Introducción a la optimización no lineal

Patricia Saavedra Barrera [1]

7 de febrero de 2011

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

Índice general

Prólogo

Esta obra está diseñada para ser una introducción a la optimización no lineal, 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 o geométrica.

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. El libro está dividido en dos partes: en la primera se estudia la optimización no lineal sin restricciones y en la segunda parte se trata la optimización no lineal con restricciones.

En el primer capítulo se presentan algunos problemas de optimización y su modelación matemática; en el capítulo 2 se dan las condiciones necesarias y suficientes para tener un punto mínimo o máximo. En el capítulo 3, se presentan los algoritmos más importantes para aproximar estos puntos. En el capítulo 4 se dan condiciones necesarias y suficientes para que un problema de optimización no lineal con restricciones de igualdad y desigualdad no lineales admita una solución, las llamdas condiciones de Kuhn-Tucker. Por último, en el capítulo 5 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 lineal con restricciones lineales en un problema de programación lineal. Cada capítulo cuenta al final con una lista de ejercicios.

Esta obra está dirigida tanto para alumnos del último año de la licenciatura como para estudiantes de posgrado. En un curso introductorio a nivel de licenciatura se reduciría el material a cubrir; por ejemplo, se estudirían los

aspectos teóricos de optimización con y sin restricciones y sólo se analizarían los aspectos numéricos de los métodos para optimización sin restricciones. En el caso de posgrado se incluiría el tema de mínimos cduadrados no lineales y el método de gradiente proyectado para resolver numéricamente los problemas no lineales con restricciones.

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 incorporan 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 nos permitan obtener resultados cualitativos y 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.

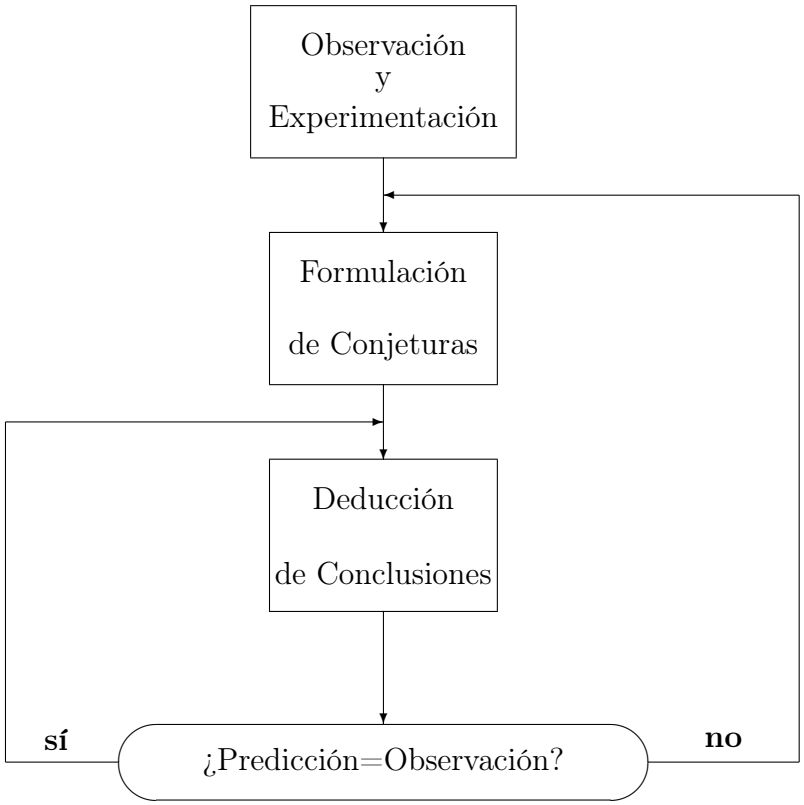

Figura 1.1. Diagrama estructural del método científico

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 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 infiramos, 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 ex-

presada 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 $(2')$ 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 del tiro parabólico; el probabilístico es aquel en el que hay incertidumbre y, por último, un modelo de optimización es aquel que busca 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 un número de horas debe 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, que consiste de 100 latas, de mango 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:

$$G(x, y, z) = 1000x + 900y + 850z.$$

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 está en cualquier lugar 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 sabemos que se requiere en cada una de ellas de un número fijo de horas 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

$$\begin{array}{rcl} \text{Max } G(x, y, z) & = & 1000x + 900y + 850z, \\ \text{sujeto a :} \quad x, y, z & \geq & 0, \\ 3x + 4y + 3z & \leq & 8, \\ 3x + 2y + 2.5z & \leq & 10, \\ 4x + 4y + 4z & \leq & 12. \end{array}$$

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 es de la forma:

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

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

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

Observemos que este es un problema de una función de $\Re^{n+1}$ a los reales. 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 son 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 se cumple si 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

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

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 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 matricial del problema anterior es

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

¿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:

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

La formulación matemática del problema fue idea de Markowitz, ganador del Premio Nobel de Economía 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

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

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:

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

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 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 386 AT. 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 20 % el precio de cada computadora, el efecto fue de un incremento del 50 % 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 maximiza las ganancias.

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 $1070 pesos, mientras que el modelo B tendrá un costo de $1350 pesos. Producir un televisor tipo A cuesta $585 pesos y del tipo B $675 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 0.4 centavos y cada vez que se vende un modelo B se reduce el precio del modelo A en 0.3 centavos. ¿Cuántas unidades de cada modelo deben producirse para maximizar la ganancia?

Capítulo 2

Optimización sin restricciones.

En este capítulo se presentan algunos resultados del cálculo diferencial de varias variables para problemas de optimización sin restricciones. Estos resultados se conocen con el nombre de condiciones de primero y segundo orden para la existencia de máximos o mínimos y se aplican a aquellos problemas en los que la función objetivo es diferenciable en un conjunto abierto S de $\Re^n$. Para aquellos lectores que les interesa estudiar el caso en que la función no es diferenciable se les recomienda consultar el libro de Fletcher [4].

2.1. Un problema de mínimos cuadrados

Según estudios médicos el número de cigarrillos que consume al año una persona incrementa el riesgo de que padezca de cancer pulmonar. Supongamos que se desea estimar el número de muertes que pueden ocurrir en la Ciudad de México por cancer pulmonar, dado que el promedio anual del consumo por persona de cigarrillos durante 1990 fue de 470. Los únicos datos que se tienen a la mano relacionan el consumo de cigarrillos, x_i, con el número de muertes por cáncer pulmonar, y_i, en los países escandinavos durante 1980. Aunque las condiciones de vida de esos países y las del nuestro son muy distintas, esos datos pueden darnos una estimación inicial.

Tabla 2.1

País	Consumo	Num. de muertes
Dinamarca	350	165
Finlandia	1100	350
Noruega	250	95
Suecia	300	120

¿Cómo se pueden usar estos datos? Al graficarlos, se observa que puede trazarse una recta que no diste mucho de ellos. Como lo que buscamos es una estimación y no el valor exacto, ¿por qué no construímos la recta $p(x) = a_1x + a_0$ que al evaluarla en cada x_i no diste mucho de y_i ?

Una manera de determinar la recta es buscar los coeficientes a_0 y a_1 que hagan que la suma de los cuadrados de la diferencia entre $p(x_i)$ y y_i sea lo más pequeña posible. Observemos que las incógnitas de nuestro problema son los coeficientes a_0 y a_1. Sea G la función que depende de a_0 y a_1, con la siguiente regla de correspondencia:

$$G(a_0, a_1) = \sum_{i=1}^{4} [y_i - a_1x_i - a_0]^2.$$

Entonces el problema a determinar puede escribirse como un problema de optimización de la forma siguiente: determinar el mínimo de G en $\Re^2$ que se denotara como

$$\min_{(a_0,a_1)\in\Re^2} G(a_0, a_1).$$

Este es un problema no lineal sin restricciones. que se conoce con el nombre de ajuste lineal por mínimos cuadrados.

Supongamos que podemos determinar una solución, entonces para estimar el número de muertes por cancer pulmonar en la Ciudad de México basta con evaluar la función $p(x)$ en 470. Esta no es la única manera de encontrar una estimación, nótese que se puede construir la función G de muy diversas maneras; por ejemplo, sea G_1 la suma del valor absoluto de $f(x_i) - y_i$, o sea

$$G_1(a_0, a_1) = \sum_{i=1}^{4} |y_i - a_1x_i - a_0|.$$

El problema de optimización correspondiente es más difícil que el anterior ya que G_1 no es diferenciable y no podemos usar cálculo para resolver el problema de optimización.

En este capítulo se estudiará la relación que existe entre la primera y segunda derivada y la existencia de puntos extremos.

2.2. Condiciones de primer orden

Veamos primero algunas definiciones que nos permitan hablar sin ambiguedad de lo que entendemos por un mínimo o por un máximo. De aquí en adelante trabajaremos con funciones f definidas en un subconjunto abierto S de $\Re^n$ con valores en los reales y supondremos que $f \in C^0(S)$ o sea es continua en S. Asimismo, denotemos con $\|\cdot\|$ la norma euclideana en $\Re^n$.

Definición 2.2.1. *Se dice que una función f tiene un punto mínimo global en S si existe una $\vec{x}^* \in S$ que satisface que*

$$f(\vec{x}^*) \leq f(\vec{x}) \qquad \forall\ \vec{x} \in S.$$

Definición 2.2.2. *Se dice que una función f tiene un punto máximo global en S si existe $\vec{x}^* \in S$ tal que*

$$f(\vec{x}) \leq f(\vec{x}^*) \qquad \forall \vec{x} \in S.$$

Definición 2.2.3. *Se dice que f tiene un mínimo local en S si existe una $\delta > 0$ tal que*

$$f(\vec{x}) \geq f(\vec{x}^*) \qquad \forall \vec{x} \in V_\delta(\vec{x}^*),$$

con $V_\delta(\vec{x}^*) = \{\vec{x} \in S |\ ||\vec{x} - \vec{x}^*|| < \delta\}$. De la misma forma se define un máximo local.

Definición 2.2.4. *Diremos que f admite un punto extremo en S si f tiene un mínimo o un máximo local en S.*

Las definiciones anteriores precisan lo que entendemos por un punto extremo pero no nos indican un procedimiento para asegurar su existencia o un procedimiento para encontrarlo.

Si f es continua en un conjunto compacto de $\Re^n$ el Teorema de Weierstrass nos garantiza que f alcanza sus valores extremos en el compacto. Si f es diferenciable en un abierto de $S \subset \Re^n$ podemos usar cálculo de varias variables para determinar los puntos extremos. En caso de que la función no sea diferenciable pero si continua, existen otros métodos para determinar los

puntos extremos. En los últimos años han aparecido varios algoritmos heurísticos como los algoritmos genéticos que no requieren el cálculo del gradiente y que han dado buenos resultados.

Definición 2.2.5. *Supongamos que f es diferenciable en S, un abierto de $\Re^n$, diremos que f admite un punto crítico $\vec{x}^*$ en S si*

$$\frac{\partial f(\vec{x}^*)}{\partial x_j} = 0 \qquad j = 1, \ldots, n.$$

El siguiente teorema nos dice que si f es diferenciable en S y tiene un punto extremo en $\vec{x}^* \in S$, necesariamente éste debe ser un punto crítico.

Teorema 2.2.6. *Supongamos que f es continuamente diferenciable en S, un abierto de $\Re^n$, y que tiene un punto extremo en $\vec{x}^* \in S \Rightarrow$*

$$\frac{\partial f(\vec{x}^*)}{\partial x_j} = 0 \qquad j = 1, \ldots, n.$$

Para cada j, definamos la función $\hat{f}_j$:$\Re \to \Re$ tal que

$$\hat{f}_j(t) = f(\vec{x}^* + t\vec{e_j}),$$

con $\vec{e_j}$ el j-ésimo vector de la base canónica de $\Re^n$. $\hat{f}_j$ es una función continua y diferenciable en $\Re$ ya que

$$\frac{d\hat{f}_j(t)}{dt} = \nabla f(\vec{x}^* + t\vec{e_j}) \cdot \vec{e_j} = \frac{\partial f(\vec{x}^* + t\vec{e_j})}{\partial x_j}.$$

Dado que $\vec{x}^*$ es un punto extremo de f, entonces $\hat{f}_j$ restringido a la recta $\vec{x}^* + t\vec{e_j}$ también alcanza un valor extremo en el mismo punto, o sea cuando $t = 0$. Por lo tanto la derivada de $\hat{f}_j$ debe anularse en $t = 0$, es decir

$$0 = \frac{d\hat{f}_j(0)}{dt} = \frac{\partial f(\vec{x}^*)}{\partial x_j}.$$

Como para cada j se cumple lo anterior, se concluye que el gradiente de f en x^* es igual al vector cero. □

Todo punto extremo de una función diferenciable es un punto crítico, pero no viceversa. Al teorema anterior se le conoce con el nombre de condiciones de primer orden. Para poder garantizar que un punto crítico es un mínimo o un máximo se requiere utilizar la información del Hessiano de la función.

2.3. Condiciones de segundo orden

Supongamos en esta sección que $f \in C^2(S)$, la matriz Hessiana de f evaluada en un punto $\vec{x}$ en $\Re^n$ es una matriz de $n \times n$ de la forma

$$H_f(\vec{x}) = \begin{pmatrix} \frac{\partial^2 f(\vec{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\vec{x})}{\partial x_1 x_n} \\ \frac{\partial^2 f(\vec{x})}{\partial x_2 x_1} & \cdots & \frac{\partial^2 f(\vec{x})}{\partial x_2 x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\vec{x})}{\partial x_n x_1} & \cdots & \frac{\partial^2 f(\vec{x})}{\partial x_n^2} \end{pmatrix}.$$

Como $f \in C^2(S)$ entonces las parciales cruzadas son iguales por lo que el Hessiano es simétrico: $H_f^t = H_f$. Denotaremos como A^t la matriz transpuesta de A.

Definición 2.3.1. *Una matriz $H \in \Re^{n\times n}$ es positiva semidefinida si*

$$\vec{x}^t H \vec{x} \geq 0 \qquad \forall \vec{x} \in \Re^n,$$

y es negativa semidefinida si

$$\vec{x}^t H \vec{x} \leq 0 \qquad \forall \vec{x} \in \Re^n. \tag{2.1}$$

Definición 2.3.2. *Una matriz $H \in \Re^{n\times n}$ es positiva definida si*

$$\vec{x}^t H \vec{x} > 0 \qquad \forall \vec{x} \in \Re^n,\ \vec{x} \neq 0$$

y es negativa definida si se cumple en (2.1) la desigualdad estricta.

