$\begin{array}{ll}
 & Min\ \frac{1}{2}w^t\ [\Sigma]\ w\\
\text{sujeto a} & w^t\overrightarrow{r} = r^*,\\
 & \overrightarrow{1}^t w = 1.
\end{array}$$

¿Qué sucede si no se permiten ventas en corto? Es decir que no se pueda pedir prestado dinero para integrar el portafolio. En este caso el problema es:

$$\begin{array}{ll}
 & Min\ \frac{1}{2}w^t\ [\Sigma]\ w\\
\text{sujeto a} & w^t\overrightarrow{r} = r^*,\\
 & \overrightarrow{1}^t w = 1,\\
 & w_i \geq 0, \quad i = 1, \ldots, n.
\end{array}$$

La formulación matemática del problema fue idea de Harry Markowitz, ganador del Premio Nobel de Economía en 1990 por su teoría de riesgo-rendimiento, entre otras cosas.

Optimización no lineal

Otro ejemplo de optimización es el siguiente: Se requiere enviar un paquete rectángular por correo. Por estipulaciones del servicio postal sólo se aceptan paquetes con dimensiones menores o iguales a 60 cm y se pide, además, que la superficie total sea a lo más de 80 cm^2. Si se desea maximizar el volumen, ¿qué dimensiones debe tener la caja?

Claramente las incógnitas del problema son las dimensiones, denotemos con la letra x al largo de la caja, con y al ancho y, por último, con z al espesor. Como se desea maximizar el volumen, denotemos con V al volumen que depende de las dimensiones de la caja de la forma siguiente $V(x, y, z) = xyz$.

V es una función de $\Re^3 \to \Re$; como en el caso anterior las dimensiones no pueden tomar cualquier valor. Por un lado, deben ser positivas y menores a 60 cm, ésto se expresa en lenguaje matemático de la forma $0 \leq x, y, z \leq 60$ y por otro lado, la superficie total no puede rebasar los 80 cm^2, o sea, $2(xy + xz + zy) \leq 80$. En suma, el problema a optimizar es el siguiente

$$\begin{array}{rcl} \text{Max} \quad V(x,y,z) & = & xyz, \\ \text{sujeto a :} \quad 0 \leq x, y, z & \leq & 60, \\ S(x,y,z) = 2(xy + yz + xz) - 80 & \leq & 0. \end{array}$$

El problema anterior es un problema de optimización no lineal con restricciones no lineales ya que tanto la superficie total como el volumen dependen en forma no lineal de las dimensiones. En forma general el problema de optimización no lineal con restricciones no lineales es de la forma:

$$\begin{array}{rcl} \text{Max } F(\vec{x}), & & \\ \text{sujeto a : } \vec{h}(\vec{x}) & \leq & 0. \end{array}$$

Como se puede observar dependiendo de las características de las restricciones como de la función objetivo, aquella que se desea maximizar o minimizar, el problema de optimización se clasifica de muy diversas maneras: si tanto la función objetivo como las restricciones son convexas se dice que se tiene un problema de programación convexa; si la función objetivo es lineal pero con dominio en los enteros se conoce con el nombre de programación entera y las técnicas que se utilizan son principalmente de combinatoria. Si la función objetivo es cuadrática se dice que el problema es de programación cuadrática, si la función es no-lineal se le llama programación no-lineal.

1.3. Ejercicios

Plantee los siguientes problemas como problemas de optimización.

1. Demuestre que de todos los rectángulos con un perímetro fijo, el cuadrado tiene máxima área y que de todos los rectángulos con área fija, el cuadrado tiene mínimo perímetro.

2. Dada una linea recta L y dos puntos A y B del mismo lado de L, encuentre el punto P sobre L que hace que la suma de las distancias AP y PB sea mínima.

3. Una lata cerrada de forma cilíndrica debe tener un volumen fijo. ¿Qué dimensiones debe tener la lata para que la superficie total sea mínima?

4. Dos caminos se intersectan en ángulo recto. Un carro A está situado en la posición P sobre uno de los caminos a S kilómetros de la intersección. Sobre el otro camino se encuentra el auto B, en la posición Q a s kilómetros de la intersección. Ellos comienza a viajar hacia la intersección al mismo tiempo, el primero con velocidad R y el segundo con velocidad r. ¿Después de qué tiempo de que comenzaron a rodar, la distancia entre los dos será mínima?

5. Una compañía aérea de transportación tiene la capacidad de mover 100,000 toneladas al día. La compañía cobra 250 dólares por tonelada. El número de toneladas que puede transportar esta limitada por la capacidad del avión que es de $50,000\ m^3$. La compañía mueve su carga a través de contenedores de distinto tamaño. La siguiente tabla muestra el peso y el volumen que cada contenedor puede llevar:

Tabla 1

Tipo de Contenedor	Peso (ton)	Volumen (m^3)
1	30	550
2	40	800
3	50	400

Determine cuántos contenedores de cada tipo deben transportarse al día para maximizar las ganancias.

6. Un productor de computadoras personales vende en promedio 10,000 unidades al mes de su modelo M1. El costo de producción de cada computadora es de 700 dólares y el precio de venta es de 1150 dólares. El administrador decidió reducir en un 10 % el precio de cada computadora y el efecto fue de un incremento del 25 % en las ventas. Por otro lado, la compañía tiene un contrato de publicidad a nivel nacional que le cuesta 50,000 dólares al mes. La agencia de publicidad afirma que si incrementan la publicidad mensual en 10,000 dólares, venderán más de 200 unidades al mes. Dado que el administrador no desea gastar más de 100,000 dólares al mes en publicidad, determine el precio en que se deben vender las computadoras y el gasto de publicidad mensual que maximizan las ganancias, si se supone que hay una relación lineal entre la disminución del precio y el incremento en las ventas.

7. Un productor de televisores desea introducir al mercado dos nuevos modelos: un aparato a colores, con una pantalla de 19 pulgadas y con sonido estereofónico que lo identificaremos como el modelo A y otro que le llamaremos el modelo B que tiene las mismas características que el anterior pero, con una pantalla de 21 pulg. El modelo A se venderá al público en \$10700 pesos, mientras que el modelo B tendrá un costo de \$13500 pesos. Producir un televisor tipo A cuesta \$5850 pesos y del tipo B \$6750 pesos. Además, al costo total de producción se le debe sumar \$400,000 de gastos fijos. La venta promedio de los televisores se reduce cada vez que se compra un televisor del mismo modelo y ésto se expresa reduciendo el precio original en un centavo por modelo vendido. Asimismo, las ventas del modelo A influyen en las ventas del modelo B y viceversa. Se estima que cada vez que se compra un televisor tipo A se reduce el precio del modelo B en 4 pesos y cada vez que se vende un modelo B se reduce el precio del modelo A en 3 pesos. ¿Cuántas unidades de cada modelo deben producirse para maximizar la ganancia?

Capítulo 2

Optimización con restricciones

2.1. Introducción

Un gran número de modelos de optimización imponen a las variables una serie de restricciones que se traducen en el que mínimo no se busca en todo el espacio sino en un subconjunto del espacio definido por las restricciones. Por ejemplo consideremos los siguientes problemas:

1. Una sonda espacial en forma esférica entra a la atmósfera de la tierra y su superficie comienza a calentarse. Supongamos que la ecuación de la esfera está dada por

   $$x^2 + y^2 + z^2 = 4$$

   y que después de diez minutos, la temperatura sobre la superficie de la sonda es

   $$T(x, y, z) = xz + y^2 + 600.$$

   Determínese el punto más caliente sobre la superficie.

   En este caso el problema se plantea de la siguiente forma: sea Ω un subconjunto de $\Re^3$ definido por

   $$\Omega = \{(x, y, z) \mid h(x, y, z) = x^2 + y^2 + z^2 - 4 = 0\},$$

   determinar

   $$\underset{x \in \Omega}{Max} \quad T(x, y, z).$$

Este es un problema cuadrático con una restricción cuadrática dada por una igualdad.

2. El problema del portafolio expuesto en el capítulo 1, sección 1.2, consiste en determinar el portafolio con ventas en corto con mínima varianza y cuyo valor esperado es mayor o igual a una r^* dada. La formulación matemática del problema es

$$\begin{array}{ll} & Min\ \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\overline{Cov}(r_i,r_j)\ w_i\ w_j \\ \text{sujeto a} & \sum_{i=1}^{n} w_i\overline{r_i} = r^*, \\ & \sum_{i=1}^{n} w_i = 1. \end{array}$$

Este es un problema cuadrático con restricciones lineales.

La principal dificultad de los problemas de minimización con restricciones reside en que no se tiene una caracterización de un punto mínimo que dependa únicamente de la función objetivo, también se requiere que se satisfagan ciertas condiciones respecto a las restricciones. A continuación presentaremos algunas definiciones que nos serán útiles en el manejo de las restricciones.

Puntos admisibles y regulares

Sea F una función de $\Re^n$ a los reales y sea Ω un conjunto distinto del vacío de $\Re^n$ definido por

$$\Omega = \{\vec{x} \in \Re^n \mid h_j(\vec{x}) = 0 \text{ para } j = 1, \dots m\},$$

donde h_j puede ser una función lineal o no lineal. Un problema de restricciones de igualdad es de la forma

$$\underset{\vec{x}\in\Omega}{Min} \quad F(\vec{x}).$$

En el caso lineal las restricciones son de la forma

$$h_j(\vec{x}) = \vec{c}_j\ ^t\vec{x} - e_j = 0.$$

Un problema de minimización con restricciones de desigualdad es un problema de minimización en el que Ω se define como

$$\Omega = \{\vec{y} \in \Re^n |\ h_j(\vec{y}) = 0,\ j = 1, \dots, m,\ g_j(\vec{y}) \leq 0,\ j = 1, \dots, s\}.$$

Definición 2.1.1. *Diremos que $\vec{x} \in \Re^n$ es un punto admisible de un problema de minimización con restricciones si $\vec{x} \in \Omega$.*

¿Qué se entiende por el mínimo de f restringido a un conjunto Ω?

Definición 2.1.2. *Un punto $\vec{x}^*$ se dice que es un mínimo de F restringido a un subconjunto $\Omega \neq \emptyset$ de $\Re^n$ si*

$$F(\vec{x}^*) \leq F(\vec{x}) \quad \forall\ \vec{x} \in \Omega.$$

Definición 2.1.3. *Sea ${}^t\vec{h} = (h_1(\vec{x}), \ldots, h_m(\vec{x}))$ una función vectorial continuamente diferenciable, definamos como la matriz jacobiana del vector a la matriz $m \times n$ con componentes*

$$J_h(\vec{x}) = \begin{pmatrix} \frac{\partial h_1(\vec{x})}{\partial x_1} & \cdots & \frac{\partial h_1(\vec{x})}{\partial x_n} \\ \frac{\partial h_2(\vec{x})}{\partial x_1} & \cdots & \frac{\partial h_2(\vec{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m(\vec{x})}{\partial x_1} & \cdots & \frac{\partial h_m(\vec{x})}{\partial x_n} \end{pmatrix}.$$

La matriz jacobiana de $\vec{h}$ tiene como renglón i al gradiente de h_i. Esta matriz nos define para cada $\vec{x}$ una transformación lineal de $\Re^n$ a $\Re^m$. Por ejemplo en el caso de la esfera, dado que sólo tenemos una restricción, J_h es una matriz de 1×3 definida por

$$J_h(x) = (2x, 2y, 2z).$$

En el caso que Ω sea el conjunto

$$\Omega = \{\vec{x} \in \Re^n \mid h_j(\vec{x}) = \vec{c}_j\ {}^t\vec{x} - e_j = 0 \text{ para } j = 1, \ldots m\}$$

entonces $J_h = C^t$ con C^t la matriz de $m \times n$ que tiene como $j-$ésimo renglón al vector $\vec{c}_j^{\,t}$.

Denotemos como $N(\vec{x})$ al espacio nulo de la transformación $J_h(\vec{x})$, es decir

$$N(\vec{x}) = \{\vec{y} \in \Re^n \mid J_h(\vec{x})\ \vec{y} = 0\}.$$

$N(\vec{x})$ es el espacio ortogonal al espacio generado por los vectores

$$\{\nabla h_1(\vec{x}), \ldots, \nabla h_m(\vec{x})\}.$$

Así en el ejemplo 1, $N(\vec{x})$ es el espacio

$$N(\vec{x}) = \{(a,b,c) \in \Re^3 |\ 2xa + 2yb + 2zc = 0\}.$$

En particular $N(\vec{0}) = \Re^3$ pues $\nabla h_1(\vec{0}) = \vec{0}$ y en cambio

$$N(1,1,1) = \{(a,b,c) \in \Re^3 | 2a + 2b + 2c = 0\}.$$

La dimensión de $N(1,1,1)$ es 2. En el caso de tener restricciones lineales $N(\vec{x})$ es igual

$$N(\vec{x}) = \{\vec{y} \in \Re^n |\ C^t\vec{y} = 0\} = EN(C^t).$$

Definición 2.1.4. *Diremos que $\vec{x}^*$ es un punto regular de Ω si el conjunto de vectores*

$$\{\nabla h_1(\vec{x}^*), \ldots, \nabla h_m(\vec{x}^*)\}$$

es linealmente independiente.

Obsérvese que si $m \leq n$ es posible que el gradiente de todas las restricciones sean linealmente independientes, pero si $m > n$ no es posible que haya un punto regular admisible. En el caso lineal o se tiene que todos los puntos son regulares o ninguno lo es, pues serán regulares si la matriz C es de rango completo o sea de rango igual a m.

2.2. Restricciones de igualdad

Sea F una función de $\Re^n$ a los reales y sea Ω el conjunto distinto del vacío de $\Re^n$ definido por

$$\Omega = \{\vec{x} \in \Re^n \mid h_j(\vec{x}) = 0 \text{ para } j = 1, \ldots m\},$$

donde h_j puede ser una función lineal o no lineal. Un problema de minimización con restricciones de igualdad no lineal (P) es de la forma

$$\underset{\vec{x}\in\Omega}{Min} \quad F(\vec{x}).$$

Denotemos como $T(\vec{x}^*)$ el plano tangente a la superficie Ω en el punto $\vec{x}^*$. Recordemos que el plano tangente está formado por todos los vectores $\vec{y} \in \Re^n$ que son rectas tangente a una curva que pasa por $\vec{x}^*$ y que está sobre Ω.