Teorema 2.3.3. *Sea $f : S \subset R^n \to \Re$, S abierto y $f \in C^2(S)$. Si $\vec{x}^*$ es un punto crítico de f en S y si el Hessiano de f evaluado en $\vec{x}^*$: $H_f(x^*)$, es positiva definida $\Rightarrow$ f tiene un mínimo local en $\vec{x}^*$.*

Como $f \in C^2(S)$ y $\vec{x}^* \in S$ entonces, por el Teorema de Taylor, para todo $\vec{x} \in S$, $f(\vec{x})$ se puede escribir de la forma

$$f(x) = f(\vec{x}^*) + (\vec{x} - \vec{x}^*)^t \nabla f(\vec{x}^*) + \frac{1}{2}(\vec{x} - \vec{x}^*)^t\ H_f(\eta)\ (\vec{x} - \vec{x}^*),$$

con $\eta \in S$. Como $\vec{x}^*$ es un punto crítico de f, por el teorema anterior $\nabla f(\vec{x}^*) = 0$ por lo que

$$f(\vec{x}) = f(\vec{x}^*) + \frac{1}{2}\ (\vec{x} - \vec{x}^*)^t H_f(\eta)\ (\vec{x} - \vec{x}^*). \tag{2.2}$$

Dado que $f \in C^2(S)$ entonces $H_f(\vec{x}^*)$ es continua y además por hipótesis es positiva definida, así que se garantiza la existencia de una $\delta > 0$ tal que $H_f(\vec{x})$ es semidefinida positiva para toda $x \in V_\delta(\vec{x}^*)$. Por la igualdad (2.2) podemos asegurar que todos los elementos de esta vecindad satisfacen que $f(\vec{x}) \geq f(\vec{x}^*)$. Por lo tanto $\vec{x}^*$ es un mínimo local de $f \in S$. □

Teorema 2.3.4. *Sea $f : S \subset R^n \to \Re$, S abierto y $f \in C^2(S)$. Sea $\vec{x}^*$ un punto de S y supóngase que $\vec{x}^*$ es un mínimo local de f en S. Entonces $H_f(x^*)$ es una matriz semidefinida positiva.*

Ver demostración en el Luenberger [6]

Para el caso de un máximo se tienen resultados similares, basta con cambiar, donde proceda, la hipótesis de que la matriz hessiana sea positiva definida por negativa definida. En la siguiente sección se aplican los resultados anteriores a problemas cuadráticos.

2.4. Funciones cuadráticas

Las funciones cuadráticas son las funciones no lineales más sencillas. Su forma general, en expresión matricial, es la siguiente

$$f(\vec{x}) = \frac{1}{2}\ \vec{x}^t\ A\vec{x} - \vec{x}^t\vec{b} + c,$$

con $A \in \Re^{n\times n}$, $\vec{b} \in \Re^n$ y $c \in \Re$. Al problema de optimización correspondiente se le conoce con el nombre de programación cuadrática siempre que A sea simétrica. En este caso, la función objetivo siempre es una función dos veces continuamente diferenciable, por lo que

$$\nabla f(\vec{x}) = A\vec{x} - \vec{b}$$

y $H_f(x) = A$. Encontrar los puntos críticos de f es equivalente a resolver el siguiente sistema de ecuaciones lineales

$$A\vec{x} = \vec{b}. \tag{2.3}$$

Este problema admite una única solución si el determinante de A es distinto de cero, por lo que un problema cuadrático con matriz A invertible

admite un único punto crítico. La matriz hessiana: $H_f(X) = A$ es una matriz constante por lo tanto si es positiva definida se tiene un mínimo o si es negativa definida un máximo. Si A es indefinida se tienen un punto silla.

Demostrar que una matriz de cualquier tamaño es positiva definida a partir de la definición no es fácil. El siguiente teorema cuya demostración puede verse en el libro de Strang, ver [12], da otros criterios equivalentes para demostrar que una matriz simétrica es positiva definida.

Teorema 2.4.1. *Las siguientes proposiciones son equivalentes:*

i) La matriz A es simétrica y positiva definida.

ii) Los valores propios de la matriz A son reales y positivos.

iii) Los determinantes de los menores principales de A son positivos.

iv) En la eliminación de Gauss todos los pivotes, sin intercambio de renglones, son estrictamente positivos.

v) La matriz A admite una descomposición tipo Cholesky; es decir existe una matriz triangular superior S invertible tal que $A = S^tS$.

Como el Hessiano de una función f dos veces continuamente diferenciable, siempre es simétrico, el anterior teorema nos da varios procedimientos para probar que éste es positivo definido. Si los valores propios de $H_f(\vec{x}^*)$ son todos positivos entonces $\vec{x}^*$ es un mínimo. Si todos los valores propios son negativos entonces $\vec{x}^*$ es un máximo; por último, si hay valores propios positivos y negativos la matriz es indefinida y se tiene un punto silla. En el caso que alguno de los valores propios sea cero, entonces el mínimo local no es mínimo local estricto ya que si $\vec{x}^*$ es el mínimo también lo es $\vec{x}^* + \vec{q}$ para todo $\vec{q} \in \{\vec{x} \in \Re^n | A\vec{x} = 0\}$.

Encontrar los valores propios de una matriz es equivalente a encontrar las raíces de un polinomio de grado n, lo cual no puede hacerse en forma exacta cuando $n \geq 5$. Si el grado del polinomio es muy alto, el problema numérico asociado es un problema mal planteado, por lo general no garantiza que las raíces que encontremos estén cerca de las exactas por lo que no se recomienda en la práctica. El procedimiento de factorizar la matriz de la forma Cholesky es fácil de implementar computacionalmente, existen numerosos paquetes de computación que realizan esta factorización.

Ejemplos

1. Consideremos la función

$$F(x) = x_1^2 - 2x_1x_2 - \frac{1}{2}(x_2^2 - 1),$$

el punto crítico es $(x_1, x_2) = (0, 0)$. El Hessiano es

$$H_F = \begin{bmatrix} 2 & -2 \\ -2 & -1 \end{bmatrix}$$

cuyos valores propios son $\lambda_1 = -2$ y $\lambda_2 = 3$ por lo que la función tiene un punto silla en $(0, 0)$. Véase la Figura 2.1 para confirmar que las curvas de nivel son hiperbolas.

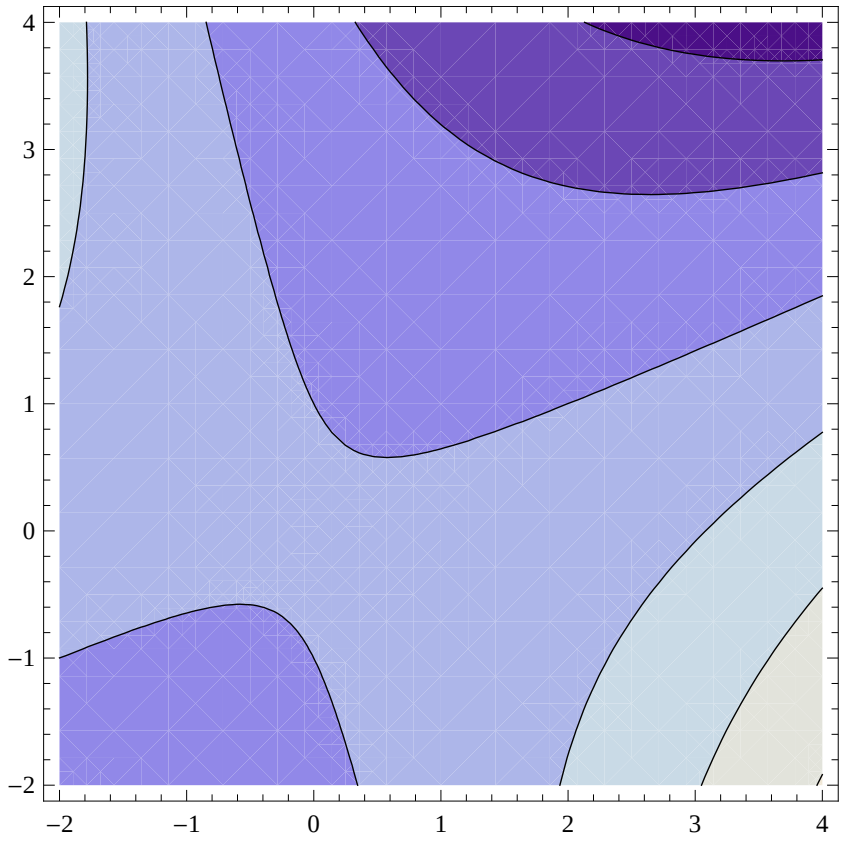

Figura 2.1: Curvas de nivel de $F(x) = x_1^2 - 2x_1x_2 - \frac{1}{2}(x_2^2 - 1)$.

2. Otro ejemplo de función cuadrática es

$$f(x_1, x_2) = \frac{1}{2}x_1^2 - 2x_1x_2 + 2x_2^2.$$

Esta función tiene una infinidad de puntos críticos. De hecho $EN(A) = \{(x_1, x_2)|x_1 = 2x_2\}$. En este caso los valores propios del Hessiano de f son $\lambda_1 = 0$ y $\lambda_2 = 5$ por lo que H_f es una matriz positiva semidefinida. En este caso no se cumplen las hipótesis del Teorema 2.3.3 por lo que hay que analizar con más cuidado la función f. Observe que $f(x_1, x_2) = \frac{1}{2}(x_1 - 2x_2)^2$ por lo que alcanza el valor mínimo en todos los puntos críticos. La Figura 2.2 nos muestra que las curvas de nivel son rectas paralelas.

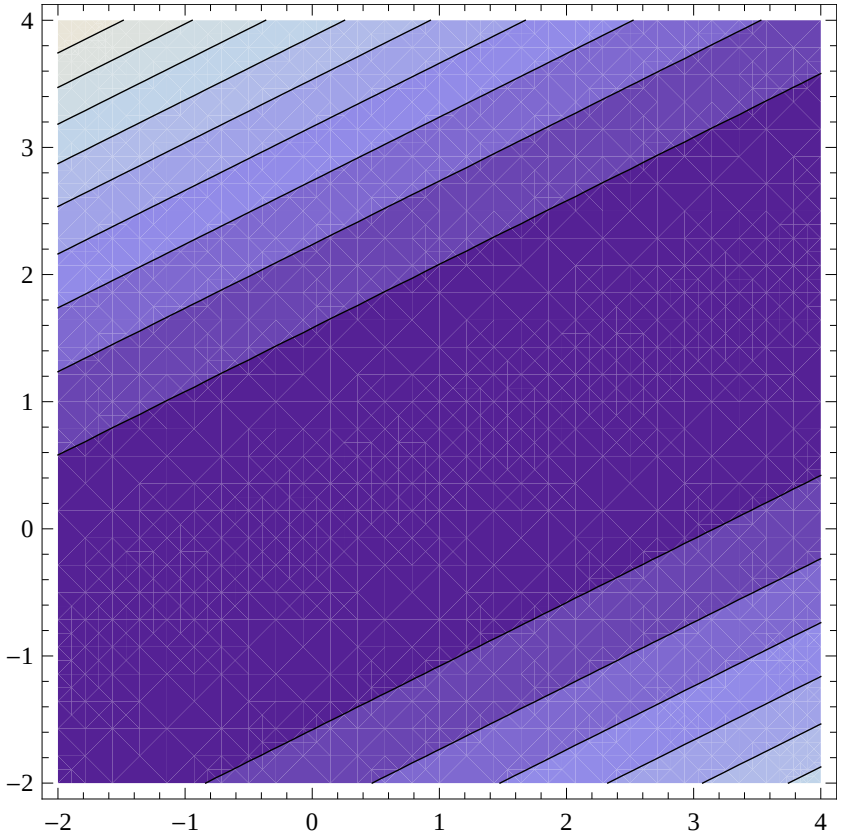

Figura 2.2: Curvas de nivel de $f(x) = \frac{1}{2}x_1^2 - 2x_1x_2 + 2x_2^2$.

3. El ejemplo 2.1 es un problema cuadrático que se expresa en forma matricial de la siguiente forma: sea $\vec{y}^t = [165, 350, 95, 120]$, $\vec{a}^t = [a_0, a_1]$ y

$$X = \begin{pmatrix} 1 & 350 \\ 1 & 1100 \\ 1 & 250 \\ 1 & 300 \end{pmatrix}$$

entonces la función G del ejemplo 3.1 se puede expresar sin pérdida de generalidad de la forma

$$\begin{aligned} G_(\vec{a}) &= \frac{1}{2}[\vec{y} - X\vec{a}]^t[\vec{y} - X\vec{a}] = \frac{1}{2}\sum_{i=1}^{4}[y_i - a_0x_i - a_1]^2, \\ &= \frac{1}{2}\vec{a}^t\ A\vec{a} - \vec{a}^t\vec{b} + c, \end{aligned}$$

con $A = X^t\ X$, $\vec{b} = X^t\vec{y}$ y $c = \frac{1}{2}\vec{y}^t\vec{y}$.

Para encontrar el punto crítico de G se resuelve el sistema

$$X^tX\vec{a} = X^t\vec{y} \tag{2.4}$$

que se conoce con el nombre de las ecuaciones normales y su solución es un mínimo siempre que X sea una matriz de rango completo, ver ejercicio 7.

Multiplicando X por su transpuesta obtenemos

$$X^t\ X = \begin{pmatrix} 4 & 2000 \\ 2000 & 1485000 \end{pmatrix}$$

y multiplicando al vector y por la transpuesta de la matriz X se obtiene el vector $\vec{b}^t = [730, 502500]$. Al resolver el sistema (2.4) se obtiene que $a_1 = 0.283503$ y $a_0 = 40.7475$. Para demostrar que este es el punto mínimo se comprueba que X^tX es positiva definida en $\Re^2$. Dado $(x, y) \in \Re^2$,

$$\begin{aligned} \vec{x}^t\ X^tX\vec{x} &= x^2 + 4000xy + 1485000y^2, \\ &= 4(x^2 + 1000xy + 250000y^2) - 1000000y^2 + 1485000y^2, \\ &= 4(x + 500)^2 + 485000y^2 \geq 0. \end{aligned}$$

Con los valores de a_0 y a_1 que determinamos, estamos en condiciones de estimar el número de muertes por cancer pulmonar en la Ciudad de México, basta evaluar la función $f(x)$ en 470.

$$\begin{aligned} f(x) &= 0.283505x + 40.7475, \\ f(470) &\approx 174. \end{aligned}$$

En la Figura 2.3 se muestran los datos y la recta que mejor los ajusta en el sentido de mínimos cuadrados.