Teorema 2.2.1. *Sea $\vec{h}$ una función continuamente diferenciable en un abierto que contenga a Ω. Si $\vec{x}^*$ es un punto regular admisible de Ω entonces*

$$N(\vec{x}^*) = T(\vec{x}^*).$$

Dem: Probemos primero que $T(\vec{x}^*) \subset N(\vec{x}^*)$. Sea $\vec{y} \in T(\vec{x}^*) \Rightarrow$ existe una curva $\vec{x}(t)$ de $\Re$ a $\Re^n$ que pasa por $\vec{x}^*$. Supongamos que $\vec{x}(0) = \vec{x}^*$ y que $\vec{x}\,'(0) = \vec{y}$. La derivada de $\vec{h}$ respecto a t es cero pues $\vec{x}(t)$ está en Ω y

$$0 = \frac{d\vec{h}(\vec{x}(t))}{dt}|_{t=0} = J_h(\vec{x}^*)\vec{x}\,'(0) = J_h(\vec{x}^*)\vec{y}$$

lo que implica que $\vec{y} \in N(\vec{x}^*)$.

Demostremos ahora que $N(\vec{x}^*) \subset T(\vec{x}^*)$. Sea $\vec{y} \in N(\vec{x}^*)$ entonces hay que demostrar que existe una $t_0 > 0$ una curva $\vec{x}(t)$ en Ω para $t \in [0, t_0]$ tal que $\vec{x}(0) = \vec{x}^*$ y $\vec{x}\,'(0) = \vec{y}$. Para ello considérese la curva

$$\vec{x}(t) = \vec{x}^* + \vec{y}t + J_h^t(\vec{x}^*)\ \vec{u}(t) \tag{2.1}$$

con $\vec{u}(t)$ un vector en $\Re^m$. Demostrar la existencia de la curva $\vec{x}(t)$ es equivalente a demostrar que existe $t_0 > 0$ tal que para cada $t \in [0, t_0]$ existe un único vector $\vec{u}(t)$ para el cual $\vec{h}(\vec{x}(t)) = 0$.

Al evaluar la curva $\vec{x}(t)$ en cero se tiene que

$$\vec{x}(0) = \vec{x}^* + J_h^t(\vec{x}^*)\vec{u}(0).$$

Como deseamos que $\vec{x}(0) = \vec{x}^*$ impongamos la condición que $\vec{u}(0) = 0$. Para que $\vec{x}(t) \in \Omega$ se tiene que cumplir que

$$\vec{h}(\vec{x}^* + \vec{y}t + J_h^t(\vec{x}^*)\vec{u}(t)) = 0. \tag{2.2}$$

Observemos que para cada t tenemos que determinar un vector $\vec{u}(t) \in \Re^m$ que satisfaga el sistema de m ecuaciones con m incógnitas dado por (2.2) ¿Tiene solución este sistema para toda t en el intervalo $[0, t_0]$, para alguna $t_0 > 0$?

El teorema de la función implícita nos dice que esta sistema puede resolverse en forma única en una vecindad de $\vec{u}(0)$ si $\vec{h}(x(0)) = 0$ y $D_u\vec{h}$ es no singular en $t = 0$.

$$D_u\vec{h}(\vec{x}^* + \vec{y}t + J_h^t(\vec{x}^*)\vec{u}(t))|_{t=0} = J_h(\vec{x}^*)\ J_h^t(\vec{x}^*)$$

es no singular pues, por hipótesis, $\vec{x}^*$ es un punto regular por lo que el rango de esta matriz es m. Por lo tanto existe una $t_0 > 0$ y un único $\vec{u}(t)$ para $|t| \leq t_0$ tal que $\vec{x}(t) \in \Omega$. Además,

$$0 = \frac{d\vec{h}}{dt}(\vec{x}^* + \vec{y}t + J_h^t(\vec{x}^*)\vec{u}(t))|_{t=0} = J_h(\vec{x}^*)[\vec{y} + J_h^t(\vec{x}^*)\vec{u}'(0)];$$

como $\vec{y} \in N(\vec{x}^*)$ entonces

$$\frac{d\vec{h}}{dt}(\vec{x}^* + \vec{y}t + J_h^t(\vec{x}^*)\vec{u}(t))|_{t=0} = [J_h(\vec{x}^*)\ J_h^t(\vec{x}^*)]\vec{u}'(0) = 0$$

lo que implica que $\vec{u}'(0) = \vec{0}$ por lo que $\vec{x}'(0) = \vec{y}$. Así que $\vec{y} \in T(\vec{x}^*)$. □

Lemma 2.2.2. *Sea F una función continuamente diferenciable en un abierto que contenga a Ω. Sea $\vec{x}^*$ un punto regular de las restricciones $\vec{h}(\vec{x}) = 0$ y sea $\vec{x}^*$ un punto extremo de F en*

$$\Omega = \{\vec{x} \in \Re^n \mid h_j(\vec{x}) = 0 \text{ para } j = 1, \dots m\},$$

entonces para toda $\vec{y} \in N(\vec{x}^*)$ se cumple

$$\nabla F^t(\vec{x}^*)\vec{y} = 0.$$

Tomemos una $\vec{y} \in N(\vec{x}^*)$, por el lema anterior, existe una curva $\vec{x}(t)$ que satisface que $\vec{x}(0) = \vec{x}^*$ y $\vec{x}'(0) = \vec{y}$. Como $\vec{x}^*$ es un punto extremo de F sobre la curva $\vec{x}(t)$ se tiene que al evaluar la derivada de F en $\vec{x}^*$ por la regla de la cadena se obtiene que

$$0 = \frac{dF}{dt}(\vec{x}^*) = \frac{dF}{dt}(\vec{x}(t))|_{t=0} = \nabla F(\vec{x}^*)^t\vec{y}.$$

Por lo tanto $\nabla F(\vec{x}^*)$ es ortogonal al espacio tangente siempre que $\vec{x}^*$ sea un punto extremo de F que es punto regular de Ω. □

Condiciones de primer orden

Teorema 2.2.3. *Si F y $\vec{h}$ son funciones continuamente diferenciables en un abierto que contenga a Ω y $\vec{x}^*$ un punto extremo, máximo o mínimo, de F sujeto a las restricciones $\vec{h}(x) = 0$. Si $\vec{x}^*$ es un punto regular de Ω entonces existe $\vec{\lambda} \in \Re^m$ tal que*

$$\nabla F(\vec{x}^*) + J_h(\vec{x}^*)^t\vec{\lambda} = 0. \tag{2.3}$$

Por el lema anterior $\nabla F(\vec{x}^*)$ es ortogonal a todo vector en el plano tangente a la superficie Ω y por el Lema 2.2.1 es ortogonal a todo vector $\vec{y} \in N(\vec{x}^*)$. Así que $\nabla F(\vec{x}^*)$ está en el espacio generado por $\{\nabla h_1(\vec{x}^*), \ldots, \nabla h_m(\vec{x}^*)\}$ y se puede escribir como una combinación lineal de estos vectores; es decir, existe un vector $\vec{\lambda} \in \Re^m$ tal que

$$\nabla f(\vec{x}^*) = -J_h^t(\vec{x}^*)\vec{\lambda}.$$

$\square$

Observemos que, como en el caso lineal, la expresión (2.3) define un sistema de n ecuaciones con $n + m$ incógnitas que junto a las m ecuaciones $\vec{h}(\vec{x}) = 0$ da lugar a un sistema de $n + m$ ecuaciones con $n + m$ incógnitas. Al vector $\vec{\lambda}$ se le conoce con el nombre de multiplicador de Lagrange.

Ejemplo

Supongamos que se desea resolver el problema presentado en el ejemplo 1 de la sección 2.1:

$$\begin{array}{ll} Max & T(x, y, z), \\ x \in \Omega & \end{array}$$

donde $T(x, y, z) = xz + y^2 + 600$ y

$$\Omega = \{(x, y, z) \mid h(x, y, z) = x^2 + y^2 + z^2 - 4 = 0\}$$

.

Para que (x, y, z) sea el máximo de T restringido a Ω debe existir $\lambda \in \Re$ tal que

$$\nabla T(x, y, z) + \lambda \nabla h(x, y, z) = (z, 2y, x) + \lambda(2x, 2y, 2z) = 0.$$

Estas ecuaciones junto con la restricción dan lugar al siguiente sistema de ecuaciones no-lineales

$$\begin{aligned} z + 2x\lambda &= 0, \\ 2y + 2y\lambda &= 0, \\ x + 2z\lambda &= 0, \\ x^2 + y^2 + z^2 &= 4. \end{aligned}$$

Este sistema tiene cinco soluciones (x, y, z, λ): $(\sqrt{2}, 0, \sqrt{2}, -\frac{1}{2})$, $(-\sqrt{2}, 0, -\sqrt{2}, -\frac{1}{2})$, $(\sqrt{2}, 0, -\sqrt{2}, \frac{1}{2})$, $(-\sqrt{2}, 0, \sqrt{2}, \frac{1}{2})$, $(0, \pm 2, 0, -1)$. Observemos que todos los puntos son puntos regulares de Ω ya que $\nabla h(\vec{x})$ en estos puntos es distinto de $\vec{0}$.

¿En cuál de estos puntos la temperatura es mayor? La respuesta se obtiene el evaluar T en cada uno de los puntos críticos y seleccionar aquel en el que alcanza el valor más grande. Observemos que $T(0, \pm 2, 0) = 604°$ y que en estos puntos alcanza su valor máximo, mientras que en $(\sqrt{2}, 0, -\sqrt{2})$ y $(-\sqrt{2}, 0, \sqrt{2})$ alcanza su valor mínimo que es $598°$. Otro procedimiento se obtendrá más adelante con las condiciones de segundo orden.

Condiciones de segundo orden

En esta sección supondremos que F y $\vec{h}$ son funciones dos veces continuamente diferenciables en un abierto que contenga a Ω.

Teorema 2.2.4. *Supongamos que $\vec{x}^*$ es un mínimo local del problema (P) y que $\vec{x}^*$ es un punto regular de Ω entonces existe un vector $\vec{\lambda} \in \Re^m$ tal que*

$$\nabla F(\vec{x}^*) + J_h^t(\vec{x}^*)\vec{\lambda} = 0.$$

Si $N(\vec{x}^*) = \{\vec{y} \in \Re^n | J_h(\vec{x}^*)\vec{y} = 0\}$ entonces la matriz

$$L(\vec{x}^*) = H_F(\vec{x}^*) + \sum_{i=1}^{m} \lambda_i H_{h_i}(\vec{x}^*)$$

es positiva semidefinida en $N(\vec{x}^*)$, es decir $\vec{y}^t L(\vec{x}^*)\vec{y} \geq 0$ para toda $\vec{y} \in N(\vec{x}^*)$.

Dem: La primera parte se demuestra por el Teorema 2.2.3. Para demostrar la segunda parte, considérese a $\vec{x}(t)$ una curva que pasa por $\vec{x}^*$ en $t = 0$ con vector tangente $\vec{y} \in N(\vec{x}^*)$ y que satisface $\vec{x}\,'(0) = \vec{y}$ entonces F restringido a esta curva es una función de variable real. Como $\vec{x}^*$ es un mínimo local de F se cumple que

$$\frac{d^2 F(\vec{x}(t))}{dt^2}|_{t=0} \geq 0,$$

lo que implica que

$$\vec{x}\,'(0)^t H_F(\vec{x}^*)\vec{x}\,'(0) + \nabla F(\vec{x}^*)\vec{x}''(0) \geq 0. \tag{2.4}$$

Por otro lado, como la curva $\vec{x}(t)$ está sobre Ω, se tiene que para toda restricción h_i

$$\vec{\lambda}_i h_i(\vec{x}(t)) = 0.$$

Al derivar dos veces la expresión anterior y al evaluarla en $t = 0$ se tiene que

$$\vec{x}\,''^t(0)\lambda_i H_{h_i}(\vec{x}^*)\vec{x}\,'(0) + \nabla h_i(\vec{x}^*)\lambda_i \vec{x}\,''(0) = 0.$$

Sumando respecto a i

$$\vec{x}\,''^t(0)\sum_{i=1}^{m}\lambda_i H_{h_i}(\vec{x}^*)\vec{x}\,'(0) + J^t_{\vec{h}}(\vec{x}^*)\vec{\lambda}\vec{x}\,''(0) = 0. \tag{2.5}$$

Recordemos que por el teorema 2.2.3

$$\nabla F(\vec{x}^*) = -J^t_{h}(\vec{x}^*)\vec{\lambda}$$

y sumando (2.4) y (2.5) se tiene que

$$\vec{x}'^t(0)[H_F(\vec{x}^*) + \sum_{i=1}^{m} H_{h_i}(\vec{x}^*)\lambda_i]\vec{x}'(0) \geq 0$$

para toda $\vec{y} \in N(\vec{x}^*)$. Lo que implica que $L(\vec{x}^*)$ es una matriz semidefinida positiva en $N(\vec{x}^*)$. $\square$

Teorema 2.2.5. *Supóngase que hay un punto $\vec{x}^*$ en Ω y una $\vec{\lambda} \in \Re^m$ tal que*

$$\nabla F(\vec{x}^*) + J^t_{h}(\vec{x}^*)\vec{\lambda} = 0. \tag{2.6}$$

Supóngase también que la matriz

$$L(\vec{x}^*) = H_F(\vec{x}^*) + \sum_{i=1}^{m} H_{h_i}(\vec{x}^*)\lambda_i$$

es positiva definida en

$$N(\vec{x}^*) = \{\vec{y} \in \Re^n | J_{\vec{h}}(\vec{x}^*)\vec{y} = 0\}$$

$\Rightarrow$ *$\vec{x}^*$ es un mínimo estricto de F en Ω.*

Supongamos que $\vec{x}^*$ no es un mínimo estricto de F en Ω, entonces existe $\vec{y} \in Omega$ tal que $F(\vec{y}) \leq F(\vec{x}^*)$. Aún más existe una sucesión $\{\vec{y}_k\} \in \Omega$ que converge a $\vec{x}^*$ y tal que $F(\vec{y}_k) \leq F(\vec{x}^*)$. De esta última afirmación se desprende que si al menos hay un punto $\vec{y}$ en el que F alcanza un valor menor que en $\vec{x}^*$, como F es continua debe entonces tomar todos los valores entre $F(\vec{y})$ y $F(\vec{x}^*)$ en puntos $\vec{y}_k$ en Ω. Esta sucesión es de la forma

$$\vec{y}_k = \vec{x}^* + \delta_k \vec{s}_k$$

con vectores $\vec{s}_k$ en la bola unitaria de $\Re^n$ y $\delta_k > 0$. Claramente $\delta_k \to 0$ cuando k tiende a infinito y además como $\vec{s}_k$ es una sucesión acotada debe tener una subsucesión convergente a un elemento $\vec{s}^* \in \Re^n$. Además

$$\lim_{k\to\infty} \frac{\vec{h}(\vec{y}_k) - \vec{h}(\vec{x}^*)}{\delta_k} = 0$$

lo que implica que

$$J_h(\vec{x}^*)\vec{s}^* = 0,$$

por lo que $\vec{s}^* \in N(\vec{x}^*)$.

Aplicando la serie de Taylor a h_i alrededor de $\vec{x}^*$ se tiene

$$0 = h_i(\vec{y}_k) = h_i(\vec{x}^*) + \delta_k\ \nabla h_i^t(\vec{x}^*)\vec{s}_k + \frac{\delta_k^2}{2}\ \vec{s}_k^t H_{h_i}(\eta_i)\vec{s}_k.$$

Multiplicando por λ_i y sumando de $i = 1$ hasta m se tiene que

$$0 = \sum_{i=1}^{m} \lambda_i h_i(\vec{y}_k) = \sum_{i=1}^{m} \lambda_i (h_i(\vec{x}^*) + \delta_k\ \nabla h_i^t(\vec{x}^*)\vec{s}_k + \frac{\delta_k^2}{2}\ \vec{s}_k^t H_{h_i}(\eta_i)\vec{s}_k). \quad (2.7)$$

Por otro lado, se tiene que

$$F(\vec{y}_k) = F(\vec{x}^*) + \delta_k \nabla F^t(\vec{x}^*)\vec{s}_k + \frac{\delta_k^2}{2}\vec{s}_k^t H_F(\xi_k)\vec{s}_k$$

y sumando esta igualdad con (2.7) se obtiene que

$$\begin{aligned} F(\vec{y}_k) &= F(\vec{x}^*) + \delta_k[\nabla F^t(\vec{x}^*) + \sum_{i=1}^{m} \lambda_i \delta h_i(\vec{x}^*)]\vec{s}_k \\ &+ \frac{\delta_k^2}{2}\vec{s}_k^t[H_F(\xi_k) + \lambda_i H_{h_i}(\eta_i)]\vec{s}_k. \end{aligned}$$

Dado que $F(\vec{y}_k) - F(\vec{x}^*) \leq 0$ y que (2.6) se cumple entonces

$$0 \geq \frac{\delta_k^2}{2}\ \vec{s}_k^{\,t}[H_F(\xi_k) + \lambda_i H_{h_i}(\eta_i)]\vec{s}_k$$

para cada k, por lo que al pasar al límite se contradice la hipótesis que la matriz L sea definida positiva en $N(\vec{x}^*)$. $\square$

Los puntos máximos de F restringidos a un conjunto Ω pueden caracterizarse de una manera similar a los puntos mínimos. La condición de primer orden es la misma que (2.6) lo que difiere es que la matriz L debe ser una matriz negativa definida en $N(\vec{x}^*)$.

Ejemplos

1. Retomemos el ejemplo de la sonda de forma esférica con ecuación $x^2 + y^2 + z^2 - 4 = 0$ y cuya temperatura está dada por la función $T(x, y, z) = xz + y^2 + 600$. Los puntos (x, y, z): $(\sqrt{2}, 0, \sqrt{2})$, $(-\sqrt{2}, 0, -\sqrt{2})$, $(\sqrt{2}, 0, -\sqrt{2})$, $(-\sqrt{2}, 0, \sqrt{2})$ y $(0, \pm 2, 0, -1)$ son candidatos a ser puntos extremos de T en la esfera dado que satisfacen las condiciones de primer orden. ¿En cuáles de ellos alcanza el valor mínimo o máximo T? Para responder calculemos la matriz L, observemos que como T y h son cuadráticas, L únicamente depende de λ

$$L(\lambda) = \begin{pmatrix} 2\lambda & 0 & 1 \\ 0 & 2+2\lambda & 0 \\ 1 & 0 & 2\lambda \end{pmatrix}.$$

En el caso de los primeros dos puntos que tienen como multiplicador de Lagrange a $\lambda = -\frac{1}{2}$, L es de la forma

$$L(-\frac{1}{2}) = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix}.$$

El espacio tangente correspondiente está definido por

$$N(\sqrt{2}, 0, \sqrt{2}) = \{\vec{y} = (a, b, c) \mid 2\sqrt{2}a + 2\sqrt{2}c = 0\}.$$

Entonces

$$\vec{y}^{\,t} L(-\frac{1}{2})\vec{y} = b^2 - 4a^2$$

que es una matriz indefinida en $N(\sqrt{2}, 0, \sqrt{2})$ por lo que no se alcanza en este punto ni el valor máximo ni el mínimo. A la misma conclusión se llega cuando se hacen los cálculos respectivos para $(-\sqrt{2}, 0, -\sqrt{2})$.

Cuando $\lambda = \frac{1}{2}$ se tiene

$$L(1/2) = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 3 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

El plano tangente a la sonda en el punto $(\sqrt{2}, 0, -\sqrt{2})$ es de la forma

$$N(\sqrt{2}, 0, -\sqrt{2}) = \{\vec{y} = (a, b, c) | 2\sqrt{2}a - 2\sqrt{2}c = 0\}$$

por lo que $\vec{y}^t\ L(1/2)\vec{y} = 4a^2 + 3b^2 \geq 0$ y la matriz $L(1/2)$ es positiva definida en $N(\sqrt{2}, 0, -\sqrt{2})$ y alcanza en ese punto un valor mínimo. Un razonamiento similar nos permite comprobar que también $(\sqrt{2}, 0, -\sqrt{2})$ es un mínimo de T en la esfera.

Para el caso de $(0, \pm 2, 0)$ que tienen multiplicador de Lagrange a $\lambda = -1$, L es de la forma

$$L(-1) = \begin{pmatrix} -2 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -2 \end{pmatrix}.$$

$N(0, \pm 2, 0) = \{\vec{y} = (a, b, c) \in \Re^3 | b = 0\}$ y

$$\vec{y}^t L(-1)\vec{y} = -2a^2 + 2ac - 2c^2 \leq -(a^2 + c^2) \leq 0.$$

Así que la matriz $L(-1)$ es negativa definida en $N(0, -2, 0)$ y $N(0, 2, 0)$, aplicando las condiciones de segundo orden se concluye que T alcanza su valor máximo en estos puntos. Observe que estas conclusiones coinciden con las que se habían obtenido al evaluar T.

2. **Caso Cuadrático**

   Sea $F(\vec{x})$ una función cuadrática de la forma

$$F(\vec{x}) = \frac{1}{2}\vec{x}^t A\vec{x} - \vec{x}^t\vec{b}$$

y sea

$$\Omega = \{\vec{x} \in \Re^n |\ C^t\vec{x} = \vec{e}\},$$

con C una matriz de $n \times m$. Determinar bajo qué condiciones el problema

$$\begin{array}{cc} \text{Min} & F(\vec{x}) \\ x \in \Omega & \end{array}$$

admite una solución única.

Observemos que en este caso todos los puntos son regulares o ninguno lo es porque $J_h = C^t$ y sólo si el rango de C es completo el punto será regular. Supongamos que este es el caso, entonces un candidato a ser punto extremo de F en Ω debe satisfacer que

$$\begin{aligned} A\vec{x} + C\vec{\lambda} &= \vec{b}, & (2.8) \\ C^t\vec{x} &= \vec{e}. & (2.9) \end{aligned}$$

Lemma 2.2.6. *El sistema (2.8) admite una solución única si A es una matriz positiva definida y C es una matriz de rango completo.*

Para demostrar que la matriz

$$\begin{pmatrix} A & C \\ C^t & 0 \end{pmatrix}$$

es una matriz no singular basta con demostrar que el sistema homogéneo asociado a (2.8) tiene como única solución a $\vec{x} = \vec{0}$ y $\vec{\lambda} = \vec{0}$.

La primera ecuación del sistema anterior nos dice que

$$A\vec{x} + C\vec{\lambda} = \vec{0}.$$

Multiplicando por $\vec{x}^t$ y usando que $C^t\vec{x} = \vec{0}$ se tiene que

$$x^tA\vec{x} + x^tC\vec{\lambda} = x^tA\vec{x} = 0$$

como A es una matriz positiva definida, sólo se cumple la igualdad a cero si $\vec{x} = \vec{0}$. Por otro lado,

$$C\vec{\lambda} = 0$$

por lo que al mutiplicar por C^t se tiene que C^tC es una matriz no singular dado que C es una matriz de rango completo, por lo que la única solución es $\vec{\lambda} = \vec{0}$. $\square$

La solución del sistema (2.8) se puede reducir a resolver el siguiente sistema de ecuaciones :

$$\begin{aligned} C^tA^{-1}C\vec{\lambda} &= C^tA^{-1}\vec{b} - \vec{e} \\ A\vec{x} &= C\vec{\lambda} + \vec{b}. \end{aligned}$$

Esta forma de resolver el sistema no es la más eficiente pues requiere del cálculo de la inversa de A. En la práctica se usa la factorización QR de la matriz C. Sistemas como el (2.8) aparecen en muchas aplicaciones como la dicretización de las ecuaciones de Navier-Stokes en mecánica de fluidos. Por ello ha recibido mucha atención de los especialistas.

3. El problema del portafolio se puede escribir en forma matricial. Sea $[\Sigma]$ la matriz de $n \times n$ cuyas componentes son iguales a

$$[\Sigma]_{ij} = \overline{Cov}(r_i, r_j)$$

a esta matriz se le conoce con el nombre de matriz de varianza-covarianza y siempre es positiva semidefinida. Denotemos como $\vec{1}$ al vector con componentes igual a 1 y $\vec{\overline{r}}$ el vector con componente i igual a $\overline{r_i}$ la media muestral de los rendimientos del activo i. Entonces el problema de minimización es

$$\begin{aligned} & Min\ \tfrac{1}{2}\vec{w}^t[\Sigma]\ \vec{w} \\ \text{sujeto a}\quad & \vec{\overline{r}}^t w = r^*, \\ & \vec{1}^t w = 1. \end{aligned}$$

Aplicando las condiciones de primer orden obtenemos el siguiente sistema de ecuaciones lineales

$$\begin{aligned} [\Sigma]\vec{w} + \lambda_1\vec{\overline{r}} + \lambda_2\vec{1} &= 0, \\ \vec{\overline{r}}^t\vec{w} &= r^*, \\ \vec{1}^t\vec{w} &= 1. \end{aligned}$$

Este sistema admite solución si la matriz $[\Sigma]$ de varianza-covarianza es positiva definida y $\vec{\bar{r}}$ y $\vec{1}$ son linealmente independientes. Esto último se satisface si los rendimientos históricos promedio de todos los activos no son iguales. Observe que esta condición garantiza que todos los puntos de Ω son regulares, con

$$\Omega = \{\vec{w} \in \Re^n | \vec{1}^t \vec{w} = 1, \quad \vec{\bar{r}}^t \vec{w} = r^*\}.$$

La solución $\vec{w}^*$ está dada por

$$[\Sigma]\vec{w}^* = -\lambda_1 \vec{\bar{r}} - \lambda_2 \vec{1}$$

con

$$\lambda_1 = \frac{B - r^* A}{\Delta} \quad \lambda_2 = \frac{r^* B - C}{\Delta},$$

y

$$A = \overrightarrow{1}^t [\Sigma]^{-1} \overrightarrow{1}, \quad B = \overrightarrow{1}^t [\Sigma]^{-1} \overrightarrow{r},$$

$$C = \overrightarrow{r^t} [\Sigma]^{-1} \overrightarrow{r}, \quad \Delta = AC - B^2.$$

Esta es la manera formal de calcular el sistema pues nunca se resuelve un sistema invirtiendo una matriz. Ver los problemas de este capítulo para aprender a resolver este sistema de una manera más eficiente.

Observemos que si la matriz $L = H_F = [\Sigma]$ es una matriz positiva definida en todo el espacio, también lo es en $N(\vec{w}^*)$.