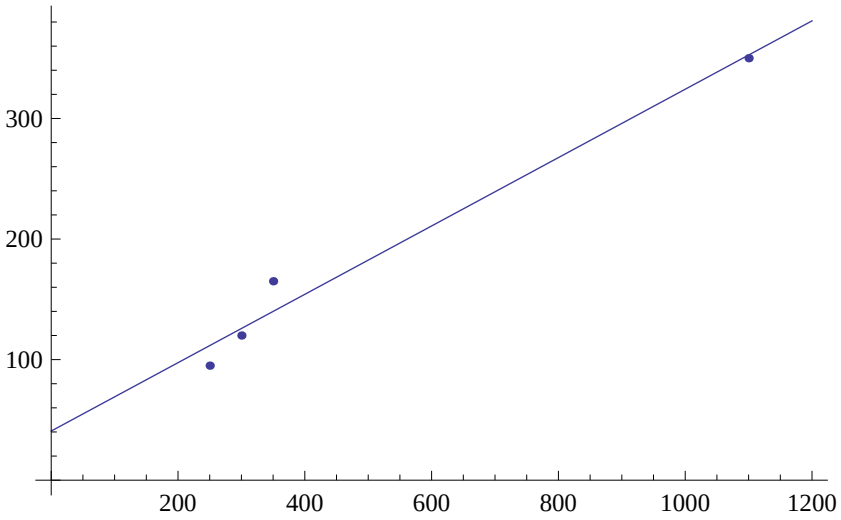

Figura 2.3: Ajuste por mínimos cuadrados

4. Apliquemos los resultados anteriores a la siguiente función que es la función objetivo del ejercicio 7 del capítulo anterior.

$$\begin{aligned} G(x,y) &= (1070 - .01x - .003y)x + (1350 - .004x - .01y)y \\ &\quad -(400000 + 585x + 675y). \end{aligned}$$

Derivando respecto a x y y e igualando a cero se obtiene el siguiente sistema de ecuaciones

$$\begin{aligned} \frac{\partial G(x)}{\partial x} &= -.02x - .007y + 465 = 0, \\ \frac{\partial G(x)}{\partial y} &= -.007x - .02y + 675 = 0. \end{aligned}$$

La solución es $x = 14173.79$ y $y = 28789.17$ La matriz Hessiana es

$$\begin{bmatrix} -.02 & -.007 \\ -.007 & -.02 \end{bmatrix}$$

cuyos valores propios son $\lambda_1 = -.027$ y $\lambda_2 = -0.013$. Por lo que la función tiene un máximo en su punto crítico. Observemos que en el problema original se tenían como restricciones que las variables fueran positivas, como esta condición se cumple para el máximo local entonces ésta es la solución del problema.

Las funciones cuadráticas son muy importantes en optimización. Dada su sencillez es fácil demostrar en su caso si un método numérico es convergente. Asimismo, permiten evaluar las ventajas de los métodos numéricos, si un método no converge para una función cuadrática, difícilmente lo hará para funciones más generales. Además, dada una función no lineal de clase C^2, en una vecindad del mínimo siempre se puede aproximar por medio de una función cuadrática que se obtiene al expandir la función alrededor de un punto en la vecindad del mínimo. Por el Teorema de Taylor se tiene que

$$F(\vec{x}) = F(\vec{x}_0) + (\vec{x} - \vec{x}_0)^t \nabla F(x_0) + \frac{1}{2}\ (\vec{x} - \vec{x}_0)^t\ H_F(\theta\vec{x}_0 + (1-\theta)\vec{x})\ (\vec{x} - \vec{x}_0).$$

Si definimos una función $\hat{F}$ como

$$\hat{F}(\vec{x}) = F(\vec{x}_0) + (\vec{x} - \vec{x}_0)^t \nabla F(x_0) + \frac{1}{2}\ (\vec{x} - \vec{x}_0)^t\ H_F(\vec{x}_0)\ (\vec{x} - \vec{x}_0),$$

es de esperarse que el mínimo de la función cuadrática $\hat{F}$ y de la original F estén "cerca". Por ello, la mayoría de los algoritmos se prueban para funciones cuadráticas, ya que cerca de la vecindad del mínimo nuestra función se comportará como una función cuadrática.

Para ilustrar el caso en que la función objetivo sea no lineal tomemos el caso de la función Shalow, llamada asi por su valle poco profundo y que fue introducida por Witte et al [2]

$$F(\vec{x}) = 100\ (x_1^2 - x_2)^2 + (1 - x_1)^2.$$

Como puede observarse la función desciende muy suavemente al punto mínimo que se alcanza en $(0, 0)$.

Observemos que la función $F(\vec{x})$ es siempre mayor o igual a cero y que alcanza un mínimo en $(1,1)$. Aplicando las condiciones de primero orden tenemos que

$$\begin{aligned}\frac{\partial F(\vec{x})}{\partial x_1} &= 400x_1^3 - 400x_1x_2 + 2x_1 - 2,\\ \frac{\partial F(\vec{x})}{\partial x_2} &= -200(x_1^2 - x_2).\end{aligned}$$

Igualando a cero ambas ecuaciones, se obtiene de la segunda que $x_1^2 = x_2$. Substituyendo en la primera ecuación nos da que $2x_1 - 2 = 0$ lo que implica que el único punto crítico es el $(1,1)$. Para demostrar que este punto es realmente un mínimo, aplicamos las condiciones de segundo orden. En este caso H_F no es una matriz constante, depende de x_1 y es de la forma

$$H_F(\vec{x}) = \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400 & 200 \end{bmatrix}$$

Evaluando el Hessiano en $(1,1)$ se tiene

$$H_F(1,1) = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}$$

cuyos valores propios son $\lambda_1 = 1001.6006$ y $\lambda_2 = 0.399360$. Por lo tanto como ambos valores propios son positivos, $(1,1)$ es el único punto mínimo de F. La Figura 2.4 nos muestra que el punto (1,1) se encuentra en un valle poco profundo, de ahí la dificultad para determinarlo numéricamente.

2.5. Mínimos globales

En muchas aplicaciones no sólo interesa determinar los mínimos locales sino el mínimo global. Existen algunos conjuntos de funciones para las cuales se puede asegurar que tienen al menos un mínimo global y estas funciones son las coercivas y las convexas.

2.5.1. Funciones coercivas

Definición 2.5.1. *Sea $F : \Re^n \to \Re$ continua, se dice que F es coerciva si*

$$\lim_{\|\vec{x}\|\to\infty} F(\vec{x}) = \infty.$$

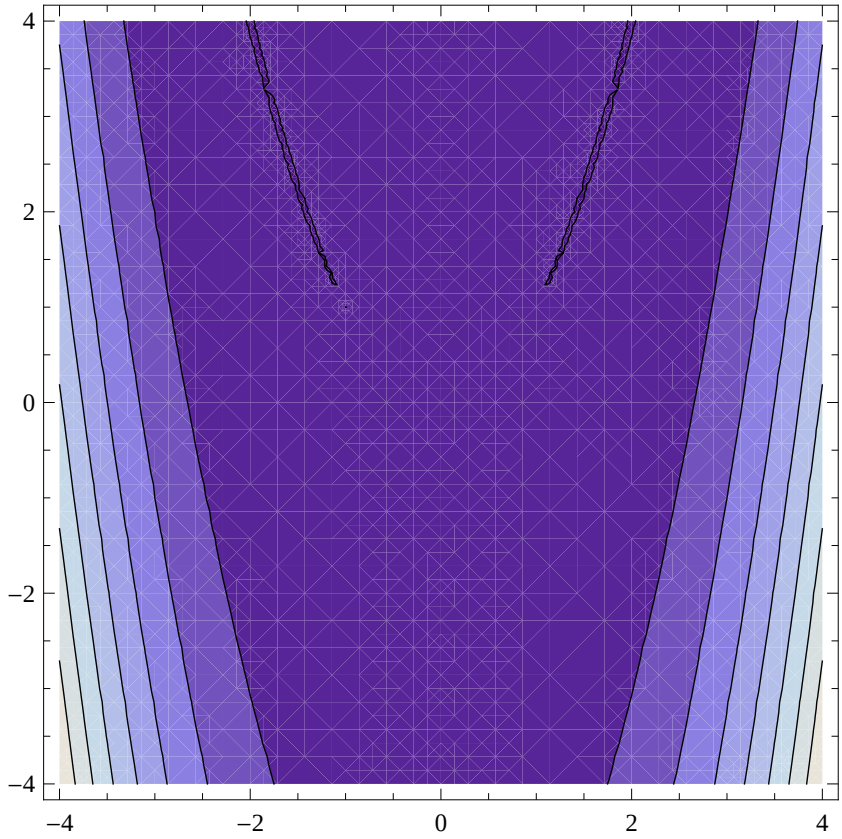

Figura 2.4: Curvas de nivel de la Función Shalow

Esta definición lo que nos dice es que dada una constante positiva M existe un número positivo R_M tal que $F(\vec{x}) \geq M$ cuando $\|\vec{x}\| \geq R_M$, o sea F no permanece acotado en un conjunto no acotado de $\Re^n$.

Ejemplos

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

$$\begin{aligned} f(x,y) &= \|\vec{x}\|^2, \\ \lim_{\|\vec{x}\|\to\infty} f(\vec{x}) &= \lim_{\|\vec{x}\|\to\infty} \|\vec{x}\|^2 = \infty. \end{aligned}$$

Por lo tanto f es coerciva.

2. Sea $f(x,y) = x^4 + y^4 - 3xy$.

$$f(x,y) = (x^4 + y^4)(1 - \frac{3xy}{x^4 + y^4}),$$
$$\lim_{\|\vec{x}\| \to \infty} f(\vec{x}) = \infty.$$

Por lo tanto f es coerciva.

3. Sea $f(x,y,z) = e^{x^2} + e^{y^2} + e^{z^2} - x^{100} - y^{100} - z^{100}$

$$f(x,y,z) = (e^{x^2} + e^{y^2} + e^{z^2})(1 - \frac{x^{100} + y^{100} + z^{100}}{e^{x^2} + e^{y^2} + e^{z^2}}),$$
$$\lim_{\|\vec{x}\| \to \infty} f(\vec{x}) = \infty.$$

f es también coerciva.

4. Las funciones lineales en $\Re^2$ no son coercivas ya que F es de la forma

$$f(x,y) = ax + by + c.$$

Si considero los puntos (x,y) tal que $ax + by = 0$ con $a \neq 0$ o $b \neq 0$ entonces $f(x,y) = c$, independientemente de los valores que tomen (x,y), por lo que no es coerciva.

5. $F(x,y) = x^2 - 2xy + y^2$

$$F(x,y) = (x-y)^2$$

y si consideramos el conjunto $S = \{(x,y) | x = y\}$ se tiene que $F(x,y) = 0$ para todo punto en S y S es un conjunto no acotado; por lo que F no es coerciva.

Teorema 2.5.2. *Sea $F : \Re^n \to \Re$ continua y coerciva $\Rightarrow$ F tiene al menos un mínimo global. Si $F \in C^1$ el mínimo global es punto crítico.*

Si F es continua y coerciva entonces

$$\lim_{\|\vec{x}\| \to \infty} F(\vec{x}) = \infty.$$

Sin pérdida general supongamos que $F(0) = M > 0$, entonces por ser F coerciva, existe una $r > 0$ tal que si $\|\vec{x}\| > r$ s cumple que $f(\vec{x}) > f(0)$. Sea

$$\overline{B}(0,r) = \{\vec{x} \in \Re^n |\ \|\vec{x}\| \leq r\}$$

como $\overline{B}(0,r)$ es un conjunto cerrado y acotado de $\Re^n$ entonces F por ser continua alcanza su valor mínimo en $\overline{B}(0,r)$. Por lo que existe $\vec{x}^* \in \overline{B}(0,r)$ tal que

$$F(\vec{x}^*) \leq F(\vec{x}) \quad \forall \vec{x} \in \overline{B}(0,r).$$

Observemos que $0 \in \overline{B}(0,r)$ por lo que $F(\vec{x}^*) \leq F(0)$. Si seleccionamos una $\vec{x}$ que no esté en $\overline{B}(0,r)$ entonces $F(\vec{x}) > F(0) \geq F(\vec{x}^*) \Rightarrow F(\vec{x}) \geq F(\vec{x}^*)$ por lo que $\vec{x}^*$ es un mínimo global.

La segunda parte del teorema se demuestra aplicando las condiciones de primer orden. □

2.5.2. Funciones convexas

Definición 2.5.3. *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 2.5.4. *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 2.5.5. *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 2.5.6. *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 2.5.7. *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 2.5.8. *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.*

2.6. Características generales de los algoritmos de descenso

Los ejemplos vistos hasta ahora tienen la cualidad que sus puntos extremos se determinan en forma exacta, pero éste no es siempre el caso. Por lo general, los problemas son no-lineales y tienen un gran número de variables. Determinar los puntos críticos equivale a resolver un sistema de ecuaciones no lineales cuya solución debe, en la mayoría de los casos, aproximarse. Dado que lo que nos interesa es determinar los extremos de una función, ¿por qué no utilizar un algoritmo apropiado para este tipo de problemas, que tome en cuenta que el punto que se busca es, por ejemplo, un mínimo.

Los algoritmos que generan una sucesión de puntos $\{x_n\}$ que satisfacen que $F(\vec{x}_{n+1}) \leq F(\vec{x}_n)$ se conocen con el nombre de métodos de descenso y funcionan de la siguiente forma: primero, se aísla el mínimo local $\vec{x}^*$ que nos interesa determinar; ésto se hace seleccionando una vecindad V_δ de $\vec{x}^*$ que no contenga otro punto extremo de F; posteriormente, se escoge un punto inicial en la vecindad que denotaremos siempre como $\vec{x}_0$. Trácese una recta que pase por $\vec{x}_0$ con una dirección $\vec{d}_0$ y escojase a lo largo de ella un punto $\vec{x}_1$ que satisfaga que $F(\vec{x}_1) \leq F(\vec{x}_0)$. Para generar un nuevo punto, se vuelve a repetir el mismo procedimiento: seleccionar una dirección d_1, trazar una recta que pasa por $\vec{x}_1$ con dicha dirección y seleccionar $\vec{x}_2$ tal que $F(\vec{x}_2) \leq F(\vec{x}_1)$. El algoritmo termina cuando determinamos un punto $\vec{x}_n = \vec{x}^*$ o cuando nos aproximamos al mínimo con la precisión deseada. En forma esquemática el algoritmo el el siguiente:

$$\begin{aligned} \text{Dado } \vec{x}_0 &\in V_\delta(\vec{x}^*), \\ \vec{x}_{n+1} &= \vec{x}_n + \alpha_n \vec{d}_n, \\ \text{con} \qquad F(x_n + \alpha_n \vec{d}_n) &\leq F(x_n + \alpha \vec{d}_n) \qquad \forall \alpha \in \Re. \end{aligned}$$

Aislar el mínimo no siempre es muy sencillo, se requiere usar todo el conocimiento que se tenga del comportamiento de la función F. Se sugiere hacer un análisis cualitativo de la función, antes de aplicar un algoritmo, para determinar una vecindad donde se encuentre el mínimo. La selección de la dirección en cada iteración distingue a los métodos. Por ejemplo si sólo se usa la información del gradiente se dice que los métodos son de tipo gradiente; si además se usa el Hessiano se dice que son métodos tipo Hessiano. Por último, determinar α_n, en cada paso, es equivalente a resolver una ecuación no lineal en una sóla variable ya que encontrar la t_n que minimiza la función a lo largo de una recta se reduce a determinar el punto crítico de una función de una sóla variable

$$\frac{dF(x_n + \alpha \vec{d}_n)}{d\alpha} = \vec{d}_n^t \ \nabla F(x_n + \alpha \vec{d}_n) = 0. \tag{2.5}$$

El problema de resolver la ecuación (2.5) se conoce con el nombre de búsqueda lineal y en la mayoría de los casos no puede encontrarse su solución exacta por lo que hay que aproximarla. En el siguiente capítulo se verán en detalle todo lo referente a los algoritmos de descenso.