2.3. Caso de restricciones de desigualdad

Cuando Ω está definido por restricciones de desigualdad e igualdad se trata el problema reduciéndolo al caso de restricciones de igualdad a través del concepto de restricciones activas y pasivas. Sea

$$\Omega = \{\vec{x} \in \Re^n | h_j(\vec{x}) \leq 0, j = 1, \ldots, m\}.$$

Definición 2.3.1. *Dado $\vec{x} \in \Omega$ se dice que $h_j(\vec{x})$ es una restricción activa en $\vec{x}$ si $h_j(\vec{x}) = 0$ y se dice que es pasiva si $h_j(\vec{x}) < 0$.*

Ejemplo 1

Consideremos el siguiente problema

$$\begin{array}{rrl} \text{Min} & F(x,y) = & (x-3)^2 + (y-3)^2, \\ \text{sujeto a} & h_1(x,y) = & x^2 + y^2 - 5 \leq 0, \\ & h_2(x,y) = & 3x + y - 5 \leq 0. \end{array}$$

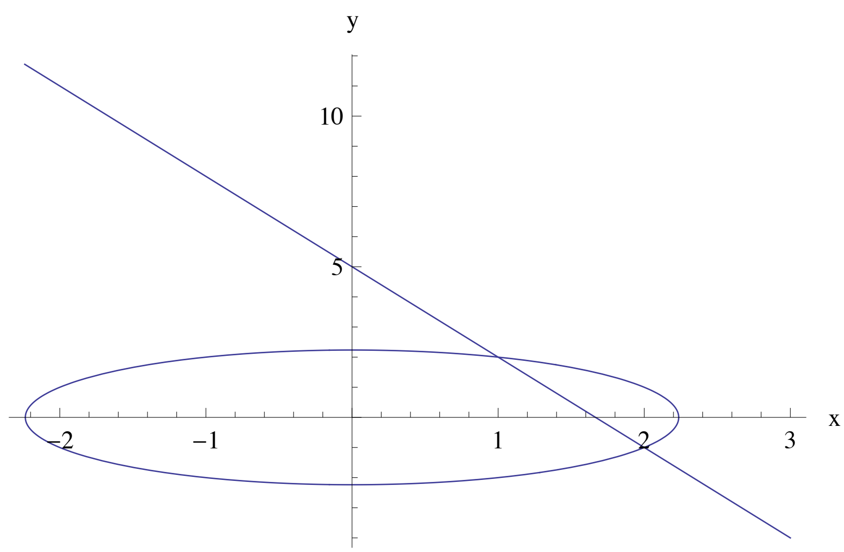

Figura 2.1: Región factible Ω.

En la Figura 2.1 se presenta el conjunto Ω que es la parte del círculo que esta por debajo de la recta, incluyendo la recta. Como se observa para todos los puntos en el interior de Ω ninguna de las dos restricciones es activa. En el caso que estemos sobre la recta $3x+y=5$ la restricción h_2 es activa mientras que h_1 no lo es, salvo para el caso de los puntos $(1,2)$, y $(2,-1)$ donde h_2 es también activa. Los puntos sobre la curva $x^2+y^2=5$ que están por debajo de la recta $3x+y=5$ tiene a h_2 como restricción activa.

Al graficar la región admisible junto con las curvas de nivel de F, ver la Figura 2.3, observamos que el mínimo debe encontrarse en los puntos cercanos a la intersección en el primer cuadrante de la recta con la circunferencia. Los puntos donde se intersectan la circunferencia y la recta son $(1,2)$ y $(2,-1)$ y

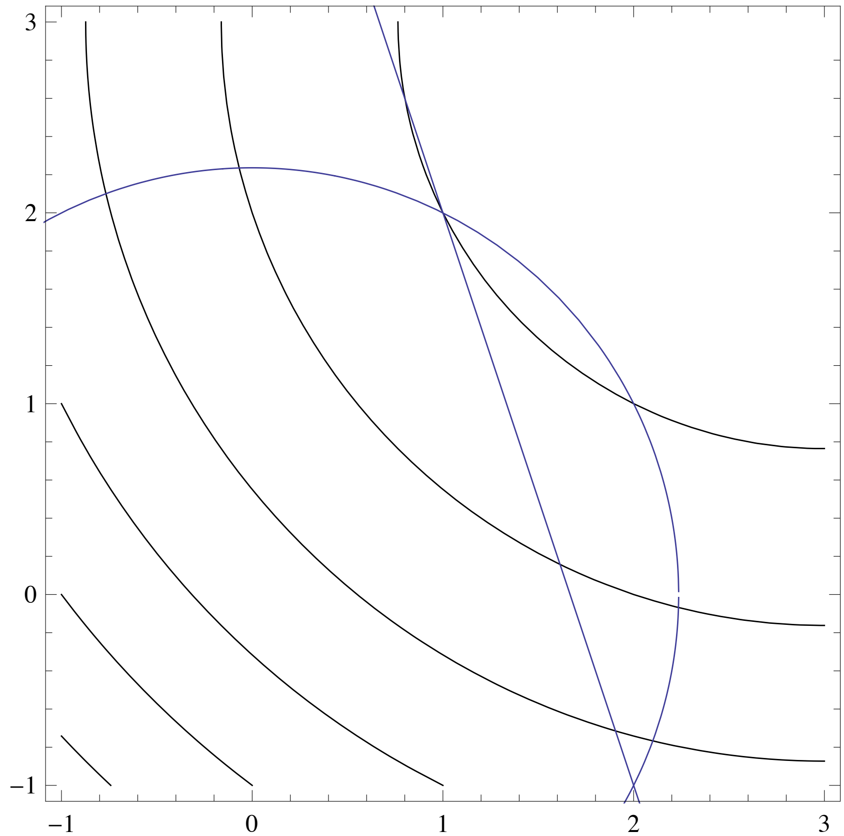

Figura 2.2: Región factible Ω y las curvas de nivel de F.

el primero es el mínimo de F restringido a Ω ya que está sobre la curva de nivel en donde F toma el valor más pequeño.

En general

$$\Omega = \{\vec{y} \in \Re^n |\ h_j(\vec{y}) = 0,\ j = 1, \ldots, m,\ f_j(\vec{y}) \leq 0,\ j = 1, \ldots, s\} \tag{2.10}$$

y las funciones h_j y f_j son funciones de $\Re^n$ a $\Re$.

Denotemos como $I(\vec{x})$ a los índices asociados a las restricciones activas en $\vec{x}$, en el caso de restricciones de desigualdad, el espacio $N(\vec{x})$ se define de la siguiente forma

$$N(\vec{x}) = \{\vec{y} \in \Re^n |\ {}^t\nabla h_j(\vec{x})\vec{y} = 0\ j = 1, \ldots, m \text{ y } {}^t\nabla f_j(\vec{x})\vec{y} = 0\ \forall j \in I(\vec{x})\}.$$

Asimismo, diremos en este caso que $\vec{x}$ es un punto regular de Ω si el conjunto de vectores formados por los gradientes de las restricciones activas son linealmente independientes.

Para el ejemplo anterior $N(0,0) = \emptyset$ porque ese punto es un punto interior de Ω. Pero

$$N(1,2) = \{\vec{y} = (a,b) \in \Re^2 |\ 3a + b = 0; \quad 2a + 4b = 0\} = \{(0,0)\}.$$

y

$$N(0, \sqrt{5}) = \{\vec{y} = (a,b) \in \Re^2 | b = 0\}.$$

Condiciones de Kuhn y Tucker para el caso no lineal

Considérese el problema

$$\begin{array}{rl} \text{Min} & F(\vec{x}) \\ \vec{x} \in \Omega & \end{array}$$

con $\Omega \neq \emptyset$ definido por

$$\Omega = \{\vec{y} \in \Re^n |\ h_j(\vec{y}) \leq 0\},$$

con $F, h_j : \Re^n \to \Re$ continuamente diferenciables en Ω.

Teorema 2.3.2. *Si $\vec{x}^*$ es un punto mínimo de F restringido a Ω y si $\vec{x}^*$ es un punto regular de Ω entonces existe $\vec{\mu} \in \Re^s$ con $\mu_j \geq 0$ para $j = 1, \ldots s$ tal que*

$$\nabla F(\vec{x}^*) + \sum_{j=1}^{s} \mu_j \nabla h_j(\vec{x}^*) = 0 \tag{2.11}$$

y

$$\mu_j [h_j(\vec{x}^*)] = 0 \quad \forall\ j = 1, \ldots, s. \tag{2.12}$$

Dem: Sea

$$S = \{\vec{x} \in \Omega |\ h_j(\vec{x}) = 0 \text{ para } j \in I(\vec{x}^*)\}$$

y $\vec{x}^*$ es el mínimo de F en Ω. Como $S \subset \Omega$, $\vec{x}^*$ también es un mínimo de F restringido a S; dado que las restricciones que definen a S son restricciones de igualdad, en $\vec{x}^*$ se deben satisfacer las condiciones de primer orden vistas en la sección anterior, por lo que existen μ_j para $j \in I(\vec{x}^*)$ tal que

$$\nabla F(\vec{x}^*) + \sum_{i \in I(\vec{x}^*)} \mu_i \nabla h_i(\vec{x}^*) = 0.$$

Si seleccionamos $\mu_i = 0$ para i no estando en $I(\vec{x}^*)$ entonces se obtienen las condiciones de Kuhn y Tucker

$$\nabla F(\vec{x}^*) + \sum_{i=1}^{m} \mu_i \nabla h_i(\vec{x}^*) = 0$$

y además se cumple que

$$\mu_i h_i(x) = 0 \quad i = 1, \ldots, m.$$

Resta por demostrar que las μ_i asociadas a las restricciones activas de $\vec{x}^*$ son no negativas. Esto lo haremos por reducción al absurdo: supongamos que existe una $\mu_k < 0$ para alguna $k \in I(\vec{x}^*)$. Sea S_k la superficie definida por todas las restricciones activas salvo la k-ésima restricción y sea $\hat{N}_k(\vec{x}^*)$ el espacio tangente asociado a S_k en el punto $\vec{x}^*$. Como $\vec{x}^*$ es un punto regular existe una $\vec{y} \in \Re^n$ tal que $\vec{y} \in \hat{N}_k(\vec{x}^*)$ y que satisface que $\nabla h_k^t(\vec{x}^*)\vec{y} < 0$. Recordemos que si $\vec{y} \in \hat{N}_k(\vec{x}^*)$ existe una curva $\vec{x}(t)$ que satisface que $\vec{x}(0) = \vec{x}^*$, $\vec{x}'(0) = \vec{y}$ y para alguna $\delta > 0$, $\vec{x}(t) \in S_k$ para $t \in < 0, \delta >$. Entonces

$$\frac{dF(x(t))}{dt}|_{t=0} = \ \nabla F^t(\vec{x}^*)\vec{y} = -\mu_k \ \nabla h_k^t(\vec{x}^*)\vec{y} < 0$$

$\Rightarrow \vec{y}$ es una dirección de descenso lo que contradice que $\vec{x}^*$ sea el mínimo. $\square$

A continuación se presentan las condiciones de segundo orden cuya demostración es similar al caso de igualdad.

Teorema 2.3.3. *Supongamos que $\vec{x}^*$ es un mínimo local del problema (P) y que $\vec{x}^*$ es un punto regular de Ω entonces existe un vector $\mu \in \Re^s$ tal que $\mu_j \geq 0$ y*

$$\nabla F(\vec{x}^*) + J_h^t(\vec{x}^*)\vec{\lambda} = 0.$$

Si $N(\vec{x}^*) = \{\vec{y} \in \Re^m | \nabla \vec{h}^t(\vec{x}^*)\vec{y} = 0 \ \forall j \in I(\vec{x}^*)\}$ entonces la matriz

$$L(\vec{x}^*) = F(\vec{x}^*) + \sum_{i=1}^{s} \mu_i H_{h_i}(\vec{x}^*)$$

es positiva semidefinida en $N(\vec{x}^*)$.

Teorema 2.3.4. *(Condiciones suficientes) Supóngase que hay un punto $\vec{x}^*$ en Ω y una $\vec{\mu} \in \Re^s$ tal que*

$$\nabla F(\vec{x}^*) + \sum_{j=1}^{s} \mu_j \ \nabla h_j(\vec{x}^*) = 0. \tag{2.13}$$

Supóngase también que la matriz

$$L(\vec{x}^*) = H_F(\vec{x}^*) + \sum_{i=1}^{s} H_{h_i}(\vec{x}^*)\mu_i$$

es positiva definida en

$$N(\vec{x}^*) = \{\vec{y} \in \Re^n |^t\nabla h_j(\vec{x}^*)\vec{y} = 0 \quad \forall j \in I(\vec{x}^*)\}.$$

$\Rightarrow$ $\vec{x}^*$ *es un mínimo estricto de* F *en* Ω*.*

La demostración de estos teoremas es similar al caso de restricciones de igualdad.

Ejemplos

1. Consideremos el siguiente problema

$$\begin{array}{rrl} \text{Min} & F(x,y) = & (x-3)^2 + (y-3)^2, \\ \text{sujeto a} & h_1(x,y) = & x^2 + y^2 - 5 \leq 0, \\ & h_2(x,y) = & 3x + y - 5 \leq 0. \end{array}$$

El gradiente de F es igual a $\nabla F(x,y) = (2(x-3), 2(y-3))$. Supongamos que h_1 y que h_2 no son activas entonces el punto $\vec{x}_1$ que hace al gradiente cero es $(3,3)$, punto que no es admisible. Por lo tanto alguna de las restricciones debe ser activa. Supongamos que h_1 es activa entonces el sistema a resolver es

$$\begin{array}{rcl} 2(x-3) + 2x\mu_1 & = & 0, \\ 2(y-3) + 2y\mu_1 & = & 0, \\ x^2 + y^2 & = & 5. \end{array}$$

La solución del sistema es $(\sqrt{5/2}, \sqrt{5/2})$ es la solución con $\mu_1 = 0.8973$. Este punto no está en Ω por lo que no es un punto admisible. Supongamos que h_2 es activa y h_1 es pasiva entonces ${}^t\nabla h_2(x,y) = (3,1)$. Entonces el sistema de ecuaciones a resolver es

$$\begin{array}{rcl} 2(x-3) + 3\mu_2 & = & 0, \\ 2(y-3) + \mu_2 & = & 0, \\ 3x + y & = & 5. \end{array}$$

La solución del sistema es $(9/10, 23/10)$ con $\mu_2 = \frac{7}{5}$. Este punto tampoco es admisible.

Supongamos ahora que h_1 y h_2 son activas, entonces el sistema correspondiente a resolver es

$$\begin{aligned} 2(x-3) + 2x\mu_1 + 3\mu_2 &= 0, \\ 2(y-3) + 2y\mu_1 + \mu_2 &= 0, \\ 3x + y &= 5, \\ x^2 + y^2 &= 5. \end{aligned}$$

Los únicos puntos que tienen estas restricciones activas son $(1, 2)$ y $(2, -1)$. Determinemos el valor de los multiplicadores de Lagrange asociados a $(1, 2)$ son $\mu_1 = 1/5$ y $\mu_2 = 6/5$, mientras que para $(2, -1)$ son $\mu_1 = -22/10$ y $\mu_2 = 18/5$. Entonces $(1, 2)$ es candidato a ser el mínimo por ser el único punto que satisface las condiciones de Kuhn-Tucker (KT). La matriz L respectiva es

$$L(1,2) = \begin{pmatrix} 2 + 2\mu_1 & 0 \\ 0 & 2 + 2\mu_1 \end{pmatrix}.$$

Dado que $\mu_1 = 1/5$ esta matriz es positiva definida para cualquier vector de $\Re^2$ distinto de cero por lo que $(1, 2)$ es el mínimo de F restringido a Ω.