2.7. Tipo de convergencia

A continuación deseamos dar algunos criterios teóricos que nos permitan comparar el desempeño de dos algoritmos numéricos. Dado un algoritmo lo que hacemos con él es generar una sucesión $\vec{x}_n$ que tienda al mínimo local que nos interesa aproximar. El n-ésimo término de la sucesión lo generamos a partir de minimizar la función a lo largo de una recta que pasa por el punto $\vec{x}_{n-1}$. La convergencia del algoritmo va a depender de cómo converja la sucesión que genera.

Diremos que un algoritmo tiene **convergencia global** si la sucesión que genera converge, independientemente del punto inicial x_0 que se seleccionó. Si, en cambio, su convergencia depende del punto inicial se dice que es un **algoritmo local**. Si F es una función cuadrática de $\Re^n$ a los reales y si un

algoritmo converge al mínimo a lo más en n iteraciones se dice que tiene **terminación cuadrática**.

Supongamos que se tienen dos algoritmos que convergen, ¿cómo decidir cuál escoger? Para poder dar una respuesta, se requiere tener un criterio para decidir, entre las sucesiones que generan, cuál converge más rápido.

Definición 2.7.1. *Diremos que una sucesión converge linealmente si existe* $K \in (0,1)$ *tal que*

$$||\vec{x}_{n+1} - \vec{x}^*|| \leq K||\vec{x}_n - \vec{x}^*||.$$

Definición 2.7.2. *Diremos que una sucesión converge cuadráticamente si*

$$||\vec{x}_{n+1} - \vec{x}^*|| \leq K||\vec{x}_n - \vec{x}^*||^2.$$

Definición 2.7.3. *Diremos que una sucesión converge superlinealmente si converge linealmente con* K_n *que tiende a cero cuando n tiende a infinito.*

Diremos que un algoritmo converge de cierta forma si la sucesión que genera converge de esa forma. Claramente un algoritmo con convergencia cuadrática converge más rápido que uno con convergencia lineal y dados dos algoritmos lineales converge más rápido aquel que tenga la menor K. A K se le conoce como la **rapidez de convergencia** cuando el orden de convergencia es lineal.

Ilustremos con un ejemplo cómo se determina el tipo de convergencia de una sucesión. ¿Qué tipo de convergencia tiene la sucesión $r_k = a^k$ con $0 < a < 1$? La sucesión converge a 0 linealmente con rapidez de convergencia $K = a$. En cambio la sucesión $r_k = a^{2^k}$ converge a cero cuadráticamente con rapidez de convergencia igual a 1.

En la práctica, es difícil determinar para cualquier problema, el valor de K y probar el tipo de convergencia. Lo que se hace es probar la convergencia y determinar su orden y el valor de K para los problemas cuadráticos. Recordemos que cerca de la vecindad de un mínimo la función se comporta como una función cuadrática.

Por otro lado ésto sólo nos da un aspecto del desempeño de un algoritmo. También hay que tomar en cuenta el número de evaluciones de la función, de su gradiente y Hessiano que se requieren en cada iteración. Asimismo, el tipo de instrumento de cálculo que se tiene a la mano. Si sólo se cuenta con una calculadora posiblemente seleccionemos a un algoritmo que requiera a lo más del gradiente de la función, de poco espacio en memoria y del menor número

de operaciones. Si el problema tiene muchas variables y sólo contamos con una PC, pues nos decidiríamos por un algoritmo intermedio que no requiriera de mucho espacio en memoria y de poco tiempo de cálculo. Pero si tenemos una supercomputadora, escogeríamos aquel algorimo que nos da la mejor precisión. La selección del algoritmo depende del número de variables, de qué tan regular es la función, del tipo de instrumento de cálculo con el que se cuenta y de la precisión que se desea.

2.8. Ejercicios

1. Clasifique los puntos críticos de las siguientes funciones:

   a) $f(x, y) = 2x^2 - 3y^2 + 2x - 3y + 7$.

   b) $f(x_1, x_2, x_3) = 2x_1^2 - 4x_3^2 + x_1x_2 - x_2x_3 - 6x_1$.

   c) $f(x, y) = x^3 + y^3 - 3x - 12y + 20$.

   d) $f(x, y) = x^4 + y^4 - x^2 - y^2 + 1$.

   e) $f(x, y) = (x^2 + y)e^{(x^2-y^2)}$.

2. Diga si las siguientes matrices son positivas definidas, negativas definidas o indefinidas

   a)
   $$\begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix},$$

   b)
   $$\begin{bmatrix} -4 & 0 & 1 \\ 0 & -3 & 2 \\ 1 & 2 & -5 \end{bmatrix},$$

   c)
   $$\begin{bmatrix} 3 & 1 & 2 \\ 1 & 5 & 3 \\ 1 & 2 & -5 \end{bmatrix}.$$

3. Demuestre que una matriz simétrica y positiva definida es invertible.

4. Demuestre que los valores propios de una matriz simétrica y positiva definida son reales y positivos.

5. Determine si las siguientes funciones son coercivas:

   - $f(x, y, z) = x^3 + y^3 + z^3 - xy$.
   - $f(x, y, z) = x^4 + y^4 + z^2 - 7xyz^2$.
   - $f(x, y, z) = ln(x^2y^2z^2) - x - y$.

6. Demuestre que $f(x, y) = x^3 + e^{3y} - 3xe^y$ tiene un único punto crítico y que este punto es un mínimo local pero no global.

7. Sea $A = B^tB$ con $B \in \Re^{m\times n}$ con $m > n$. Demuestre que si B es una matriz de rango completo entonces B es positiva definida. Hint: $(By)^tBy = x^tx$ con $By = x$.

8. Dada la siguiente tabla

t_i	r_i
1/13	.0863
3/13	.0863
6/12	.0860
1	.0861
5	.0917
10	.1012
20	.1010

   Determine el polinomio lineal que mejor ajuste los datos en el sentido de mínimos cuadrados.

9. Factorice la matriz asociada al siguiente problema cuadrático por medio de Cholesky. ¿Es positiva definida la matriz?

$$f(x, y, z) = 2x^2 + xy + y^2 + yz + z^2 - 6x - 7y - 8z + 9.$$

10. Demuestre que si A es una matriz simétrica existe una matriz Q ortogonal: $Q^tQ = I$ tal que $D = QAQ^t$ es una matriz diagonal cuyos elementos en la diagonal son los valores propios de A.

11. Grafique las curvas de nivel de la función f del ejercicio 9. Para ello, encuentre los valores y vectores propios de la matriz A. Dado que A es simétrica aplique el ejercicio anterior y defina $\vec{v} = Q\vec{x}$. Determine la función g que se obtiene al aplicar el cambio de variable, complete

cuadrados y grafique las curvas de nivel de esta función. Posteriormente, grafique las de la función original.

Capítulo 3

Métodos de descenso

En este capítulo veremos algunos algoritmos que nos permiten aproximar el mínimo de una función. En particular se verá el método de máximo descenso, el método de Newton y algunas variantes de éste conocidas con el nombre de métodos cuasi-Newton. Al final se presenta el método de gradiente conjugado. A lo largo de todo este capítulo vamos a suponer que F es una función de un abierto S de $\Re^n$ a los reales, que tiene un mínimo relativo en $\vec{x}^*$ y que es una función dos veces continuamente diferenciable en una vecindad de $\vec{x}^*$.

3.1. Introducción

Supongamos que se desea determinar la solución del siguiente problema

$$\min_{\vec{x}\in\Re^2} F(\vec{x})$$

con $F(x,y) = 3x^2 + y^2 - x^4 - 12$. Este es un problema de minimización no lineal. Aplicando el criterio de la primera derivada se obtiene que la solución (x,y) debe satisfacer

$$\begin{aligned} \frac{\partial F(x,y)}{\partial x} &= 6x - 4x^3 = 0, \\ \frac{\partial F(x,y)}{\partial y} &= 2y = 0. \end{aligned}$$

De la segunda ecuación podemos deducir que $y = 0$ y de la primera se tiene que $x(6 - 4x^2) = 0$, lo que implica que son puntos críticos: $(0,0)$, $(\sqrt{\frac{3}{2}}, 0)$

y $(-\sqrt{\frac{3}{2}}, 0)$. Calculando el Hessiano se puede comprobar que $(0,0)$ es un mínimo y los otros dos son puntos silla, véase Figura 3.1.

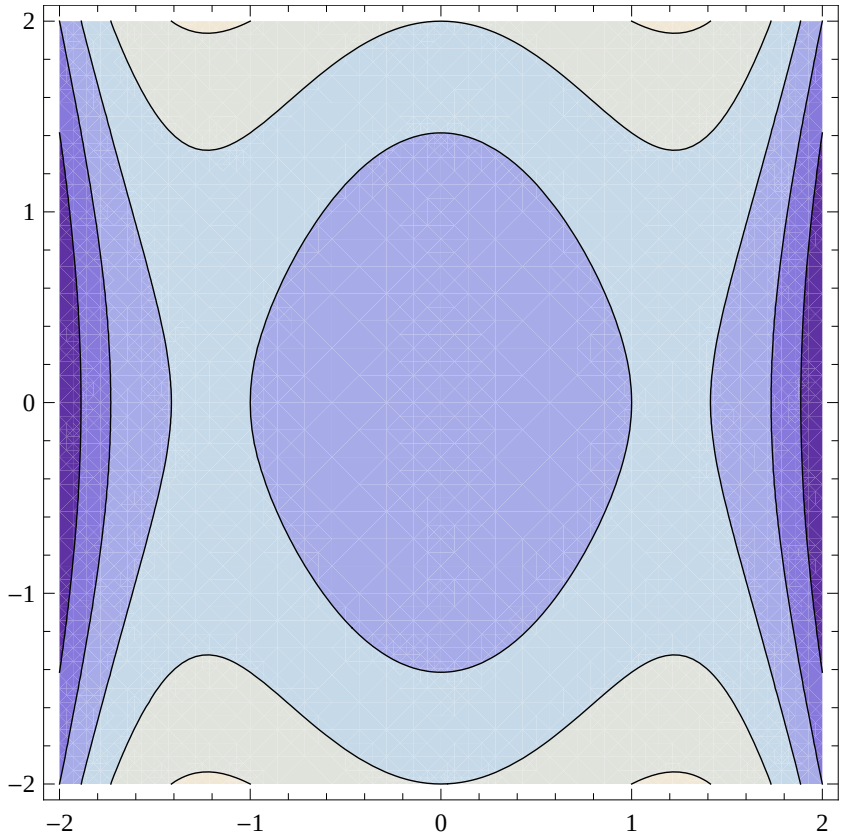

Figura 3.1: Curvas de nivel de $F(x,y) = 3x^2 + y^2 - x^4 - 12$.

Recordemos que un método de descenso consiste en lo siguiente:

$$\begin{aligned} & \text{Dado } \vec{x}_0 && \in V_\delta(\vec{x}^*), \\ & && \vec{x}_{n+1} = \vec{x}_n + \alpha_n \vec{d}_n, \\ \text{con} \qquad & F(x_n + \alpha_n \vec{d_n}) && \leq F(x_n + \alpha \vec{d_n}) \qquad \forall \alpha \in \Re. \end{aligned}$$

Para simplificar la notación denotemos como $\vec{g}_k$ al gradiente de F evaluado en $\vec{x}_k$ o sea $\vec{g}_k = \nabla F(\vec{x}_k)$. Supongamos que se han generado k términos por medio del algoritmo anterior, si se expande F en una serie de Taylor en el

punto $\vec{x}_k$ se tiene que para alguna θ entre $(0,1)$ se cumple

$$\begin{aligned} F(\vec{x}_{k+1}) &= F(\vec{x}_k) + \vec{g}_k^t(\vec{x}_{k+1} - \vec{x}_k) \\ &\quad + \frac{1}{2}\,(\vec{x}_{k+1} - \vec{x}_k)^t H_F(\theta\vec{x}_k + (1-\theta)\vec{x}_{k+1})(\vec{x}_{k+1} - \vec{x}_k). \end{aligned}$$

Como el término $\vec{x}_{k+1}$ fue generado por medio de un método de descenso entonces satisface que $\vec{x}_{k+1} - \vec{x}_k = \alpha_k \vec{d}_k$, por lo que se tiene que

$$F(\vec{x}_{k+1}) \approx F(\vec{x}_k) + \alpha_k \vec{g}_k^t \vec{d}_k$$

y para $\alpha_k > 0$, $F(\vec{x}_{k+1}) \leq F(\vec{x}_k)$ siempre que

$$\vec{g}_k^t \vec{d}_k \leq 0. \tag{3.1}$$

Esta última relación nos dice, que en forma aproximada, la condición que debe satisfacer la dirección $\vec{d}_k$ para que la sucesión $\vec{x}_k$ sea una sucesión de descenso es $\vec{g}_k^t \vec{d}_k \leq 0$; en consecuencia, cuando una dirección $\vec{d}_k$ satisface (3.1) se le llama una dirección de descenso. Recordemos que si $\vec{x}_k$ y $\vec{x}_{k+1}$ están suficientemente cerca del mínimo, el Hessiano de F debe tomar valores positivos en la recta que une a estos dos puntos por lo que la única forma que disminuya el valor de F en $\vec{x}_{k+1}$ es que se cumpla esta relación.

Observemos que no es suficiente que F tenga un mínimo en $\vec{x}^*$ y que sea una función acotada inferiormente en una vecindad del punto $\vec{x}_0$, para que cualquier sucesión de descenso $\vec{x}_k$ converja al mínimo; por ejemplo, en el caso que $F(x) = x^4$, si $x_0 = 3/2$ y $x_k = x_{k-1} - 1/9^k$, se tiene una sucesión de descenso que converge a $11/8$ en lugar de al mínimo que es $x^* = 0$. En este caso las α_k son cada vez más pequeñas por lo que no se logra alcanzar al mínimo. Otra situación que puede impedir que una sucesión de descenso converja es que las direcciones $\vec{d}_k$ tiendan a ser ortogonales al gradiente de F en $\vec{x}_k$, sin que el gradiente $\vec{g}_k$ converja a cero; en este caso, los puntos que se generen quedan atrapados en una curva de nivel de F. Por ejemplo, cuando $F(x,y) = x^2 + y^2$, $\vec{x}_0 = (-2,1)$ y $\vec{d}_k = (1, -x_k/y_k - 1/2^k)$. Por último, puede suceder que se tome un punto $\vec{x}_0$ que este sobre una curva de nivel no cerrada, por ejemplo cerca de un punto silla, en ese caso la sucesión no logrará entrar a la región donde las curvas de nivel se vuelven elípticas o sea cerca del mínimo. El siguiente resultado nos da condiciones para poder garantizar que una sucesión de descenso converja a un mínimo.

Teorema 3.1.1. *Si F es una función continuamente diferenciable que alcanza un mínimo relativo en $\vec{x}^*$ y si F esta acotada inferiormente, las siguientes*

condiciones nos permiten asegurar que una sucesión de descenso converge al mínimo:

i).- $\vec{x}_0$ *se selecciona en una curva de nivel cerrada y acotada.*

ii).- La función F *decrece suficientemente en cada paso.*

ii).- Las direcciones de descenso $\vec{d}_k$ *no satisfacen que*

$$\lim_{k\to\infty} {}^t\vec{g}_k\vec{d}_k = 0,$$

sin que $\vec{g}_k \to 0$ *en el límite.*

El resultado formal y su demostración pueden verse en el Luenberger [6].

Criterios para detener un algoritmo de descenso

Aunque el algoritmo de máximo descenso sea convergente al mínimo de una función F, no se sabe si el número de iteraciones que se requieren para ello sea finito o no. En la práctica se necesitan de criterios que nos permitan decidir cuándo hemos aproximado al mínimo con la precisión deseada. Hay dos criterios que se usan: el criterio del gradiente y un criterio que estima el error relativo que llamaremos el criterio de la sucesión. Recordemos que el error que se comete al estimar el mínimo $\vec{x}^*$ por medio del k-ésimo término de una sucesión se puede calcular de dos maneras distintas: estimando el error absoluto, que es igual a $||\vec{x}^* - \vec{x}_k||$ o por medio del error relativo que se define por

$$E_k = \frac{||\vec{x}^* - \vec{x}_k||}{||\vec{x}^*||}$$

y que es menos sensible a los cambios de escala.

El criterio del gradiente consiste en evaluar en cada iteración k la norma del gradiente de la función en $\vec{x}_k$ o sea $||\vec{g}_k||$. Si la norma es cercana a cero es de esperarse que $\vec{x}_k$ esté cerca del mínimo. El criterio de la sucesión está basado en la idea de que en $\Re^n$ toda sucesión convergente es una sucesión de Cauchy, es decir al aproximarnos al límite los elementos de la sucesión distan cada vez menos unos de otros, por ello si la expresión

$$\frac{||\vec{x}_i - \vec{x}_{i-1}||}{||\vec{x}_i||}, \tag{3.2}$$

es pequeña es de esperarse que estemos cerca del límite.

¿Qué significa estar "cerca"del cero y que (3.2) sea pequeño? Ambas propiedades dependen de la precisión del equipo de cálculo que se utilice, es decir el número de dígitos que almacena en la mantisa en la notación de punto flotante, y del grado de precisión que se desea. Para cada problema se debe fijar, de antemano, el valor de un parámetro $\varepsilon > 0$, que dependerá del número de cifras significativas que se deseen obtener; por ejemplo, si se desean dos cifras significativas, ε se escoge igual a 10^{-2}. La prueba del gradiente o de la sucesión consisten en verificar, para cada iteración $\vec{x}_k$, si la norma $||\vec{g}_k||$ o la expresión (3.2) son mayores o menores que ε; si son mayores, se genera $\vec{x}_{k+1}$; si son menores o iguales se detiene el algoritmo.

3.2. Búsqueda lineal

Llamamos búsqueda lineal a la determinación de α_k en un método de descenso. Para el caso cuadrático, cuando $F(\vec{x}) = \frac{1}{2}{}^t\vec{x}A\vec{x} - {}^t\vec{x}\vec{b} + c$, es posible determinar el valor de α_k en forma exacta para cualquier valor que tome la dirección $\vec{d}_k$; α_k debe satisfacer

$$\frac{dF(\vec{x}_k + \alpha\vec{d}_k)}{d\alpha}|_{\alpha=\alpha_k} = 0.$$

Al calcular la derivada de F se obtiene

$$\begin{aligned}\frac{dF(\vec{x}_k + \alpha\vec{d}_k)}{d\alpha} &= \vec{d}_k^t \nabla F(\vec{x}_k + \alpha\vec{d}_k) \\ &= \vec{d}_k^t[A(\vec{x}_k + \alpha\vec{d}_k) - \vec{b}], \\ &= \vec{d}_k^t(A\vec{x}_k - \vec{b}) + \alpha\vec{d}_k^t A\vec{d}_k.\end{aligned}$$

Como $\vec{g}_k = A\vec{x}_k - \vec{b}$ se tiene que α_k debe satisfacer que

$$\alpha_k = -\frac{{}^t\vec{d}_k\vec{g}_k}{{}^t\vec{d}_k A\vec{d}_k}. \tag{3.3}$$

Determinar α_k en el caso no lineal puede ser muy engorroso. Como el cálculo de α_k tiene como objetivo generar una aproximación mejor al mínimo, no vale la pena gastar tiempo y esfuerzo en calcular el valor exacto, basta con obtener una buena aproximación que no introduzca un error que pueda

a larga hacer diverger al proceso. A continuación se verán algunos algoritmos para estimar el valor de α_k en un número pequeño de iteraciones. Estos algoritmos se pueden calcular independientemente de cómo se determina la dirección de descenso en el paso k, por ello se aplicarán para cualquier vector $\vec{d_k}$.

Para ilustrar las dificultades que se presentan en el cálculo de α_k, apliquemos un método de descenso a la siguiente función objetivo $F(x,y) = 16x^4 - 16x^3 + 6x^2 - x + \frac{1}{16} + 3x^2y^2$.

El gradiente de la función es

$$\begin{aligned} \frac{\partial F(x,y)}{\partial x} &= 64x^3 - 48x^2 + 12x - 1 + 6xy^2 = 0, \\ \frac{\partial F(x,y)}{\partial y} &= 6x^2y = 0. \end{aligned}$$

Supongamos que aplicamos un método de descenso para aproximar el mínimo de la función con $\vec{d_0} = -\vec{g_0}$ y $\vec{x}_0 = (0,0)$. Para determinar la α_0 hay que resolver la siguiente ecuación

$$\frac{dF(\vec{x}_0 + \alpha\vec{d_0})}{d\alpha} = 64\alpha^3 - 48\alpha^2 + 12\alpha - 1 = 0.$$

Esta ecuación es de orden cúbico por lo hay que usar un esquema numérico para aproximar la solución. ¿Por qué mejor no buscar un algoritmo poco costoso que nos dé una aproximación razonable del valor de α_k sin que tengamos en cada paso k que calcular explícitamente el valor de la derivada de F respecto a α?

3.2.1. Búsqueda lineal no exacta

Deseamos encontrar un algoritmo que nos permita estimar, en cada paso k de un método de descenso, el valor de α_k sin tener que calcular $dF/d\alpha$. Los algoritmos deben al menos tener las siguientes características:

1.- El algoritmo debe darnos una buena aproximación al valor exacto de α_k en un número finito de pasos.

2.- En el caso cuadrático la sucesión que se genere debe converger al valor exacto de α_k.

Sea φ_k una función de variable real definida por

$$\varphi_k(\alpha) = F(\vec{x}_k + \alpha\vec{d}_k),$$

claramente $\varphi_k(0) = F(\vec{x}_k)$ y $\varphi'_k(0) = \vec{d}_k^t\vec{g}_k$; además, para cualquier método de descenso se tienen que $\varphi'_k(0) < 0$. Lo que se desea es obtener un procedimiento que en cada paso k nos determine una aproximación $\hat{\alpha}_k$ al valor exacto de α_k que satisfaga que $\varphi_k(\hat{\alpha}_k) < \varphi_k(0)$.

3.2.2. Algoritmo de Armijo

Trácese la recta que pasa por $\alpha(0)$ con pendiente $\varepsilon\varphi'_k(0)$ para alguna $\varepsilon \in (0, 1)$. La ecuación de esta recta es $y(\alpha) = \varphi_k(0) + \alpha\varepsilon\varphi'_k(0)$.

El algoritmo de Armijo determina un intervalo dónde se encuentran los valores de α que son buenas estimaciones de α_k; a este intervalo le llamaremos intervalo de valores admisibles de α_k. Para ello aplica dos criterios. El primero consiste en determinar, para cada paso k, una $\hat{\alpha}_k$ que satisfaga que $\varphi_k(\hat{\alpha}_k)$ está por debajo de la recta $\varphi_k(0) + \varepsilon\varphi'_k(0)$, es decir

$$\varphi_k(\hat{\alpha}_k) \leq \varphi(0) + \varepsilon\hat{\alpha}_k\varphi'_k(0). \tag{3.4}$$

Si $\hat{\alpha}_k$ satisface (3.4) entonces es una buena aproximación al valor exacto α_k. Para evitar que el nuevo punto $\vec{x}_{k+1} = \vec{x}_k + \hat{\alpha}_k\vec{d}_k$ esté muy cerca de $\vec{x}_k$, y, con ello, avancemos muy lentamente hacia el mínimo, Armijo sugiere comprobar si para múltiplos de $\hat{\alpha}_k$ se sigue satisfaciendo (3.4). El procedimiento que se sigue es el siguiente: determínese la mínima j en los enteros para la cual

$$\varphi_k(2^j\hat{\alpha}_k) > \varphi_k(0) + \varepsilon\ 2^j\ \hat{\alpha}_k\varphi'_k(0). \tag{3.5}$$

Entonces el intervalo admisible de α_k es $(\hat{\alpha}_k, 2^j\hat{\alpha}_k)$ y se sugiere escoger como estimación a α_k el valor de $2^{j-1}\hat{\alpha}_k$.

Apliquemos el criterio de Armijo para determinar una estimación de α_0 para la función no lineal $F(x, y) = 16x^4 - 16x^3 + 6x^2 - x + \frac{1}{16} + 3x^2y^2$, con $\vec{x}_0 = (0, 0)$, $\vec{g}_0 = (-1, 0)$ y $\varphi_0(\alpha) = 16\alpha^4 - 16\alpha^3 + 6\alpha^2 - \alpha + 1/16$. En este caso el valor exacto de $\alpha_0 = 1/4$. Tomemos a $1/10$ como el valor de ε. Como primer paso, busquemos una α tal que

$$\varphi_0(\alpha) \leq \frac{1}{16} - \frac{\alpha}{10}.$$

Observemos que para $\alpha = 1$ no se cumple esta desigualdad pero que sí se satisface para $\hat{\alpha}_0 = 1/10$. Como siguiente paso para determinar el intervalo admisible, encontremos la j para la cual

$$\varphi_0(\frac{2^j}{10}) > \frac{1}{16} - \frac{2^j}{100}.$$

Para $j = 3$ ya no se cumple la desigualdad anterior, por lo tanto cualquier valor de $\alpha \in (0, .4)$ es una buena estimación. Seleccionemos $\alpha_0 = .4$, entonces $\vec{x}_1 = (.4, 0)$ y $F(x_1) = -.0015$ que es un valor menor a $1/16 = F(0, 0)$.

El algoritmo de Armijo es muy sencillo de implementar en una instrumento de cálculo pero, no cumple con la condición de que para cualquier valor de $\varepsilon \in (0, 1)$, el mínimo de una función cuadrática debe estar en el intervalo admisible de α_k. Este inconveniente del algoritmo de Armijo es muy fácil de comprobar: supongamos que $F(\vec{x}) = \frac{1}{2}{}^t\vec{x}A\vec{x} - {}^t\vec{x}\vec{b} + c$ y expandamos en serie de Taylor a $\varphi_k(\alpha)$ alrededor de cero, $\varphi_k(\alpha)$ puede escribirse como

$$\varphi_k(\alpha) = \varphi_k(0) + \alpha \vec{d}_k^t \vec{g}_k + \frac{\alpha^2}{2}\, \vec{d}_k^t A \vec{d}_k,$$

substituyendo el valor exacto α_k dado por (3.3) y simplificando se tiene

$$\varphi_k(\alpha_k) = \varphi_k(0) + \frac{\alpha_k}{2} \vec{d}_k^t \vec{g}_k = \varphi_k(0) + \frac{\alpha_k}{2} \varphi_k'(0),$$

y la desigualdad

$$\varphi_k(\alpha_k) \leq \varphi_k(0) + \varepsilon \alpha_k \varphi_k'(0)$$

se cumple siempre que $\varepsilon \in (0, 1/2)$ ya que $\vec{d}_k^t \vec{g}_k \leq 0$. Por lo que hay que restringir el valor de ε para el algoritmo de Armijo.

3.2.3. Interpolación cuadrática

Otro procedimiento para aproximar el valor de α_k en la búsqueda lineal es determinar tres puntos que estén en la región admisible de α_k y construir un polinomio cuadrático que aproxime el valor de $\varphi_k(\alpha)$ en este intervalo. Como el polinomio es una parábola que se abre hacia arriba, dado que $\varphi_k'(0) < 0$, tiene un mínimo que se utiliza como aproximación de α_k.

Sea $p(x)$ un polinomio cuadrático de la forma $p(x) = ax^2 + bx + c$, el problema consiste en determinar a, b y c tales que $p(0) = \varphi_k(0)$, $p'(0) = \varphi_k'(0) = \vec{d}_k^t \vec{g}_k$ y, por último, $p(\alpha_0) = \varphi_k(\alpha_0)$ con α_0 un punto para el cual

$\varphi_k'(\alpha_0) > 0$; esto último para asegurarnos que en $(0, \alpha_0)$ se encuentra el valor mínimo de $\varphi_k(\alpha)$.

Este valor se determina proponiendo un valor para α_0; si la derivada es negativa se incrementa y se busca un nuevo punto, si el valor es positivo se comprueba si se puede hacer más pequeño el intervalo. Los coeficientes del polinomio satisfacen:

$$c = \varphi_k(0), \quad b = \varphi_k'(0) \quad \text{y} \quad a = \frac{\varphi_k(\alpha_0) - b\alpha_0 - c}{\alpha_0^2}.$$

El mínimo del polinomio se alcanza en $\frac{-b}{2a}$ por lo que se toma $\alpha_k = \frac{-b}{2a}$.

Ejemplo

Apliquemos este algoritmo para obtener una aproximación a α_0 en el caso que $F(x,y) = 16x^4 - 16x^3 + 6x^2 - x + \frac{1}{16} + 3x^2y^2$, con $\vec{x}_0 = (0,0)$, $\vec{g}_0 = (-1,0)$ y $\varphi_0(\alpha) = 16\alpha^4 - 16\alpha^3 + 6\alpha^2 - \alpha + 1/16$. Para construir el polinomio cuadrático $p(x)$ usamos la información que $p(0) = \varphi_0(0) = 1/16$, $p'(0) = \varphi_0'(0) = -1$ y determinanos un valor de α para el cual $\varphi_0'(\alpha) \geq 0$. En este caso se cumple para $\alpha = 1$ por lo que sabemos que el mínimo de $\varphi(\alpha) \in (0,1)$. Tratemos de reducir el intervalo: $\varphi_0'(1/2) = 1$ y $\varphi_0'(1/4) = 0$ por lo que se ha encontrado el valor mínimo. Si hubiéramos detenido el proceso en $\alpha = 1/2$ el polinomio cuadrático tendría como coeficientes:

$$c = \varphi_0(0) = 1/16, \; b = \varphi_0'(0) = -1 \text{ y a} = \frac{(\varphi_0(1/2) - (1/2)\text{b} - \text{c})}{1/4} = 2$$

y $\alpha_0 = \frac{-b}{2a} = 1/4$ por lo que obtenemos el valor exacto.