2. El problema de optimización de portafolios sin ventas en corto es un problema de optimización cuadrática con desigualdades lineales. En este caso las condiciones KT correspondientes son: existen λ, $\mu \in \Re$ y $\nu_i \geq 0$, con $i = 1, \ldots, n$ tal que

$$\Sigma w + \lambda \overrightarrow{r} + \mu \overrightarrow{1} - \sum_{i=1}^{n} \nu_i \overrightarrow{e}_i = 0, \tag{2.14}$$

$$w^t \overrightarrow{r} = r^*, \tag{2.15}$$

$$\overrightarrow{1}^t w = 1, \tag{2.16}$$

$$\nu_i w_i = 0, \quad i = 1, \ldots, n \tag{2.17}$$

con $\overrightarrow{e}_i$ el i-ésimo vector de la base canónica de $\Re^n$.

Si denotamos como w^* el punto admisible que satisface las condiciones de KT entonces las condiciones de segundo orden para que este punto sea un mínimo de la función objetivo en Ω están dadas por

$$y^t[\Sigma]y > 0$$

para toda $y \in N(w^*)$ con $y \neq 0$. Como $[\Sigma]$ es positiva definida en todo $\Re^n$ también lo es en $N(w^*) \subset \Re^n$.

El algoritmo a seguir en este caso es clasificar todos los puntos de

$$\Omega = \left\{ \begin{array}{ll} w \in \Re^n | & h_1(w) = w^t \overrightarrow{r} - r^* = 0, \\ & h_2(w) = \overrightarrow{1}^t w - 1 = 0, \\ & h_{i+2} = -w_i \leq 0, \ i = 1, \ldots, n \end{array} \right\}$$

dependiendo de si las restricciones h_i son activas o pasivas. Es decir h_i es activa si $w_i = 0$ y es pasiva si $w_i > 0$. Para cada subconjunto hay que comprobar si en algún punto se satisfacen las condiciones de Kuhn y Tucker.

Consideremos el ejemplo anterior cuando se tienen tres activos no correlacionados, el problema a minimizar es el siguiente

$$\begin{gathered} Min\ \tfrac{1}{2}[0.2w_1^2 + 0.18w_2^2 + 0.15w_3^2] \\ \text{sujeto a } 0.2w_1 + 0.25w_2 + 0.15w_3 = r^*, \\ \textstyle\sum_{i=1}^3 w_i = 1, \\ w_i \geq 0, \quad i = 1, \ldots, 3. \end{gathered}$$

Como primer paso clasifiquemos los puntos de Ω dependiendo de que las restricciones $h_i(w) = w_i$ sean pasivas o activas. Para ello, definamos el conjunto

$$\hat{\Omega} = \{w \in \Re^n | w^t \overrightarrow{r} = r^*, w^t \overrightarrow{1} = 1\}.$$

Entonces los puntos admisibles se pueden clasificar en los siguientes conjuntos

$$\begin{aligned}
S_1 &= \{w \in \hat{\Omega} | w_i > 0, i = 1, \ldots, 3\}, \\
S_2 &= \{w \in \hat{\Omega} | w_1 = 0\}, \\
S_3 &= \{w \in \hat{\Omega} | w_2 = 0\}, \\
S_4 &= \{w \in \hat{\Omega} | w_3 = 0\}, \\
S_5 &= \{w \in \hat{\Omega} | w_1 = w_2 = 0\} = \{(0,0,1)\}, \\
S_6 &= \{w \in \hat{\Omega} | w_1 = w_3 = 0\} = \{(0,1,0)\}, \\
S_7 &= \{w \in \hat{\Omega} | w_2 = w_3 = 0\} = \{(1,0,0)\}.
\end{aligned}$$

Observemos que en S_5, S_6 y S_7 ningún punto es regular, por haber más restricciones que incógnitas, por lo que no se cumplen las condiciones de KT. Analicemos si existe algún punto w de S_1 que satisfaga las condiciones de KT correspondientes: existen λ y $\mu \in \Re$, tal que

$$\Sigma w + \lambda \overrightarrow{r} + \mu \overrightarrow{1} = 0, \tag{2.18}$$

$$w^t \overrightarrow{r} = r^*, \tag{2.19}$$

$$\overrightarrow{1}^t w = 1. \tag{2.20}$$

Al resolver este sistema en términos de r^* obtenemos que si $r^* \in [0.163636, 0.234483]$ entonces

$$\begin{aligned}
w_1^* &= 0.53097345 r^* + 0.18584071, \\
w_2^* &= 9.73451327 r^* - 1.59292035, \\
w_3^* &= 2.40707965 - 10.2654867 r^*.
\end{aligned}$$

Para S_3 se cumplen las condiciones de KT para

$$w_1^* = 20r^* - 3, \quad w_2^* = 0 \quad y\ w_3^* = 4 - 20r^*,$$

siempre que $r^* \in [.15, .163636]$.

Para S_4 se cumplen las condiciones de KT para

$$w_1^* = 5 - 20r^*, \quad w_2^* = 20r^* - 4 \quad y\ w_3^* = 0.$$

si $r^* \in [0.234482, 0.25]$.

3. Consideremos el caso de que la función objetivo sea una función no lineal con restricciones no lineales. Consideremos el problema

$$Min\ e^{-(x+y)}$$
$$\text{sujeto a } e^x + e^y \leq 20,$$
$$x \geq 0.$$

En este caso el conjunto admisible es

$$\Omega = \{(x,y) \in \Re^2 |\ h_1(x,y) = e^x + e^y - 20 \leq 0;\ h_2(x,y) = -x \leq 0\}.$$

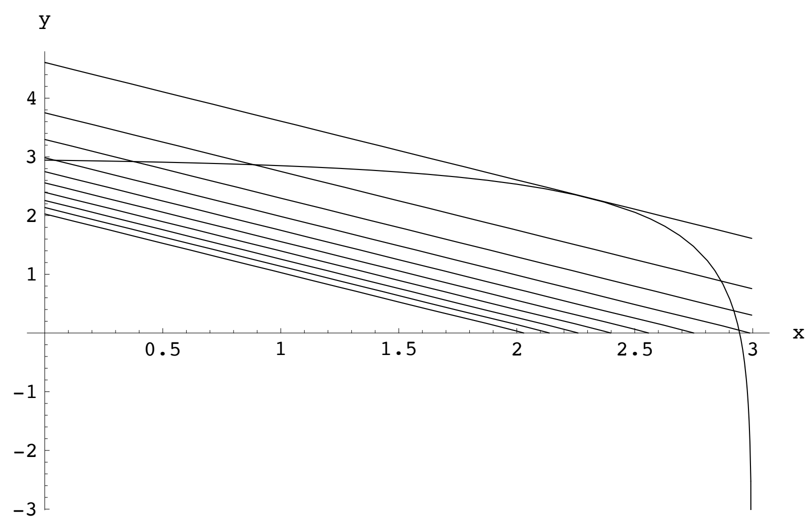

Figura 2.3: Región factible Ω con las curvas de nivel de F.

En la Figura 2.3 se presenta la solución gráfica de este problema. Observemos que el punto $\vec{x}$ de Ω que esta en la curva de nivel de menor valor es aquel que está satisface $h_1 = 0$.

Clasifiquemos los puntos dependiendo de si las restricciones son activas o pasivas.

$$\begin{aligned}
S_1 &= \{(x,y) \in \Omega|\ h_1(x,y) = e^x + e^y - 20 = 0;\ h_2(x,y) = -x < 0\}, \\
S_2 &= \{(x,y) \in \Omega|\ h_1(x,y) < 0;\ h_2(x,y) = -x = 0\}, \\
S_3 &= \{(0,y) \in \Omega|\ h_1(x,y) = 0\} = \{(0, \ln(19)\}.
\end{aligned}$$

Chequemos para cada subconjunto si se cumplen las condiciones de Kuhn-Tucker en algún punto. Un punto es regular en S_1 si $\{(e^x, e^y)\}$ es linealmente independiente, lo cual se cumple para todo $(x, y) \in S_1$. Las condiciones de Kuhn-Tucker correspondientes son:

$$\begin{aligned} \nabla F(x,y) + \mu_1 \nabla h_1(x,y) &= (0,0), \\ h_1(x,y) &= 0. \end{aligned}$$

Esto se reduce a resolver el siguiente sistema de ecuaciones no-lineal

$$\begin{aligned} -e^{-x} + \mu_1\ e^{2y} &= 0, \\ -e^{-y} + \mu_1\ e^{2x} &= 0, \\ e^x + e^y &= 20. \end{aligned}$$

cuya solución es $x = y = -1/3 \ln(\mu_1)$ con $\mu_1 = .001$; μ_1 es mayor que cero y $x = \ln(10) = y$ es un punto admisible de S_1. Calculemos la matriz L

$$L(\vec{x}, \mu_1) = H_F(\vec{x}) + \mu_1 H_{h_1}(\vec{x})$$

lo que es igual a

$$L(\vec{x}, \mu_1) = \begin{pmatrix} e^{-(x+y)} + \mu_1 e^x & e^{-(x+y)} \\ e^{-(x+y)} & e^{-(x+y)} + \mu_1 e^y \end{pmatrix}$$

que al evaluarla en $(\ln(10), \ln(10), .001)$ da una matriz positiva definida en $\Re^2$ por lo que este punto es el mínimo de F restringido a Ω. $F(\ln(10), \ln(10)) = .01$. Cheque el lector que las condiciones de Kuhn-Tucker no se cumplen en ningún punto de S_2 y S_3.

2.3.1. Funciones convexas

Definición 2.3.5. *Sea $\Omega \subset \Re^n$ se dice que Ω es convexo si para cualquier $\vec{x}$ y $\vec{y}$ en Ω y $\lambda \in (0,1)$ tal que $\vec{x} + \lambda(\vec{y} - \vec{x}) \in \Omega$.*

La interpretación geométrica de esta definición es que cualquier recta que una a dos puntos del conjunto se encuentra totalmente contenida en el conjunto.

Definición 2.3.6. *Sea $\Omega \subset \Re^n$ y sea $F : \Omega \to \Re^n$ si para cualesquiera $\vec{x}, \vec{y} \in \Omega$ y $\lambda \in (0,1)$ se cumple*

$$F(\lambda\vec{x} + (1-\lambda)\vec{y}) \leq \lambda F(\vec{x}) + (1-\lambda) F(\vec{y})$$

y se dice que es estrictamente convexa si se cumple

$$F(\lambda\vec{x} + (1-\lambda)\vec{y}) < \lambda F(\vec{x}) + (1-\lambda)F(\vec{y}).$$

La interpretación geométrica en $\Re$ es que la función evaluada sobre algún punto de la recta que une a x con y es mayor o igual al valor que toma la recta que une a los puntos $(x, f(x))$ y $(y, f(y))$.

Ejemplos

1. En $\Re$, $f(x) = x$, $f(x) = x^2$ y $f(x) = e^x$.

2. En $\Re^2$, $f(x, y) = x^2 + y^2$, $f(x, y) = e^{x+y}$.

3. En $\Re^n$, $f(\vec{x}) = \vec{c}^t\vec{x}$.

Demostrar que f es convexa usando la definición no siempre es sencillo, pero el siguiente resultado nos permite probarlo más f'acilmente.

Lemma 2.3.7. *Sean $f, g : \Omega \subset \Re^n \to \Re$ con Ω convexo en $\Re^n$*

1. *Si f es convexa y $\alpha > 0 \Rightarrow \alpha f$ es convexa.*

2. *Si f y g son convexas $\Rightarrow f + g$ es convexa.*

3. *Si f es convexa y g es creciente $\Rightarrow g \circ f$ es convexa.*

4. *Si f y g son convexas $\Rightarrow g \circ f$ es convexa.*

Ejemplos

1. $f(x, y, z) = x^2 + y^2 + z^2$ es convexa.

2. $f(x, y, z) = e^{x^2+y^2+z^2}$ es convexa.

Teorema 2.3.8. *Cualquier mínimo local de F definido en un convexo Ω es un mínimo global en Ω.*

Dem: Si $\vec{x}^*$ es un mínimo local entonces existe $r > 0$ tal que $V_r(\vec{x}^*) \subset \Omega$ y para toda $\vec{x} \in V_r(\vec{x}^*)$ se cumple $F(\vec{x}^*) \leq F(\vec{x})$. Sea $\vec{y}$ cualquier otro elemento de Ω y sea $\lambda \in (0,1)$ tal que $\vec{x}^* + \lambda(\vec{y} - \vec{x}^*) \in V_r(\vec{x}^*)$ entonces

$$F(\vec{x}^*) \leq F((1-\lambda)\vec{x}^* + \lambda\vec{y}) \leq (1-\lambda)F(\vec{x}^*) + \lambda F(\vec{y})$$

por ser F convexa. Entonces

$$0 \geq \lambda(F(\vec{y}) - F(\vec{x}^*))$$

y de aquí se concluye que $F(\vec{x}^*) \leq F(\vec{y})$.

Las funciones convexas en un convexo pueden caracterizarse a través de la primera derivada.

Teorema 2.3.9. *Si $F : \Re^n \to \Re$ es continuamente diferenciable y convexa en $\Omega \subset \Re^n$ convexo $\Rightarrow$ para toda $\vec{x}$ y $\vec{y} \in \Omega$ se cumple*

$$F(\vec{x}) + \nabla F(\vec{x})^t(\vec{y} - \vec{x}) \leq F(\vec{y}).$$

También el hessiano puede dar información sobre la convexidad de la función.