Si el valor α_k que se obtiene al usar interpolación cuadrática no satisface la desigualdad de Armijo, (3.4), entonces se usa interpolación cúbica que aproxima mejor a las funciones con cambios pronunciados en la curvatura. Para profundizar el tema de búsqueda lineal, consultar el libro de Nocedal, ver [9]

3.3. Método de máximo descenso

En la sección anterior vimos algunos aspectos sobre los métodos de descenso en general. En esta sección comenzaremos a estudiar algunos de los

algoritmos más usados en la optimización no lineal. Los algoritmos de descenso se distinguen entre sí por la forma en la que se calcula en cada paso k la dirección de descenso.

Dado un punto $\vec{x}_0$ y una función objetivo F, ¿cuál es la dirección en la que F decrece mas? Por el curso de cálculo de varias variables sabemos que F disminuye más si nos movemos en la dirección de menos el gradiente de F en $\vec{x}_0$ ya que

$$\vec{g}_0^t \vec{d}_0 = ||\vec{g}_0||\ ||\vec{d}_0|| \cos\theta$$

y el mínimo valor que puede tomar cos θ es cuando $\theta = \pi$ o sea cuando $d_0 = -g_0$. El método de máximo descenso o descenso pronunciado consiste en seleccionar para cada paso k a $-\vec{g}_k$ como la dirección $\vec{d}_k$.

Algoritmo de máximo descenso

El algoritmo consiste en los siguiente:

i).- Dado $\vec{x}_0 \in V_\delta(\vec{x}^*)$ y $\varepsilon > 0$

ii).- $\vec{x}_{n+1} = \vec{x}_n - \alpha_n \vec{g}_n$, con

$$F(x_n - \alpha_n \vec{g}_n) \leq F(x_n - \alpha \vec{g}_n) \qquad \forall \alpha \in \Re.$$

iii).- si $||\vec{g}_{n+1}|| \leq \varepsilon$ y $\frac{||\vec{x}_{n+1} - \vec{x}_n||}{||\vec{x}_n||} \leq \varepsilon$ entonces $x^* \approx x_{n+1}$ y se detiene el algoritmo.

iv).- Si no se cumplen las condiciones del inciso iii) regresar a ii) con $\vec{x}_n = \vec{x}_{n+1}$.

Ejemplo

Apliquemos este algoritmo para el caso en que la función objetivo es una función cuadrática de la forma

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

El algoritmo de máximo descenso para este caso es de la forma:

$$\begin{aligned} \text{Dado } \vec{x}_0 &\in V_\delta(\vec{x}^*), \\ \vec{x}_{n+1} &= \vec{x}_n - \alpha_n \vec{g}_n, \\ \text{con} \qquad \alpha_n &= \frac{\vec{g}_n^t g_n}{\vec{g}_n^t A \vec{g}_n}. \end{aligned}$$

Apliquemos este algoritmo al cuarto ejemplo de la sección 2.4. que consiste en determinar el mínimo de

$$F(x,y) = .01x^2 + .01y^2 + .007xy - 485x - 675y + 400,000.$$

El gradiente de F esta dado por

$$\nabla F(x,y) = (.02x + .007y - 485,\ .007x + .02y - 675)$$

y la solución exacta con dos cifras decimales es $(14173.789, 28789.17)$. Supongamos que $\vec{x}_0 = (10000, 20000)$ y que $\varepsilon = 10^{-2}$ entonces aplicando el método de máximo descenso se obtienen los siguientes valores

Tabla 3.1

i	$\vec{x}_i$	$\vec{g}_i$	α_i	$\|\|\vec{g}_i\|\|$	R_i
0	(10,000, 20,000)	(-145, 205)	63,05	205.09	-
1	(15,450.5, 27,706.6)	(17.95, 12.7)	74.64	22	.29
2	(14,110.17, 28,656.6)	(-2.20, -3.11)	37.58	3.82	.05
3	(14,193., 28,772.7)	(0.26, -0.19)	74.82	0.3321	.003
4	(14,172.82, 28,787.13)	(-.032, -0.045)	8.34	.058	.0007
5	(14,174.08, 28788.92)	(.004, -.0029)	-	0.0051	.00006

donde $R_i = \frac{||\vec{x}_i - \vec{x}_{i-1}||}{||\vec{x}_i||}$. Observemos que si sólo hubiéramos usado el criterio de la sucesión, el proceso se hubiera suspendido en la tercera iteración mientras que con el criterio del gradiente se requiere generar hasta cinco iteraciones. El error relativo al tomar a $\vec{x}_5$ como aproximación al mínimo es de

$$\frac{||\vec{x}_5 - \vec{x}^*||}{||\vec{x}^*||} = .00001,$$

por lo que en este caso el criterio (3.2) es una mejor estimación del error relativo.

3.3.1. Convergencia del método de máximo descenso

Dada una función objetivo F, ¿bajo qué condiciones converge el método de máximo descenso? y ¿con qué rapidez converge? Las respuestas a estas preguntas nos permitirán comparar el desempeño de este método respecto a la de otros métodos que se verán más adelante. Las respuestas que daremos

son para el caso en que F sea una función cuadrática con matriz A simétrica y positiva definida, es decir F es de la forma

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

Supongamos que F alcanza su mínimo en $\vec{x}^*$ y definamos una nueva función $E(\vec{x})$ de la forma

$$E(\vec{x}) = \frac{1}{2}(\vec{x} - \vec{x}^*)^t A(\vec{x} - \vec{x}^*).$$

Observemos que el mínimo de la función E se alcanza en $\vec{x}^*$.

Lemma 3.3.1. *Sea $\{\vec{x}_k\}$ una sucesión generada por el método de descenso pronunciado para aproximar el mínimo de una función F cuadrática con matriz A simétrica y positiva definida, entonces*

$$E(\vec{x}_{k+1}) = \{1 - \frac{(\vec{g}_k^t\vec{g}_k)^2}{\vec{g}_k^t A\vec{g}_k \ \vec{g}_k^t A^{-1}\vec{g}_k}\}E(\vec{x}_k).$$

Ver demostración en [6].

Lemma 3.3.2. *(Desigualdad de Kantorovich) Sea A una matriz $n \times n$ simétrica, positiva definida con λ_n y λ_1 como los valores propios más grande y más pequeño de A, respectivamente entonces*

$$\frac{(\vec{x}^t\vec{x})^2}{(\vec{x}^t A\vec{x})(\vec{x}^t A^{-1}\vec{x})} \geq \frac{4(\lambda_1\lambda_n)}{(\lambda_1 + \lambda_n)^2}.$$

Ver demostración en el [6].

Teorema 3.3.3. *(Convergencia método de máximo descenso para el caso cuadrático) Para cualquier $\vec{x}_0 \in \Re^n$, el método de máximo descenso converge al mínimo $\vec{x}^*$ de F y*

$$E(\vec{x}_{k+1}) \leq \{\frac{(\lambda_n - \lambda_1)}{(\lambda_n + \lambda_1)}\}^2 E(\vec{x}_k). \tag{3.6}$$

con λ_n y λ_1 como los valores propios más grande y más pequeño de A, respectivamente.

Dem: La demostración se obtiene de combinar los resultados de los dos lemas anteriores

$$E(\vec{x}_{k+1}) = \{1 - \frac{(\vec{g}_k^t \vec{g}_k)^2}{\vec{g}_k^t A \vec{g}_k \vec{g}_k^t A^{-1} \vec{g}_k}\} E[x_k]$$

$$E(\vec{x}_{k+1}) \leq \{1 - \frac{4\lambda_1 \lambda_n}{(\lambda_1 + \lambda_n)^2}\} E(\vec{x}_k),$$

al hacer algebra se obtiene el resultado del teorema.

Dado que A es positiva definida y simétrica, la función $E(\vec{x}_k)$ nos mide el error que cometemos al aproximar $\vec{x}^*$ por la k-ésima iteración en una norma que depende de la matriz A y que se define por

$$\|\vec{x}\|_A^2 = \vec{x}^t A \vec{x}.$$

El resultado anterior nos permite asegurar que al cumplirse las hipótesis del teorema

$$\|\vec{x}_{k+1} - \vec{x}^*\|_A \leq \frac{(\lambda_n - \lambda_1)}{(\lambda_n + \lambda_1)} \|\vec{x}_k - \vec{x}^*\|_A.$$

Como A es una matriz positiva definida, todos sus valores propios son positivos, por lo que

$$K = \frac{(\lambda_n - \lambda_1)}{(\lambda_n + \lambda_1)} < 1.$$

La desigualdad anterior nos permite garantizar que el método de descenso pronunciado converge linealmente con rapidez de convergencia K bajo la norma $\|.\|_A$. Recordemos que en $\Re^n$ todas las normas son equivalentes por lo que obtenemos tambien la convergencia en la norma euclideana. □

Cuando el máximo y el mínimo valor propio distan mucho entre sí, el valor de este cociente es cercano a uno, y en consecuencia la convergencia es muy lenta. Si los valores propios están muy cercanos, este método puede funcionar bien. Otra conclusión importante es que el método de máximo descenso converge *globalmente* pues el valor de K no depende de cuál es el punto inicial $\vec{x}_0$.

Al aplicar el resultado de este teorema para el ejemplo de los televisores. Los valores propios de la matriz A respectiva son: $\lambda_2 = .027$ y $\lambda_1 = .013$. La rapidez de convergencia K está dada por

$$K = \frac{(\lambda_2 - \lambda_1)}{(\lambda_1 + \lambda_2)} = .35.$$

Este valor nos permite estimar el número de iteraciones que se requieren para obtener la precisión ε que se desea. De la expresión (3.6) se obtiene una cota del error que se comete en la n ésima iteración que depende de K y el error inicial

$$\|\vec{x}_n - \vec{x}^*\|_A \leq K^n \|\vec{x}_0 - \vec{x}^*\|_A.$$

Si se desea obtener una precisión ε, hay que calcular cuál valor debe tomar n para que

$$K^n \|\vec{x}_0 - \vec{x}^*\|_A \leq \varepsilon.$$

Al despejar n se obtiene

$$n \geq \frac{\ln \varepsilon - \ln(\|\vec{x}_0 - \vec{x}^*\|_A)}{\ln(K)}. \tag{3.7}$$

En el caso de los televisores si $\varepsilon = 10^{-2}$ entonces $n \geq 11.38$. Se requieren de al menos doce iteraciones para obtener la precisión deseada. Esta cota es conservadora: en la práctica, en un número menor de iteraciones se obtiene la precisión deseada.

3.3.2. Aplicación al caso no lineal

Si la función objetivo es cualquier función no lineal, la convergencia del método de máximo descenso depende fuertemente de las características de la función. En caso de que haya convergencia, ésta será lineal y la rapidez de convergencia se puede estimar a partir de los valores propios del Hessiano de la función objetivo, evaluados en el mínimo $\vec{x}^*$. Si λ_n y λ_1 son los valores propios máximo y mínimo, respectivamente, de $H_F(\vec{x}^*)$ entonces

$$K \approx \frac{(\lambda_n - \lambda_1)}{(\lambda_n + \lambda_1)}.$$

Los valores propios se pueden estimar por medio del Hessiano evaluado en $\vec{x}_k$. Apliquemos este algoritmo a la función $F(x, y) = 3x^2 + y^2 - x^4 - 12$ con búsqueda lineal inexacta aplicando interpolación.

Tabla 3.2

i	$\vec{x}_i$	$\|\|\vec{g}_i\|\|$
0	(1/3,1.)	2.8284
1	(.038,.6)	1.2218
2	(.053,.12)	.4
3	(-.010,.072)	.15
4	(.0021,.043)	.08
5	(-.00296,.0086)	.024
6	(.00059,.005)	.01
7	(-.0008,.001)	.0054
8	(.00016,.0008)	.0015

El punto mínimo de este problema es $(0, 0)$. Observemos que el error absoluto en la octava iteración es de .0008 mientras que el criterio del gradiente nos da .0015. Los valores propios del Hessiano en $(0, 0)$ son $\lambda_1 = 6$ y $\lambda_2 = 2$. La rapidez de convergencia es 1/2. Por lo que para tener tres cifras significativas se requiere de al menos 8 iteraciones.

Apliquemos este algoritmo a otra función un poco más compleja: $F(x, y) = 16x^4 - 16x^3 + 6x^2 - x + \frac{1}{16} + 3x^2y^2$, con $\vec{x}_0 = (1, 1)$ y con búsqueda lineal inexacta.

Tabla 3.3

i	$\vec{x}_i$	$\|\|\vec{g}_i\|\|$	R_i
0	(1,1)	57.05	1.4142
1	(.94,.998)	49.59	.041
2	(.89,.996)	43.84	.037
3	(.84,.994)	39.25	.033
4	(.81,.993)	35.53	.030
5	(.77,.991)	32.44	.028
10	(.635,.986)	22.62	.020
20	(.5097,.982)	16.31	.0014
30	(.5052,.9824)	16.12	.00014

El mínimo de esta función esta en $(1/4, 0)$. Observemos que en la treintava iteración el error absoluto es de 1.01 que es enorme. Los valores propios del Hessiano en la solución son: $\lambda_1 = 24$ y $\lambda_2 = 0.375$; así que la rapidez de convergencia es aproximadamente de 0.9692. Al usar la expresión (3.10) para estimar el número de iteraciones que son necesarias para tener dos cifras significativas, obtenemos que n es mayor o igual a 113. Esto en el caso de

búsqueda lineal exacta. Por lo que no es de sorprenderse que en el método de máximo descenso no sea muy usado para estimar la solución de problemas reales.

3.4. Método de Newton

Como vimos en el capítulo anterior, el método de máximo descenso es muy fácil de implementar computacionalmente pero, desgraciadamente, converge muy lentamente debido a que su orden de convergencia es lineal. A continuación se estudiará el método de Newton que tiene convergencia cuadrática cerca del mínimo.

El método de Newton es un método diseñado para converger en una sóla iteración cuando la función a minimizar es cuadrática. El método consiste en minimizar en cada iteración k una función cuadrática $G_k(\vec{x})$ que se obtiene al expander en serie de Taylor a $F(\vec{x})$ alrededor de $\mathbf{x}_k$ hasta el segundo término. Es decir, $G_k(\vec{x})$ es igual a

$$G_k(\vec{x}) = F(\vec{x}_k) + (\vec{x} - \vec{x}_k)^t \vec{g}_k + \frac{1}{2}\ (\vec{x} - \vec{x}_k)^t H_F(\vec{x}_k)(\vec{x} - \vec{x}_k).$$

Para obtener el mínimo de $G_k(\vec{x})$ obtenga su gradiente e iguálelo a cero

$$\nabla G_k(\vec{x}) = \vec{g}_k + H_F(\vec{x}_k)(\vec{x} - \vec{x}_k) = 0;$$

entonces, el mínimo se alcanza en

$$\vec{x} = \vec{x}_k - H_F(\vec{x}_k)^{-1}\vec{g}_k,$$

siempre que el Hessiano de F en x_k sea positivo definido. El método de Newton selecciona en cada paso como el punto $\vec{x}_{k+1}$ al mínimo de la función $G_k(\vec{x})$.