Teorema 2.3.10. *Si $F : \Re^n \to \Re$ es dos veces continuamente diferenciable y convexa en $\Omega \subset \Re^n$ convexo $\Rightarrow$ para toda $\vec{x} \in \Omega$ se cumple $H_F(\vec{x})$ es positiva definida.*

2.4. Ejercicios

1. Plantee y resuelva analíticamente el siguiente problema. El Sol de Mérida fue recientemente adquirido por Televisa. Este se vende a \$2.00 el ejemplar y tiene una circulación diaria de 20,000 números. Por cuestión de venta de anuncios gana $\$1,000$ por página y el periódico vende 15 páginas diarias. La nueva administración desea incrementar sus ganancias y desea reducir sus gastos semanales. El periódico gasta $\$60,000$ en su departamento editorial (escritores, reporteros, fotógrafos,etc), $\$20,000$ en su departamento de publicidad y suscripciones y $\$50,000$ de gastos fijos a la semana. Si se reduce el presupuesto del departamento editorial se ahorraría dinero, pero afectaría la calidad del periódico. El mínimo presupuesto con el que puede funcionar este departamento es

de \$40,000. Estudios demuestran que por cada 10 % de reducción de presupuesto de este departamento se pierde un 2 % de suscriptores y uno por ciento por venta de anuncios. Recientemente, otro periódico, incrementó su presupuesto del departamento de publicidad en un 20 % y como consecuencia se incrementó en un 15 % el número de páginas de anuncios vendidas. Los nuevos dueños del Sol de Mérida están dispuestos a gastar hasta \$40,000 en su departamento de publicidad, ¿Qué estrategia hay que seguir para maximizar las ganacias dado que el monto total de gastos no puede exceder los \$140,000 pesos a la semana?

2. Una compañía planea fabricar cajas rectangulares cerradas con un volumen de 8lt. El material para la base y la tapa cuesta el doble que el material para los lados. Encuentre las dimensiones para las cuales el costo es mínimo.

3. El cono $z^2 = x^2 + y^2$ está cortado por el plano $z = 1 + x + y$. Hállense los puntos sobre esta sección más próximos al origen.

4. Se trata de montar un radiotelescopio en un planeta recien descubierto. Para minimizar la interferencia se desea emplazarlo donde el campo magnético sea más débil. Supongamos que se modela el planeta usando una esfera con un radio de 6 unidades. Se sabe que la fuerza magnética esta dada por $G(x, y, z) = 6x - y^2 + xz + 60$, considerando un sistema coordinado cuyo origen está en el centro del planeta. ¿Dónde hay que ubicar al radiotelescopio?

5. Dados n números positivos $a_1, a_2, \ldots, a_n$, hállese el valor máximo de la expresión
$$w(x) = \sum_{i=1}^{n} a_i x_i$$
si $\sum_{i=1}^{n} x_i^2 = 1$.

6. Sea $A \in \Re^{n \times n}$ una matriz positiva definida. Sea $B \in \Re^{n \times m}$ con $m < n$ una matriz de rango completo entonces el sistema
$$B^t A^{-1} B \lambda = c$$
con $\lambda \in \Re^m$ admite una solución única.

7. Resuelva analíticamente el siguiente problema

$$\begin{array}{rcl} \text{Min} & & x^2 - xy + y^2 - 3x \\ \text{sujeto a } x, y & \geq & 0 \\ x + y & \leq & 1. \end{array}$$

Bosqueje el conjunto Ω admisible.

8. Resuelva analíticamente

$$\begin{array}{rcl} \text{Min} & & x_1^3 + x_2^2 \\ \text{sujeto a} & & x_1^2 + x_2^2 - 10 = 0, \\ 1 - x_1 & \leq & 0, \\ 1 - x_2 & \leq & 0, \end{array}$$

Grafique la región admisible.

9. Se tiene un portafolio con tres activos con los siguientes datos

	A_1	A_2	A_3
r_i	.4	.8	.8
σ_i^2	.2	.25	.2
σ_{ij}	σ_{12}=.1	σ_{13}=0.1	$\sigma_{23} = 0.05$

determine la composición del portafolio que minimiza el riesgo, con rendimiento esperado igual a r^*, determine los posibles valores que puede tomar r^*, con y sin ventas en corto, si la suma de las w_i debe ser igual a 1.

10. Sea

$$\begin{array}{rl} Min & F(\vec{x}) \\ \text{sujeto a} & g_i(\vec{x}) \leq 0 \quad i = 1, \ldots m. \end{array}$$

se dice que este problema es de programación convexa si $F, h_i : C \to \Re^n$ con C conjunto convexo y F, h_i, para $i = 1, \ldots m$ son convexas. Demuestre que el siguiente problema es un problema de programación convexa y que $(0, 0)$ es la única solución.

$$
\begin{array}{rc}
Min & x^4 + y^4 \\
\text{sujeto a:} & x^2 - 1 \leq 0 \\
 & y^2 - 1 \leq 0 \\
 & e^{x+y} - 1 \leq 0, \quad (x,y) \in \Re^2.
\end{array}
$$

11. Determine la solución del siguiente problema y demuestre también que es un problema de programación convexa: Minimice $F(x,y) = 2x^2+y^2-2xy-5x-2y$ sujeto a $h_1(x,y) = 3x+2y \leq 20$, $h_2(x,y) = 5x-2y \geq -4$, y $x, y \geq 0$.

12. (Factorización QR)

    Dada una matriz A de $n \times m$ de rango m con $m \leq n$ exite una matriz Q de $n \times n$ ortogonal, es decir $Q^t = Q^{-1}$, y una matriz R de $n \times m$ triangular superior en los primeros m renglones y con elementos igual a cero en todos los $n-m$ restantes renglones tal que

    $$A = QR.$$

    Una forma de construir las matrices Q y R es a través de las transformaciones de Householder. Dada una matriz A con columnas formadas por los vectores $\vec{A}_1, \ldots, \vec{A}_m$ de dimensión n la matriz P_1 de la forma

    $$P_1 = I - \frac{2}{\vec{v}_1^t \vec{v}_1} \vec{v}_1 \ \vec{v}_1^t,$$

    con $\vec{v}_1 = \vec{A}_1 + sign(A_{11})\alpha_1 \vec{e}_1$ y $\alpha_1 = \{\sum_{i=1}^n A_{i1}^2\}^{1/2}$.
    Entonces

    $$P_1 A = \begin{pmatrix} A_{11}^1 & \ldots & A_{1m}^1 \\ 0 & A_{22}^1 \ldots & A_{2m}^1 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & A_{nm}^1 \end{pmatrix}.$$

    Para cualquier columna $i > 1$, la matriz P_i es igual a

    $$P_i = I - \frac{2}{\vec{v}_i^t \vec{v}_i} \vec{v}_i \vec{v}_i^t,$$

con $\vec{v}_i = (0, \ldots, A_{ii}^{i-1} + sign(A_{ii}^{i-1})\alpha_i, A_{i+1i}^{i-1}, \ldots, A_{in}^{i-1})$ y $\alpha_i = \{\sum_{j=i}^{n}(A_{j1}^{i-1})^2\}^{1/2}$. Entonces

$$Q^t A = P_n P_{n-1} \ldots P_1 A = R$$

y R es una matriz con las características deseadas.

Aplique el siguiente procedimiento para factorizar las matrices

$$A = \begin{pmatrix} 2 & 1 & 1 & 4 \\ 0 & 1 & 2 & 1 \end{pmatrix},$$

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

13. Sea A una matriz que satisface las hipótesis del ejercicio anterior, si Q es la matriz ortogonal de $n \times n$ que aparece en su factorización QR entonces si Q se descompone de la forma

$$Q = [Q_1 | Q_2]$$

con Q_1 una matriz de $n \times m$ y Q_2 una matriz de $n \times (n-m)$, las columnas de Q_2 generan el espacio $N(A)$. Entonces para cualquier vector $\vec{z} \in \Re^{n-m}$, $Q_2 z \in N(A)$. Usando lo anterior se puede aplicar la factorización de Cholesky para demostrar que una matriz G es positiva definida en $N(A)$, basta aplicarlo a la matriz ${}^tQ_2\ G\ Q_2$. Demostrar que

$$cond(Q_2^t\ G\ Q_2) \leq cond(G)$$

y aplicar este procedimiento para demostrar si el problema

$$\begin{array}{rcl} \text{Min} & & x_1^3 + x_2^2 \\ \text{sujeto a} : 1 - x_1 & = & 0. \end{array}$$

admite un mínimo.

14. Demostrar que la matriz de proyección al Nucleo de A: $P_{N(A)}$ puede definirse en términos de las matriz Q_2 como

$$P_{N(A)} = Q_2 (Q_2^t\ Q_2)^{-1} Q_2^t,$$

entonces la matriz de proyección al $R(A^t)$ está dada por

$$P_{R(A^t)} = I - Q_2 (Q_2^t\ Q_2)^{-1}\ Q_2^t.$$

Capítulo 3

Método de gradiente proyectado

En este capítulo se verá el método de gradiente proyectado para resolver numéricamente los problemas con restricciones de igualdad y desigualdad cuando las restricciones son lineales. En la primera sección se aplicarán para el caso de restricciones de igualdad. En la segunda sección cuando se tienen restricciones de desigualdad y en la última sección se presenta el Método de Wolfe. Cabe mencionar que hay numerosos métodos de aproximación que buscan sacar provecho de las características de la función objetivo y de las restricciones. Hay métodos específicos para funciones convexas definidas en conjuntos convexos o funciones cuadráticas con restricciones lineales. Como en el caso sin restricciones, el tipo de métodos que se presentan son de descenso, pero con la diferencia que los elementos de la sucesión deben satisfacer las restricciones.

3.1. Método de gradiente proyectado

Consideremos que tenemos el siguiente problema

$$\min_{\vec{x}\in\Omega} F(\vec{x}),$$

con Ω un subconjunto distinto del vacío de $\Re^n$. Supongamos que el problema admite una solución.

Los métodos de descenso para resolver estos problema consisten en lo siguiente:

$$\begin{aligned}
\text{dado } \vec{x}_0 &\in \Omega, \\
\text{se genera } &\vec{x}_{n+1} \in \Omega \text{ tal que} \\
\vec{x}_{n+1} &= \vec{x}_n + \alpha_n \vec{d}_n, \\
\text{con} \qquad F(x_n + \alpha_n \vec{d}_n) &\leq F(x_n).
\end{aligned}$$

3.1.1. Caso de restricciones lineales de igualdad

Consideremos el caso que F sea una función dos veces continuamente diferenciable en un abierto que contenga a Ω y que Ω sea de la forma

$$\Omega = \{\vec{x} \in \Re^n | A^t \vec{x} = \vec{e}\},$$

con A una matriz en $\Re^{n\times m}$, con $m \leq n$, de rango completo.

Punto inicial

El primer problema que surge es cómo seleccionar un punto admisible. La forma más eficiente es usar la factorización QR de la matriz A. Si $A \in \Re^{n\times m}$ es una matriz de rango completo igual a m entonces existe una matriz ortogonal $Q \in \Re^{n\times n}$ y una matriz $\widetilde{R} \in \Re^{n\times m}$, con transpuesta igual a $[R^t, 0]$ y $R \in \Re^{m\times m}$ triangular superior, tal que $A = Q\widetilde{R}$.

Observemos que $Q = [Q_1, Q_2]$ con $Q_1 \in \Re^{n\times m}$ y $Q_2 \in \Re^{n\times n-m}$ tal que $Q_1^t A = R$ y $Q_2^t A = 0$. Entonces resolver el sistema $A^t\vec{x} = \vec{e}$ es equivalente a resolver primero

$$R^t \vec{z} = \vec{e}$$

y posteriormente a determinar $\vec{x}$ por

$$\vec{x} = Q_1 \vec{z}.$$

El primer sistema tiene solución única porque R es una matriz invertible.

Direcciones admisibles

Dado $\vec{x}_0 \in \Omega$, ¿cómo garantizamos que el punto $\vec{x}_1 = \vec{x}_0 + \alpha\vec{d}_0$ también esté en Ω?

$$A^t(\vec{x}_0 + \alpha\vec{d}_0) = A^t x_0 + \alpha A^t \vec{d}_0 = \vec{e} + \alpha A^t \vec{d}_0$$

y $\vec{x}_1 \in \Omega$ siempre que $A^t\vec{d}_0 = 0 \Rightarrow$ que $\vec{d}_0 \in EN(A^t)$ con

$$EN(A^t) = \{\vec{y} \in \Re^n | A^t\vec{y} = 0\}.$$

Si la dimensión de $EC(A) = m < n$ se tiene que todos los puntos de Ω son regulares y la dimensión de $EN(A^t)$ =dim $EC(A)^\perp = n - m$. Basta con escoger como direcciones de descenso a vectores $\vec{d}_k$ en $EN(A^t)$ para que la sucesión generada por el método de descenso permanezca en Ω. A estas direcciones se les conoce con el nombre de direcciones admisibles.

Condiciones de primero y segundo orden

Las condiciones de primero y segundo orden, vistas en el capítulo anterior, nos permiten asegurar que si A es una matriz de rango completo y existe una solución $(\vec{x}^*, \vec{\lambda}^*)$ del sistema

$$\begin{aligned} \nabla F(\vec{x}^*) + A\vec{\lambda}^* &= 0, && (3.1) \\ A^t\vec{x}^* &= \vec{e}, && (3.2) \end{aligned}$$

que satisface

$$\vec{y}^t H_f(\vec{x}^*) y > 0 \quad \forall\, \vec{y} \neq 0,\ \vec{y} \in EN(A^t)$$

$\Rightarrow \vec{x}^*$ es un mínimo estricto de F en Ω.

Con objeto de desacoplar las ecuaciones (3.1), usaremos el Lema 2.2.2. de la sección anterior: si $\vec{x}^*$ es un punto admisible y regular de Ω que es punto extremo de F restringido a Ω entonces resolver el sistema anterior es equivalente a determinar primero $\vec{x}^*$ que satisfaga

$$\nabla F(\vec{x}^*)^t\ \vec{y} = 0 \quad \forall\, \vec{y} \in EN(A^t) \tag{3.3}$$

y posteriormente

$$A\lambda^* = -\nabla F(\vec{x}^*).$$

3.1.2. Método de Newton

Para determinar numéricamente un punto $\vec{x}^* \in \Omega$ que satisfaga la ecuación (5.2) y (3.3) usaremos el método de descenso; en cada iteración $k+1$ se genera un punto $\vec{x}_{k+1}$ que satisfaga (3.3) para la aproximación lineal de $\nabla F(\vec{x}_{k+1})$ por la serie de Taylor alrededor de $\vec{x}_k$.

$$\nabla F(\vec{x}_{k+1}) \approx \nabla F(\vec{x}_k) + H_F(\vec{x}_k)(\vec{x}_{k+1} - \vec{x}_k).$$

Es decir, se busca un punto $\vec{x}_{k+1}$ que satisfaga

$$0 = \vec{y}^t \ \nabla F(\vec{x}_{k+1}) \approx \vec{y}^t \ \nabla F(\vec{x}_k) + \vec{y}^t \ H_F(\vec{x}_k) \ (\vec{x}_{k+1} - \vec{x}_k) \tag{3.4}$$

para toda $\vec{y} \in EN(A^t)$.

Para garantizar que (3.4) se cumple para toda $\vec{y} \in EN(A^t)$ basta hacerlo para los elementos de una base. Sea $\{z_1, z_2, \ldots, z_{n-m}\}$ una base de $EN(A^t)$ y sea $Z \in \Re^{n \times n-m}$ la matriz con i-ésima columna igual a z_i entonces en la iteración $k+1$ se debe cumplir que

$$0 = Z^t \nabla F(\vec{x}_{k+1}).$$

Recordemos que la dirección de descenso en cada iteración debe ser una dirección admisible lo que implica que existe $\vec{b}_k \in R^{n-m}$ tal que $Zb_k = d_k$ y por lo tanto

$$0 = Z^t \nabla F(\vec{x}_k + \alpha_k Z\vec{b}_k) \approx Z^t \nabla F(\vec{x}_k) + \alpha_k Z^t H_F(\vec{x}_k) Z\vec{b}_k$$

$\Rightarrow$

$$Z^t H_F(\vec{x}_k) Z \ \vec{b}_k = -Z^t \nabla F(\vec{x}_k).$$

3.1.3. Algoritmo de Newton

1. Dado $x_0 \in \Omega$.
2. Para $k = 1, 2, \ldots$ determine la dirección de descenso $\vec{d}_k$ por

$$Z^t H_F(\vec{x}_k) Z \ \vec{b}_k = -Z^t \nabla F(\vec{x}_k), \tag{3.5}$$

$$\vec{d}_k = Z\vec{b}_k. \tag{3.6}$$

   y

$$\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{d}_k$$

   con α_k seleccionada para que $F(\vec{x}_{k+1}) < F(\vec{x}_k)$.

3. Si $\|Z^t \nabla F(\vec{x}_{k+1})\| < rtol$ y $\frac{\|\vec{x}_{k+1} - \vec{x}_k\|}{\|\vec{x}_k\|} \leq rtol \Rightarrow$

$$A^t A \lambda_{k+1} = -A^t \nabla F(\vec{x}_{k+1}). \tag{3.7}$$

   y $\vec{x}^* \approx \vec{x}_{k+1}$, $\lambda \approx \lambda_{k+1}$.

4. Si no se satisface el criterio de paro regresar a 2 y calcular otra iteración.

Observaciones:

1. Para comprobar en cada iteración que las condiciones de segundo orden se cumplen, el sistema (3.5) debe resolverse por Cholesky.

2. El cálculo del multiplicador de Lagrange asociado al mínimo $\vec{x}^*$ se obtiene al resolver
$$A^t A\lambda = -A^t \nabla F(\vec{x}^*).$$
En el algoritmo anterior λ se estima por λ_{k+1}, solución de
$$A^t A\lambda_{k+1} = -A^t \nabla F(\vec{x}_{k+1}).$$
Este valor se mejora cuando $\nabla F(\vec{x}^*)$ se aproxima por medio de los dos primeros términos de la serie de Taylor cuando se expande alrededor de $\vec{x}_{k+1}$
$$A^t A\lambda_{k+1} = -A^t[\nabla F(\vec{x}_{k+1}) + H_F(\vec{x}_{k+1})(\vec{x}_{k+1} - \vec{x}_k)]. \qquad (3.8)$$

El algoritmo anterior se simplifica si contamos con la factorización QR de la matriz A.

Algoritmo de Newton con factorización QR

1. Dado $x_0 \in \Omega$ y $A = QR$, matriz de rango completo.

2. Para $k = 1, 2, \ldots$ determine la dirección de descenso $\vec{d}_k$ por
$$\begin{aligned} Q_2^t H_F(\vec{x}_k) Q_2\, \vec{b}_k &= -Q_2^t \nabla F(\vec{x}_k), \\ \vec{d}_k &= Q_2 \vec{b}_k. \end{aligned}$$
y
$$\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{d}_k$$
con α_k seleccionada para que $F(\vec{x}_{k+1}) < F(\vec{x}_k)$.

3. Si $\|Q_2^t \nabla F(\vec{x}_{k+1})\| < rtol$ y $\frac{\|\vec{x}_{k+1} - \vec{x}_k\|}{\|\vec{x}_k\|} \leq rtol \Rightarrow$
$$R\lambda_{k+1} = Q_1^t \nabla F(\vec{x}_{k+1}). \qquad (3.9)$$
y $\vec{x}^* \approx \vec{x}_{k+1}$, $\lambda \approx \lambda_{k+1}$.

4. Si no se satisface el criterio de paro regresar a 2.

Ejemplos

1. Dada una función cuadrática

$$F(x) = \frac{1}{2}\vec{x}^t G\vec{x} - \vec{x}^t\vec{f} + c$$

con G matriz positiva definida, supongamos que

$$\Omega = \{\vec{x} \in \Re^n | A^t\vec{x} = \vec{e}\},$$

con $A \in \Re^{n\times m}$ de rango completo igual a m.

En este caso el algoritmo de Newton converge en una iteración ya que

$$\nabla F(\vec{x}^*) = \nabla F(\vec{x}_0) + \alpha G d_0$$

por ser F cuadrática. Dado $\vec{x}_0 \in \Omega$, la solución exacta $(\vec{x}^*, \lambda)$ se obtiene al resolver

$$\begin{array}{rcl} Q_2^t\ G\ Q_2\ \vec{b}_0 & = & -Q_2^t\ \nabla F(\vec{x}_0), \\ \vec{x}^* & = & x_0 + Q_2\vec{b}_0, \end{array}$$

con $\nabla F(\vec{x}_0) = G\vec{x}_0 - \vec{f}$ y

$$R\vec{\lambda}^* = -Q_1^t \nabla F(\vec{x}^*).$$

2. Apliquemos lo anterior para determinar el punto en

$$\Omega = \{\vec{x} \in \Re^3 | 2x + 3y - z = 4; x - y - z = 1\}$$

cuya distancia al origen es mínima.

La función objetivo es $F(\vec{x}) = x^2 + y^2 + z^2$ y A^t es de la forma

$$A^t = \begin{pmatrix} 2 & 3 & -1 \\ 1 & -1 & -1 \end{pmatrix}$$

con rango 2, por lo que todos los puntos de Ω son regulares.

Al resolver el problema por multiplicadores de Lagrange se obtiene que $\vec{x}^* = \frac{1}{21}(19, 11, -13)$ y $\vec{\lambda} = (\frac{-4}{7}, \frac{-2}{3})$ es solución de las condiciones de primero y segundo orden por lo cual $\vec{x}^*$ es un mínimo de F restringido a Ω.

Apliquemos el algoritmo de Newton: sea $x_0 = (0, 3/4, -7/4)$, $\nabla F(\vec{x}_0) = G\vec{x}_0 = (0, \frac{3}{4}, \frac{-7}{4})$ y

$$EN(A^t) = \{\vec{x} \in \Re^3 | \vec{x} = (-4t, t, -5t) \quad t \in \Re\}.$$

Sea $Z^t = (-4, 1, -5)$ entonces $b_0 \in \Re$ debe satisfacer

$$Z^t G Z b_0 = Z^t \nabla F(\vec{x}_0);$$

por lo que $b_0 = -19/84$ y

$$\vec{d}_0 = Zb_0 = \begin{pmatrix} 0.904761905 \\ -0.226190476 \\ 1.130952381 \end{pmatrix}.$$

Entonces

$$\vec{x}^* = \vec{x}_0 + \vec{d}_0 = (0.904761905, 0.523809524, -0.619047619).$$

Por otro lado,

$$A^t A \vec{\lambda} = -A^t \nabla F(x_1) \Rightarrow \vec{\lambda} = \begin{pmatrix} -0.571428571 \\ -0.666666667 \end{pmatrix}$$

que coincide con la solución obtenida analíticamente.

La factorización QR de la matriz A es

$$Q = \begin{pmatrix} -0.534522484 & 0.577350269 & 0.6172134 \\ -0.801783726 & -0.577350269 & -0.15430335 \\ 0.267261242 & -0.577350269 & 0.77151675 \end{pmatrix}$$

$$R = \begin{pmatrix} -3.741657387 & 1.11022E-16 \\ 0 & 1.732050808 \\ 0 & 1.1395E-16 \end{pmatrix}$$

$$Q_1 = \begin{pmatrix} -0.534522484 & 0.577350269 \\ -0.801783726 & -0.577350269 \\ 0.267261242 & -0.577350269 \end{pmatrix}$$

y $Q_2^t = (0.6172134, -0.15430335, 0.77151675)$. En este caso $b_0 = 1.465881825$ y $d_0 = (0.904761905, -0.226190476, 1.130952381)$ que coincide con los cálculos anteriores.

3.2. Caso de restricciones de desigualdad

Consideremos que tenemos el siguiente problema

$$\min_{\vec{x}\in\Omega} F(\vec{x}),$$

con

$$\Omega = \{\vec{x} \in \Re^n | \vec{h}(\vec{x}) = A^t\vec{x} - \vec{e} \leq 0\}$$

un subconjunto distinto del vacío de $\Re^n$. Supongamos que el problema admite una solución.

El algoritmo de gradiente proyectado puede generalizarse para el caso de restricciones de desigualdad. Dado un punto $\vec{x} \in \Omega$, denotemos por $I(\vec{x})$ el conjunto de índices asociados a las restricciones activas, supongamos que q es la cardinalidad de $I(\vec{x})$. Sea A_q la matriz $\Re^{n\times q}$ cuyas columnas son los gradientes de las restricciones activas. Si el rango de A_q es igual a q puede factorizarse de la forma

$$A_q = Q_q\widehat{R}_q = [Q_q^1, Q_q^2]\begin{pmatrix} R_q \\ 0 \end{pmatrix}$$

y Q_q^2 es una base para $EN(A_q^t)$.

Uno de los aspectos que se requiere modificar para adaptar el algoritmo proyectado cuando se tiene restricciones de desigualdad es que dada una dirección admisible $\vec{d}_k$, el punto

$$\vec{x}_k + \alpha\vec{d}_k$$

puede no estar en Ω para todo valor de α ya que las restricciones pasivas pueden ser violadas dado que en el cálculo de $\vec{d}_k$ sólo se toma en cuenta las restricciones activas. Para garantizar que esto no sucede se calcula, para cada una de ellas, para cuál valor de α

$$h_j(\vec{x}_k + \alpha\vec{d}_k) = h_j(\vec{x}_k) + \alpha A_j^t\vec{d}_k = 0$$

y se selecciona el valor más pequeño para toda j; es decir si

$$\beta_t = \min_{i\notin I(x_k)}\{\beta_i = \frac{-h_j(\vec{x}_k)}{A_j^t\vec{d}_k}\},$$

y $\beta_t \neq 0$, entonces se selecciona $\alpha_k = \beta_t$; si además $F(x_{k+1}) < F(x_k)$ entonces la restricción t se vuelve activa por lo cual debe incluirse en la matriz A y, en consecuencia, actualizarse la factorización QR antes de checar si se cumple la condición de paro. Si $\alpha_k = 0$ no hay puntos admisibles que se puedan obtener a partir de la dirección admisible por lo que $\vec{x}_{k+1} = \vec{x}_k$ y $F_{k+1} = F_k$ y $\vec{g}_{k+1} = \vec{g}_k$.

Algoritmo de gradiente proyectado con restricciones de desigualdad

1. Dado $x_0 \in \Omega$, sea $I(x_0)$ y $A_{q,0}$ matriz de rango completo con factorización QR que denotaremos por $Q_{q,0}\widehat{R}_{q,0}$.

   Para $k = 1, 2, \ldots$ determine:

2. Calcular la dirección admisible $\vec{d}_k$ por

$$\begin{aligned} Q_{q,k}^{2\ t} H_F(\vec{x}_k) Q_{q,k}^2\ \vec{b}_k &= -Q_{q,k}^{2\ t} \vec{g}_k, \\ \vec{d}_k &= Q_{q,k}^2 \vec{b}_k. \end{aligned}$$

3. Para seleccionar α_k se hace lo siguiente: sea

$$\beta_t = \min_{j \notin I(x_k)} \{\beta_j = -\frac{h_j(\vec{x}_k)}{\vec{A}_j^t d_k}\},$$

   $\Rightarrow \alpha_k = \beta_t$ y

$$\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{d}_k.$$

   a) Si $\alpha_k = 0$ entonces $\vec{x}_{k+1} = \vec{x}_k$, $F_{k+1} = F_k$ y $\vec{g}_{k+1} = \vec{g}_k$ y es necesario quitar alguna de las restricciones activas. Ir al Paso 6.

   b) Si $\alpha_k \neq 0$ entonces $F_{k+1} = F(\vec{x}_{k+1})$ y $\vec{g}_{k+1} = \nabla F(\vec{x}_{k+1})$.