3.4.1. Algoritmo de Newton

El algoritmo de Newton es el siguiente: Dada una función F dos veces continuamente diferenciable en una vecindad V_δ del mínimo x^*, $\vec{x}_0$ en $V_\delta(\vec{x}^*)$ y $rtol > 0$ como la tolerancia:

1. Determine $\vec{d_k}$ resolviendo primero el sistema

$$H_F(\vec{x}_k)\vec{d_k} = -\vec{g}_k. \tag{3.8}$$

2. Calcúlese $\vec{x}_{k+1}$ por medio de la expresión

$$\vec{x}_{k+1} = \vec{x}_k + \vec{d}_k. \tag{3.9}$$

3. Si $||\vec{g}_{k+1}|| \leq rtol$ y $\frac{||\vec{x}_{k+1} - \vec{x}_k||}{||\vec{x}_{k+1}||} \leq rtol$ entonces se toma a $\vec{x}^* \approx \vec{x}_{k+1}$

4. Si no se cumple lo anterior se regresa a 1 y se calcula $\vec{x}_{k+2}$.

¿Es el método de Newton un método de descenso? Observemos que la expresión (3.8) nos indica que en cada iteración se escoge como dirección a $\vec{d}_k$ a $-H_F(\vec{x}_k)^{-1}\vec{g}_k$. Newton es efectivamente un método de descenso ya que

$${}^t\vec{g}_k\vec{d}_k = -\vec{g}_k^t H_F^{-1}(\vec{x}_k)\vec{g}_k \leq 0$$

y esto se cumple siempre que $H_F(\vec{x}_k)$ sea una matriz semipositiva definida para cada iteración k. Esta última condición debe restringirse a que $H_F(\vec{x}_k)$ sea estrictamente positiva definida para garantizar que el sistema de ecuaciones a resolver admite una única solución. Por lo tanto, Newton convergerá siempre que la vecindad del mínimo que se seleccione sea suficientemente pequeña como para poder garantizar que el Hessiano, evaluado en cualquier punto de esa vecindad, es positivo definido.

El cálculo de la dirección requiere conocer el Hessiano, por lo que se clasifica a Newton como un método tipo Hessiano en contraste con el de descenso pronunciado que es un método de gradiente, pues sólo requiere esta información para calcular la dirección.

Por otro lado, el método de Newton se puede modificar para controlar el paso en cada iteración. En este caso el paso 2 se cambia a

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

con α_k que satisface

$$F(\vec{x}_k + \alpha_k\vec{d}_k) \leq F(\vec{x}_k + \alpha\vec{d}_k), \quad \forall\alpha \in \Re.$$

3.4.2. Caso cuadrático

Definición 3.4.1. *Diremos que un algoritmo tiene terminación cuadrática si converge para todo punto inicial $\vec{x}_0$ al mínimo de una función cuadrática en al menos n de pasos, donde n es el número de incognitas del problema.*

Lemma 3.4.2. *El método de Newton tiene terminación cuadrática.*

Supongamos que F es una función cuadrática de la forma $F(\vec{x}) = \frac{1}{2}\,\vec{x}^t A\vec{x} - \vec{x}^t\vec{b} + c$ con A matriz simétrica y positiva definida; dado $\vec{x}_0$ y $\alpha_0 = 1$ se tiene que

$$A\vec{d_0} = -\nabla F(\vec{x}_0) = -(A\vec{x}_0 - \vec{b}),$$

o sea $A(\vec{x}_1 - \vec{x}_0) + A\vec{x}_0 = \vec{b}$, lo que implica que

$$A\vec{x}_1 = \vec{b}.$$

Por lo que $\vec{x}_1$ satisface que el gradiente evaluado en $\vec{x}_1$: $\vec{g}_1 = A\vec{x}_1 - \vec{b}$ es igual a cero y por lo tanto es el mínimo de F. Así que el método de Newton converge globalmente al mínimo en una sóla iteración siempre que la matriz A sea positiva definida.

En el caso del método de descenso no puede asegurarse que para todo punto inicial $\vec{x}_0$ se convergerá en n iteraciones, por lo que no tiene terminación cuadrática. Vea el lector el ejemplo 3.1 de la sección anterior.

3.4.3. Caso general

Iniciaremos esta sección demostrando la convergencia del método de Newton cuando F es una función no lineal.

Teorema 3.4.3. *Sea $\vec{x}^*$ un mínimo local de una función F. Supóngase que*

1. *F es tres veces diferenciable en una vecindad de radio δ de $\mathbf{x}^*$: $V_\delta(x^*)$.*
2. *El Hessiano de F, $H_F(\vec{x}^*)$, es positivo definido.*
3. *El punto inicial $\vec{x}_0$ está en una $V_\rho(\vec{x}^*)$ para ρ pequeña y menor o igual que δ.*

Entonces la sucesión $\vec{x}_{k+1}$, definida por

$$\vec{x}_{k+1} = \vec{x}_k - H_F(\vec{x}_k)^{-1}\vec{g}_k,$$

converge a $\vec{x}^*$ y el orden de convergencia es dos.

Dado que $F \in C^3(V_\delta(\vec{x}^*))$, existe una constante positiva β_1 tal que para toda $\vec{x} \in V_\delta(\vec{x}^*)$ se cumple que

$$||H_F^{-1}(\vec{x})|| < \beta_1. \tag{3.10}$$

Además, usando la misma hipótesis y la serie de Taylor, se tiene que

$$\begin{aligned} \nabla F(\vec{x}^*) &= \nabla F(\vec{x}) + H_F(\vec{x})(\vec{x}^* - \vec{x}) \\ &+ \frac{1}{2}\ (\vec{x}^* - \vec{x})^t D_3 F(\theta\vec{x} + (1-\theta)\vec{x}^*)(\vec{x}^* - \vec{x}), \end{aligned}$$

para alguna $\theta \in (0,1)$. Por lo que existe $\beta_2 > 0$ tal que

$$||\nabla F(\vec{x}^*) - \nabla F(\vec{x}) - H_F(\vec{x})(\vec{x}^* - \vec{x})|| \leq$$

$$\frac{1}{2}\ ||D_3 F(\theta\vec{x} + (1-\theta)\vec{x}^*)||\ ||(\vec{x}^* - \vec{x})||^2 \leq \beta_2\ ||\vec{x}^* - \vec{x}||^2. \tag{3.11}$$

Supongamos que se elige $\rho > 0$ tal que para toda $\vec{x}$ estando en $V_\rho(\vec{x}^*)$ se cumple que

$$\beta_1\ \beta_2\ ||\vec{x}^* - \vec{x}|| < 1.$$

Para demostrar convergencia cuadrática hay que probar que existe una $K > 0$ tal que

$$\vec{e}_{k+1} \leq K\vec{e}_k^2.$$

Si en la k-ésima iteración del método de Newton $\vec{x}_k$ está en $V_\rho(\vec{x}^*)$, se tiene que

$$\vec{e}_{k+1} = ||\vec{x}_{k+1} - \vec{x}^*|| = ||\vec{x}_k - \vec{x}^* - H_F(\vec{x}_k)^{-1}\nabla F(\vec{x}_k)||.$$

Factorizando la inversa del Hessiano se tiene que

$$\vec{e}_{k+1} \leq ||H_F(\vec{x}_k)^{-1}||\ ||H_F(\vec{x}_k)(\vec{x}_k - \vec{x}^*) - \nabla F(\vec{x}_k)||$$

y usando el hecho que $\nabla F(\vec{x}^*) = 0$

$$e_{k+1} \leq ||H_F(\vec{x}_k)^{-1}||\ ||\nabla F(\vec{x}^*) - \nabla F(\vec{x}_k) - H_F(\vec{x}_k)(\vec{x}^* - \vec{x}_k)||.$$

Por último, al usar las desigualdades (3.10) y (3.11) se obtiene que

$$\vec{e}_{k+1} \leq ||H_F(\vec{x}_k)^{-1}||\beta_2||\vec{x}^* - \vec{x}_k||^2 \leq \beta_1\ \beta_2\ ||\vec{x}^* - \vec{x}_k||^2.$$

Por lo que el método converge y la convergencia es cuadrática para una vecindad de diámetro menor que ρ. La suposición que $\vec{x}_k \in V_\rho(\vec{x}^*)$ se cumple por la hipótesis que se hizo de que $\vec{x}_0$ estuviera lo suficientemente cerca de $\vec{x}^*$ y de la manera en que se determinó ρ.

El teorema anterior nos permite garantizar convergencia cuadrática siempre que H_F sea positiva definida para alguna vecindad de radio ρ de $\vec{x}^*$. Por ello, la convergencia en el caso general es local pues depende fuertemente de la ρ que se seleccione. El problema principal al que se enfrenta uno al tratar de aplicar Newton es el determinar una ρ que nos garantice las hipótesis del teorema anterior.

Recordemos que una matriz positiva definida es invertible y que su inversa es positiva definida. ¿Cómo checar que una matriz es positiva definida al mismo tiempo que se resuelve el sistema? Por medio del método de Cholesky. Si al factorizar la matriz $H_F(\vec{x}_k)$ por medio de Cholesky el algoritmo falla, la matriz no es positiva definida.

3.4.4. Ejemplos

1. Aplicar el método de Newton para aproximar el mínimo de $F(x, y) = 3x^2 + y^2 - x^4 - 12$ tomando como $\vec{x}_0 = (1/3, 1)$.

Tabla 3.4

i	$\vec{x}_i$	$\|\|\vec{g}_i\|\|$
0	(1/3,1.)	2.725
1	(-.405,0)	2.164
2	(-.205,0)	1.19
3	(-.064,0)	.3833
4	(-.0074,0)	.0446
5	(-.0001,0)	.00065
6	(-2.3×10^{-8},0)	1.42×10^{-7}

   En este caso la convergencia es cuadrática porque $\vec{x}_0$ se encuentra en la región de convergencia de Newton que es de radio 1/2. Compare el lector para este ejemplo el desempeño de Newton con el de descenso pronunciado.

2. Usar Newton para determinar el mínimo de $F(x, y) = 16x^4 - 16x^3 + 6x^2 - x + \frac{1}{16} + 3x^2y^2$, con $\vec{x}_0 = (1, 1)$.

Tabla 3.5

i	$\mathbf{x_i}$	$\|\|\mathbf{g}_i\|\|$	R_i
0	(1.,1.)	.687677	12.8097
1	(.8327,-.1607)	.6876	3.8339
2	(.6412,.0138)	.09333	3.8339
3	(.5108,.0009)	.0417	1.1360
4	(.4239,.0001)	.02897	.3366
5	(.3659,.00004)	.01870	.099
6	(.3272,.00001)	.01115	.029
7	(.3015,.000003)	.00619	.0087
8	(.2843,.000001)	.0007	.0002
15	$(.2529, 3.6 \times 10^{-9})$	3.7×10^{-5}	1.5×10^{-6}

En este caso, la convergencia de Newton no es cuadrática y esto se debe a que en la vecindad de radio uno del mínimo no se cumple las condiciones del teorema 3.4.3. La matriz es semidefinida positiva en el origen e indefinida en otros puntos.

En suma el método de Newton tiene como gran ventaja la de tener convergencia cuadrática pero se consigue a un alto costo: por un lado, se requiere calcular en cada iteración el Hessiano y resolver un sistema de ecuaciones. Otro inconveniente más es el determinar una vecindad ρ donde se garantice convergencia. Para ilustrar estas dificultades, trate el lector de aproximar la solución del ejemplo 1 iniciando con $\vec{x}_0 = (1/2, 1)$ o el origen. La matriz en ese caso no es positiva definida. Otra dificultad que puede presentarse es que la función a minimizar no sea dos veces diferenciable en la vecindad del mínimo. Con objeto de superar estos problemas se sugiere hacer las siguientes modificaciones al método de Newton.

3.4.5. Modificaciones al método de Newton

1. Supongamos que iniciamos con un punto $\vec{x}_0$, en una vecindad V_δ en la que el Hessiano sea positivo definido para todo punto en ella. ¿Que hacer si en la k-ésima iteración el Hessiano se vuelve indefinido o semidefinido positivo? Lo que se sugiere es resolver un sistema distinto a (3.8), construyendo una matriz M_k de la forma

$$M_k = H_F(\vec{x}_k) + \varepsilon I, \tag{3.12}$$

con ε un real positivo. Observemos que si $\varepsilon \to 0$ la matriz M_k converge al Hessiano pero, si $\varepsilon \to \infty$ la dirección converge al gradiente. El punto delicado de esta modificación es determinar la ε adecuada para que la matriz M_k sea positiva definida. Hay varias estrategias, entre ellas usar la factorización de Cholesky del Hessiano.

2. Otra dificultad que puede presentarse en la aplicación del método de Newton es que la función a minimizar no sea dos veces diferenciable o que el cálculo de sus segundas derivadas sea sumamente costoso. En este caso las derivadas pueden aproximarse usando diferencias finitas: suponga que se tiene $\vec{x}_k$ entonces

$$H_F(\vec{x}_k)_{ij} \approx \frac{\nabla F(\vec{x}_k + h\vec{e}_j)_i - \nabla F(\vec{x}_k)_i}{h},$$

En este caso se introduce un error debido a la discretización y es del orden de h.

Debido a las dificultades que presenta Newton, muchos trabajos se han publicado respecto a cómo obtener un método que tenga convergencia superior al de descenso pronunciado sin requerir en cada paso del cálculo del Hessiano. Estas investigaciones han dado lugar a los métodos Cuasi-Newton, los cuales son los métodos más usados en la práctica. Para aquellos lectores que les interese conocer estos métodos se sugiere el libro de Dennis [2], Fletcher [4], Gill et al [3] y el Scales [11].

3.5. Método de gradiente conjugado

El método de gradiente conjugado tiene como objetivo tener terminación cuadrática.

Definición 3.5.1. *Se dice que $\{\vec{d}_k\}_{k=1}^n$ son vectores mutuamente conjugados respecto a una matriz G simétrica y positiva definida si*

$$\vec{d}_k^{\,t}\, G\, \vec{d}_j = 0 \quad j \neq k. \tag{3.13}$$

Ejemplo

Sea

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

observemos que la matriz G es simétrica y positiva definida, $(1,0)$ y $(1/2,1)$ son mutuamente conjugados respecto a G.

Dada una matriz G positiva definida y simétrica y un vector $\vec{v}$ ¿cómo se construye un conjunto de vectores conjugados respecto a la matriz G ? Por un procedimiento similar al de Gramm-Schimdt que nos permite construir, a partir de un conjunto de vectores linealmente independientes, un conjunto de vectores ortogonales. Observemos que ser ortogonales es lo mismo que ser conjugados cuando la matriz G es la identidad. El procedimiento es el siguiente: dado $\vec{d_1} = \vec{v}_1$ y $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ linealmente independientes desde $i = 1, \ldots, n$

$$\vec{d}_{i+1} = \vec{v}_{i+1} - \sum_{k=1}^{i} \frac{\vec{v}_{i+1}^t \; G \; \vec{d}_k}{\vec{d}_k^t \; G \; \vec{d}_k} \; \vec{d}_k.$$

Es fácil probar que $\{\vec{d_1}, \ldots, \vec{d_n}\}$ forma un conjunto mutuamente conjugado respecto a la matriz A.