4. Si $\|Q_{k+1}^{2,t} \vec{g}_{k+1}\| < rtol$ y $\frac{\|\vec{x}_{k+1} - \vec{x}_k\|}{\|\vec{x}_k\|} \leq rtol \Rightarrow$

$$R_{k+1} \vec{\lambda}_{k+1} = -Q_{k+1}^{1,t} \vec{g}_{k+1} \tag{3.10}$$

   y $\vec{x}^* \approx \vec{x}_{k+1}$, $\lambda \approx \lambda_{k+1}$. Se detiene el algoritmo.

5. Si no se satisface el criterio de paro regresar al Paso 2 a calcular una nueva dirección.

6. (Paso para quitar restricciones) Determinar los multiplicadores de Lagrange para el punto $\vec{x}_k$

$$R_{q,k}\vec{\lambda}_{q,k} = -Q^{1,t}_{q,k}\vec{g}_k.$$

Determinar la componente negativa t de $\vec{\lambda}_{q,k}$ que cumple

$$(\lambda_{q,k})_t = \min\{(\lambda_{q,k})_i < 0\}$$

y a la matriz A se le quita la t-ésima columna para obtener

$$A_{q-1,k} = [a_1....a_{t-1}a_t...a_q].$$

Se factoriza de la forma

$$A_{q-1,k} = [Q^1_{q-1,k}, Q^2_{q-1,k}]\begin{pmatrix} R_{q-1,k} \\ 0 \end{pmatrix},$$

con $Z_{q-1,k} = Q^2_{q-1,k}$; se regresa al Paso 2.

7. (Paso para incluir restricciones) Sea t a nueva restricción activa, entonces $A_{q+1,k+1} = [A_{q,k}\ a_t]$ y al factorizarla de la forma QR se obtiene

$$A_{q+1,k+1} = [Q^1_{q+1,k+1}, Q^2_{q+1,k+1}]\begin{pmatrix} R_{q+1,k+1} \\ 0 \end{pmatrix}.$$

Calcular nueva dirección. Ir al Paso 4.

Ejemplo

1. Consideremos el siguiente problema del portafolio sin ventas en corto

$$\begin{array}{rrl} \text{Min} & \frac{1}{2}\vec{w}^t\ [\Sigma]\ w, & \\ \text{sujeto a} & h_1(\vec{w}) = & 0.2w_1 + 0.25w_2 + 0.15w_3 = .24, \\ & h_2(\vec{w}) = & w_1 + w_2 + w_3 = 1, \\ & -w_i \leq & 0 \qquad i = 1, \ldots, 3 \end{array}$$

con

$$[\Sigma] = \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.18 & 0 \\ 0 & 0 & 0.15 \end{pmatrix}.$$

La solución de este problemas es $\vec{w}^* = (.2, .8, 0)$. Seleccionamos como $\vec{w}_0$ a $(0.1, 0.85, 0.05)$. En este caso $\vec{g}_0 = (0.02, 0.153, 0.0075)$,

$$A_0 = \begin{pmatrix} 0.2 & 1 \\ 0.25 & 1 \\ 0.15 & 1 \end{pmatrix},$$

$$EN(A^t) = \{\vec{y} \in \Re^3 | .2y_1 + .25y_2 + .15y_3 = 0;\ y_1 + y_2 + y_3 = 1\}$$

y $\vec{Z}_0^t = (1, -.5, -.5)$. Como $|\vec{Z}_0^t \vec{g}_0| = 0.06025$ generamos la dirección $\vec{d}_0$ al resolver

$$\vec{Z}_0^t [\Sigma] \vec{Z}_0 b_0 = -\vec{Z}_0^t \vec{g}_0$$

cuya solución es $b_0 = 0.213274$ y

$$\vec{d}_0 = b_0 \vec{Z}_0 = \begin{pmatrix} 0.213274 \\ -0.106637 \\ -0.106637 \end{pmatrix}.$$

Para determinar el valor de α_0 determinamos

$$\beta = \min_{\beta} \{0.1 + 0.2132\beta_1,\ 0.85 - 0.106637\beta_2,\ 0.05 - 0.106637\beta_3\}$$

cuyo valor es $\beta_3 = 0.468880$. Por lo que $\alpha_0 = 0.468880$ y $\vec{x}_1 = (0.2, 0.8, 0)$. Como $F(x_1) = 0.0616 < F(x_0) = .06621$ entonces se procede a calcular el multiplicador de Lagrange, pero antes se actualiza la matriz A_1 a

$$A_1 = \begin{pmatrix} 0.2 & 1 & 0 \\ 0.25 & 1 & 0 \\ 0.15 & 1 & -1 \end{pmatrix},$$

y resolvemos $A_1 \vec{\lambda}_1 = -\vec{g}_1$ con $\vec{g}_1 = (0.04, 0.144, 0)$. Entonces

$$\vec{\lambda} = (-2.08, 0.376, 0.064)$$

y como $\lambda_3 > 0$ el algoritmo se detiene por haber encontrado la solución óptima.

2. Apliquemos el algoritmo a una función objetivo que no sea cuadrática.

$$\begin{aligned} \min F(x, y) \quad &= \quad \frac{1}{xy} \\ \text{sujeta a} \quad & \quad x + y = 2 \\ & \quad x, y > 0 \end{aligned}$$

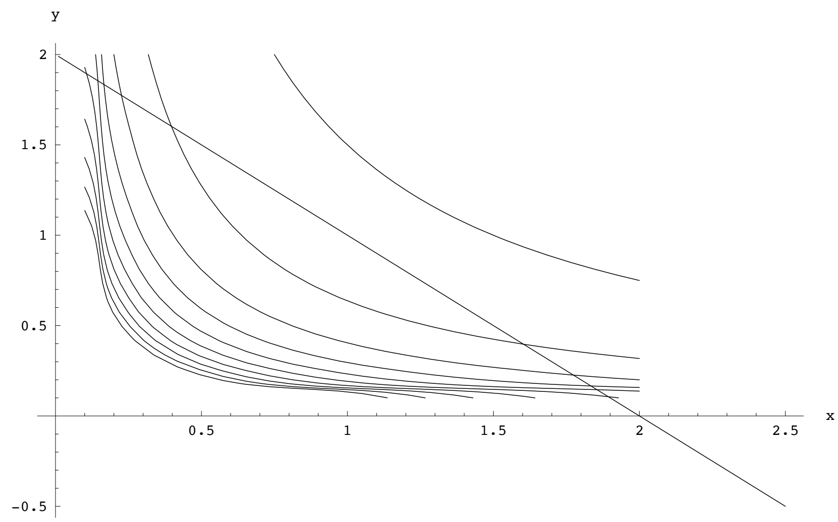

Figura 3.1: Región factible Ω con las curvas de nivel de F.

En la Figura 3.1 se presenta Ω y las curvas de nivel de la función F; un candidato a ser punto extremo de F en Ω es $(1,1)$ por estar en la curva de nivel $F(x,y)=1$. Las otras curvas de nivel que cortan al segmento de recta corresponden a valores mayores.

Dado $\vec{x}_0 = (\frac{3}{2}, \frac{1}{2})$, se tiene que $\vec{g}_0 = (-\frac{16}{9}, -\frac{8}{3})$ y

$$H_F(3/2, 1/2) = \begin{pmatrix} \frac{32}{27} & \frac{16}{9} \\ \frac{16}{9} & \frac{32}{3} \end{pmatrix}.$$

En este caso el plano tangente en $\vec{x}_0$ es

$$N((3/2,1/2)) = \{\vec{y} \in \Re^2 | \ y_1 + y_2 = 0\}$$

Una base de $N((3/2,1/2))$ es $\{(1,-1)\}$.

Aplicando el algoritmo anterior se obtiene

$$b_0 = -3/28 \quad \vec{d}_0 = (-3/28, 3/28).$$

Para determinar el tamaño de α no se puede aplicar el paso 3 del algoritmo porque la función F no está definida cuando alguna de las coordenadas se anula. En este caso se aplica búsqueda lineal exacta determinando el α que satisface:

$$\frac{dF(\vec{x}_0 + \alpha\vec{d}_0)}{d\alpha} = \nabla F(\frac{3}{2} - \frac{3}{28}\alpha, \frac{1}{2} + \frac{3}{28}\alpha)^t \vec{d}_0 = 0.$$

Al resolver se obtiene $\alpha = 14/3$ por lo que $\vec{x}_1 = (1,1)$. Por último, para checar las condiciones de optimalidad se calcula $(1,-1)^t\vec{g}_1$, como es igual a cero, $(1,1)$ es el mínimo de F restringido a Ω.

3.3. Método de Frank-Wolfe

El método de Frank-Wolfe o de gradiente reducido es un método propuesto por Marguerite Frank and Phil Wolfe in 1956. Es un método iterativo para problemas de optimización no lineales con restricciones lineales y fue inicialmente concebido para el caso de funciones objetivo cuadráticas. El método consiste en lo siguiente: en cada iteración se aproxima la función objetivo por una función cuadrática y se obtienen las condiones de Kuhn-Tucker respectivas. A partir de estos elementos se construye un problema de programación lineal en el que las condiciones de Kuhn-Tucker aparecen como restricciones y la función objetivo a minimizar es una combinación lineal de variables artificiales, cuyo valor mínimo se obtiene cuando éstas son cero. El método no es competitivo respecto a otros métodos más sofisticados, pero sigue siendo muy usado en problemas de muchas dimensiones como los que aparecen en los problemas de asignación de tráfico. Ilustremos este algoritmo con el siguiente problema cuadrático.

Sean $A \in \Re^{m\times n}$ y $H \in \Re^{n\times n}$, determinar $\vec{x} \in \Re^n$ y $\vec{\lambda} \in \Re^m$ tal que

$$\begin{aligned} \text{Min}\{F(\vec{x}) &= \tfrac{1}{2}\vec{x}^t H\vec{x} - \vec{c}^t\vec{x}\} \\ \text{sujeto a}: A^t\vec{x} &= b, \\ \vec{x} \geq 0. & \end{aligned}$$

Las condiciones de Kuhn-Tucker correspondientes son:

$$\begin{aligned} H\vec{x} + A\lambda &= \vec{c}, \\ A^t\vec{x} &= b, \\ \vec{x} &\geq 0. \end{aligned}$$

El problema de programación lineal asociado se construye de la siguiente forma: las condiciones de Kuhn y Tucker se imponen como restricciones y de esta manera una solución admisible automáticamente cumple las condiciones de Kuhn y Tucker. Por otro lado, se usa el procedimiento de variables artificiales de programación lineal para encontrar una solución admisible. Se introducen tantas variables artificiales y_i, no negativas, como restricciones de igualdad se tengan: $n+m$; por último, se propone como función objetivo la suma de las y_i. En suma el problema a minimizar es

$$\begin{aligned} Min \sum_{i=1}^{n+m} y_i \\ \text{sujeto a } H\vec{x} + A\lambda + I_1\vec{y} &= \vec{c}, \\ A^t\vec{x} + I_2\vec{y} &= b \\ x_i, y_i &\geq 0, \end{aligned}$$

con $I_1 \in \Re^{n\times m+n}$ y $I_2 \in \Re^{m\times m+n}$ matrices por bloques de la forma $I_1 = [I_{n\times n}\ 0]$ y $I_2 = [0\ I_{m\times m}]$.

Ejemplo

En el caso del problema del portafolio de optimización sin ventas en corto el problema de programación lineal asociado es

$$\begin{aligned} &Min\ \textstyle\sum_{i=1}^{n+2} y_i \\ \text{sujeto a}\quad &\Sigma w + \lambda \overrightarrow{r} + \mu \overrightarrow{1} + \textstyle\sum_{i=1}^{n} (y_i - \nu_i) \overrightarrow{e}_i = 0, \\ &w^t \overrightarrow{r} + y_{n+1} = r^*, \\ &\overrightarrow{1}^t w + y_{n+2} = 1, \\ &w_i, \nu_i, y_i \geq 0. \end{aligned}$$

Se puede usar cualquier software para resolver este problema. En particular Excel tiene el módulo Solver, Mathematica y Matlab tienen subrutinas específicas de programación lineal.

3.4. Ejercicios

1. Resuelva por el método de Frank-Wolfe el problema del portafolio sin ventas en corto que aparece en la sección 2.3.

2. Resuelva por el método de Frank-Wolfe el ejercicio 11 del capítulo 2.

3. Resuelva el ejercicio 2 por medio del algoritmo de gradiente proyectado y compare sus resultados con los obtenidos en el ejercicio anterior.

4. Adapte el algoritmo de gradiente proyectado para resolver el problema de programación convexa que aparece en el ejercicio 10 del capítulo 2.

Bibliografía

[1] Bazaraa M. and Sherali H. Nonlinear programming: Theory and algorithms. Wiley. Third Edition. 2006.

[2] Dennis J. E. and Schnabel R. Numerical methods for unconstrained optimization and nonlinear equations. Classic in Applied mathematics 16. SIAM. 1996.

[3] Gill, Murray & Saunders. Practical Optimization. Academic Press. 1981.

[4] R. Fletcher. Practical Methods of Optimization. Wiley 1987.

[5] Diego Bricio Hernández.Ciencia y Matemáticas: los modelos de predicción. Ciencia 31, 103-121. 1980.

[6] D. Luenberger. Programación Lineal y no Lineal. Addison Wesley - Editorial Iberoamericana. 1989. (2 edición).

[7] Mark Meerschaert. Mathematical Modelling. Academic Press. 1993.

[8] J. Mathews y K. Fink. Métodos Numéricos con Mathlab. Tercera edición.Pearson. Prentice Hall. 2007.

[9] Jorge Nocedal y Stephen J. Wright. Numerical Optimization. Second Edition. Springer. 2000.

[10] Peressini A., Sullivan F. y Uhl J.J. The mathematics of Nonlinear problems. Springer. 2000.

[11] L.E. Scales Int. to Non linear Optimization. Springer Verlag. 1985

[12] Gilbert Strang. Algebra lineal y sus aplicaciones. Addison-Wesley Iberoamericana. 1986.

[13] Sundaram Rangarajan. A First Course in Optimization theory. Cambridge university Press. 1996.