Apliquemos este procedimiento para construir un conjunto de vectores conjugados respecto a la matriz G

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}.$$

Tomemos la base canónica $\{\vec{e}_i\}$ de $\Re^4$; $\vec{d_1} = \vec{e}_1$ y $\vec{d_2} = \vec{e}_2 - 1/2 \; \vec{d_1} = (1/2, 1, 0, 0)$,

$$\vec{d_3} = \vec{e}_3 + \frac{2}{3}\vec{d_2} = \frac{1}{3}(1,2,1,0)$$

y

$$\vec{d_4} = \vec{e}_4 + \frac{3}{4}\vec{d_3} = \frac{1}{4}(1,2,3,4).$$

Lemma 3.5.2. *Todo conjunto de vectores conjugados a una matriz G son linealmente independientes.*

Tomemos una combinación lineal de n vectores $\vec{v}_i$ mutuamente conjugados respecto a una matriz G y supongamos que existen contantes $c_i \in \Re$ tales

que

$$\sum_{i=1}^{n} c_i \vec{v}_i = 0.$$

Entonces si denotamos como $<,>$ el producto escalar usual en $\Re^n$,

$$0 =< \sum_{i=1}^{n} c_i G\vec{v}_i, \vec{v}_j >= \sum_{i=1}^{n} c_i \ \vec{v}_j^t \ G \ \vec{v}_i = c_j \ \vec{v}_j^t \ G \ \vec{v}_j$$

y esto implica que $c_j = 0$ para toda j desde uno hasta n. Por lo tanto, el conjunto $\{\vec{v}_1, \ldots, \vec{v}_n\}$ son linealmente independientes. De este lema se desprende que dada una matriz $n \times n$ a lo más hay n vectores conjugados respecto a la matriz G.

La idea del método de gradiente conjugado es tomar un método de descenso en que las direcciones son conjugadas respecto al Hessiano de la función cuadrática que se desea minimizar. Es decir, si F es una función cuadrática de la forma:

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

El algoritmo de descenso con direcciones conjugadas $\vec{d}_k$ respecto a la matriz A es de la forma:

- Dado $\vec{x}_0 \in V_\delta(\mathbf{x}^*)$
- $$\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{d}_k,$$

  con α_k que satisface

  $$\alpha_k = -\frac{\vec{g}_k^t \ \vec{d}_k}{\vec{d}_k^t A d_k},$$

  con $\vec{d}_k^t \ A \ \vec{d}_j = 0$ si $j \neq k$ para $j = 1, \ldots, k-1$ y $\vec{d}_k^t \ A \ \vec{d}_k > 0$.

A estos algoritmos se les conoce con el nombre de algoritmos de dirección conjugada.

Lemma 3.5.3. *El algoritmo anterior tiene terminación cuadrática si hay búsqueda lineal exacta y F es cuadrática.*

Definamos como

$$\Delta\vec{x}_k = \vec{x}_{k+1} - \vec{x}_k = \alpha_k \vec{d}_k. \tag{3.14}$$

Si $\Delta\vec{g}_k = \vec{g}_{k+1} - \vec{g}_k$ entonces por la serie de Taylor

$$\vec{g}_{k+1} = \vec{g}_k + A\ \Delta\vec{x}_k$$

y

$$\Delta\vec{g}_k = A\ \Delta\vec{x}_k = \alpha_k\ A\ \vec{d}_k. \tag{3.15}$$

Supongamos que estamos en la k-ésima iteración y aún no determinamos el mínimo, escriba a $\vec{g}_k$ como

$$\begin{aligned} \vec{g}_k &= \vec{g}_k - \vec{g}_{k-1} + \vec{g}_{k-1} - \cdots - \vec{g}_{j+1} + \vec{g}_{j+1} \\ &= \vec{g}_{j+1} + \sum_{i=j+1}^{k-1} \Delta\vec{g}_i, \end{aligned}$$

para $j = 0, 1, \ldots, k-1$. Multiplicando por $\vec{d}_j$ se tiene

$$\vec{d}_j^t \vec{g}_k = \vec{d}_j^t \vec{g}_{j+1} + \sum_{i=j+1}^{k-1} \vec{d}_j^t\ \Delta\vec{g}_i.$$

Substituyendo $\Delta\vec{g}_i = \alpha_i\ A\ \vec{d}_i$ se tiene que

$$\vec{d}_j^t \vec{g}_k = \vec{d}_j^t \vec{g}_{j+1} + \sum_{i=j+1}^{k-1} \alpha_i \vec{d}_j^t A \vec{d}_i = \vec{d}_j^t \vec{g}_{j+1},$$

por ser las direcciones conjugadas desde $j = 0$ hasta $k-1$.

Por otro lado, recordemos que para el caso cuadrático, si hay búsqueda lineal exacta, se tiene que

$$\frac{dF(\vec{x}_{k+1})}{d\alpha} = \vec{d}_k^t \vec{g}_{k+1} = 0$$

y esto es válido para toda k. Por lo tanto $\vec{d}_j^t \vec{g}_k = 0$ para toda j desde 1 hasta $k-1$.

Supongamos que $k = n$ entonces se cumple que

$$\vec{d}_j^t \vec{g}_n = 0 \quad j = 0, \ldots, n-1.$$

Como $\vec{g}_n \in \Re^n$ y es ortogonal a n vectores linealmente independientes, la única posibilidad es que $\vec{g}_n$ sea cero. Por lo que el mínimo debe ser $\vec{x}_n$ y hemos demostrado que a lo más en n iteraciones converge al mínimo. □

De esta forma se liga el concepto de vectores conjugados respecto a una matriz A y la terminación cuadrática. Resta para tener un algoritmo que determine en cada paso k una dirección conjugada respecto a las anteriores.

3.5.1. Algoritmo de Gradiente Conjugado

1. Dado $\vec{x}_0 \in V_\delta(\vec{x}^*)$ y $\vec{d_0} = -\vec{g}_0$
2. Para $k = 0, 1, ..$
$$\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{d_k},$$
con
$$\alpha_k = -\frac{\vec{g}_k^t \vec{d_k}}{\vec{d_k^t}\ A\ \vec{d_k}}. \tag{3.16}$$
3. La dirección $\vec{d}_{k+1}$ se calcula por
$$\vec{d_{k+1}} = -\vec{g}_{k+1} + \beta_k \vec{d_k}, \tag{3.17}$$
con
$$\beta_k = \frac{\vec{g}_{k+1}^t\ A\ \vec{d_k}}{\vec{d_k^t}\ A\ \vec{d_k}}. \tag{3.18}$$
4. Si $||\vec{g}_{k+1}|| \leq rtol$ y $\frac{||\vec{x}_{k+1} - \vec{x}_k||}{||\vec{x}_k||} \leq rtol$ entonces se toma a $\vec{x}^* \approx \vec{x}_{k+1}$
5. Si no se cumple lo anterior se regresa a 2 y se calcula $\vec{x}_{k+2}$.

Apliquemos el algoritmo para determinar el mínimo de $F(x, y) = x^2 + xy + y^2$. El gradiente es $\nabla F(x, y) = (2x + y, x + 2y)$. Tomemos como $\vec{x}_0$ a $(2, -1)$ entonces $\vec{g}_0 = (3, 0)$, $\alpha_0 = 1/2$ y $\vec{x}_1 = (1/2, -1)$. Además $\beta_1 = 1/4$, $\vec{d_1} = (-3/4, 3/2)$ y $\alpha_1 = 2/3$, por lo que $\vec{x}_2 = (0, 0)$ que es la solución. Observemos que en dos iteraciones alcanzamos el mínimo.

Teorema 3.5.4. *El algoritmo de gradiente conjugado es un algoritmo de direcciones conjugadas.*

Este resultado se demuestra a partir del siguiente lema; para denotar que un conjunto de n vectores genera un espacio se usará la notación:

$$Gen\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}.$$

Lemma 3.5.5. *Las siguientes relaciones se cumplen en la k-ésima iteración*

1.
$$Gen\{\vec{g}_0, \vec{g}_1, \ldots, \vec{g}_k\} = Gen\{\vec{g}_0, A\vec{g}_0, \ldots, A^k\vec{g}_0\}. \tag{3.19}$$

2.
$$Gen\{\vec{d_0}, \ldots \vec{d_k}\} = Gen\{\vec{g}_0, A\vec{g}_0, \ldots, A^k\vec{g}_0\}. \tag{3.20}$$

3.
$$\vec{d_j^t} \;\; A \;\vec{d_k} = 0 \quad j = 0, \ldots k-1. \tag{3.21}$$

Observemos que el α_k en la expresion (3.16) es el valor exacto de la búsqueda lineal en el caso cuadrático.

Demostremos el lema por inducción. Es muy fácil ver que se cumple para $k = 0$. Supongamos entonces que se cumple para $n = k$ y probémoslo para $n = k + 1$ para la igualdad (3.19), es decir que $\vec{g}_{k+1} \in \{\vec{g}_0, A\vec{g}_0, \ldots, A^{k+1}\vec{g}_0\}$.

Por la igualdad (3.14) se tiene que

$$\vec{g}_{k+1} = \vec{g}_k + \alpha_k A\vec{d_k}.$$

Los vectores $\vec{g}_k$ y $\vec{d_k}$ están en $\{\vec{g}_0, A\vec{g}_0, \ldots, A^k\vec{g}_0\}$ por hipótesis de inducción, por lo que

$$A \;\vec{d_k} \in \;\; Gen\{A\vec{g}_0, A^2\vec{g}_0, \ldots, A^{k+1}\vec{g}_0\}$$

y por lo tanto

$$\vec{g}_{k+1} \in Gen\{A\vec{g}_0, A^2\vec{g}_0, \ldots, A^{k+1}\vec{g}_0\}.$$

Así que

$$Gen\{\vec{g}_0, \vec{g}_1, \ldots, \vec{g}_{k+1}\} \subset Gen\{\vec{g}_0, A\vec{g}_0, \ldots, A^{k+1}\vec{g}_0\}.$$

Para demostrar la igualdad, observemos primero que por el lema 3.5.3, $\vec{d_j^t}\vec{g}_{k+1} = 0$ para $j = 0, \ldots, k$. Por hipótesis de inducción como la igualdad (3.19) se cumple para i desde cero hasta k, entonces $\vec{g}_{k+1}$ es ortogonal al subespacio

$$Gen\{\vec{g}_0, A\vec{g}_0, \ldots, A^k\vec{g}_0\}$$

y por lo tanto para que se cumpla (3.19) debe existir una $c \neq 0$ tal que

$$\vec{g}_{k+1} = c\ A^{k+1}\ \vec{g}_0$$

lo que implica que

$$A^{k+1}\ \vec{g}_0 \in Gen\{\vec{g}_0, \vec{g}_1, \ldots, \vec{g}_{k+1}\}.$$

Así se concluye que se cumple la igualdad (3.19).

Pasemos a demostrar la segunda igualdad haciendo uso de la igualdad anterior (3.19) para $k+1$ y que la igualdad (3.20) se cumple para k. La nueva dirección $\vec{d}_{k+1}$ se calcula por

$$\vec{d}_{k+1} = -\vec{g}_{k+1} + \beta\vec{d}_k.$$

Por hipótesis de inducción $\vec{d}_k \in Gen\{\vec{g}_0, \vec{g}_1, \ldots, \vec{g}_k\}$ y a su vez $\vec{g}_{k+1}$ está en

$$Gen\{\vec{g}_0, \vec{g}_1, \ldots, \vec{g}_k, \vec{g}_{k+1}\}.$$

Para demostrar la igualdad, sabemos por el inciso anterior que

$$A^{k+1}\vec{g}_0 \in Gen\{\vec{g}_0, \vec{g}_1, \ldots, \vec{g}_k, \vec{g}_{k+1}\},$$

entonces por hipótesis de inducción

$$A^{k+1}\vec{g}_0 \in Gen\{\vec{d}_0, \vec{d}_1, \ldots, \vec{d}_k, \vec{g}_{k+1}\}$$

y usando la relacion (3.17) concluimos que

$$A^{k+1}\vec{g}_0 \in Gen\{\vec{d}_0, \vec{d}_1, \ldots, \vec{d}_k, \vec{d}_{k+1}\}.$$

Por último para demostrar que las direcciones son conjugadas, retomemos la igualdad (3.17) y multipliquemos ambos lados de la igualdad por la matriz A

$$A\vec{d}_{k+1} = -A\vec{g}_{k+1} + \beta A\vec{d}_k,$$

y multiplicando por $\vec{d}_j^t$ para $j = 0, \ldots, k-1$ se obtiene

$$\vec{d}_j^t\ A\ \vec{d}_{k+1} = -\vec{d}_j^t\ A\ \vec{g}_{k+1}$$

porque, por la hipótesis de inducción, las direcciones son conjugadas para $j = 1, \ldots, k$. Usemos que la matriz A es simétrica, entonces

$$\vec{d}_j^t\ A\ \vec{g}_{k+1} = \vec{g}_{k+1}^t\ A\ \vec{d}_j$$

y como

$$A\ \vec{d}_j \in \{A\vec{g}_0, A^2\vec{g}_0, \ldots, A^{j+1}\vec{g}_0\}$$

al usar de nuevo que $\vec{g}_{k+1}$ es ortogonal a este último conjunto para $j = 0, \ldots, k-1$ se tiene que

$$\vec{d}_j^{\,t}\ A\ \vec{g}_{k+1} = 0$$

y por lo tanto $\vec{d}_{k+1}$ es conjugada a todas las direcciones desde $\vec{d}_0$ hasta $\vec{d}_{k-1}$. Para demostrar que también esto se cumple para $\vec{d}_k$ basta con substituir el valor de β_k en la igualdad (3.17). Por lo tanto el algoritmo es de direcciones conjugadas y por ende tiene terminación cuadrática. □

Observemos que α y β pueden expresarse en forma más sencilla y más barata cuando hay búsqueda lineal exacta. Retomando la igualdad (3.16) se tiene que

$$\alpha_k = \frac{-\vec{g}_k^t\vec{d}_k}{\vec{d}_k^t A\vec{d}_k} = \frac{-\vec{g}_k^t(-\vec{g}_k + \beta_k\vec{d}_{k-1})}{\vec{d}_k^t A\vec{d}_k} = \frac{\vec{g}_k^t\vec{g}_k}{\vec{d}_k^t A\vec{d}_k},$$

ya que $\vec{g}_k^t\vec{d}_{k-1} = 0$. Substituyamos esta expresión en (3.18)

$$\beta_k = \alpha_k\frac{\vec{g}_{k+1}^t A\vec{d}_k}{\vec{g}_k^t\vec{g}_k} = \frac{\vec{g}_{k+1}^t\ \Delta\vec{g}_k}{\vec{g}_k^t\ \vec{g}_k}. \tag{3.22}$$

Esta expresión se puede simplificar si recordamos que $\vec{g}_{k+1}$ y $\vec{g}_k$ son ortogonales entonces

$$\beta_k = \frac{\vec{g}_{k+1}^t\ \vec{g}_{k+1}}{\vec{g}_k^t\ \vec{g}_k}. \tag{3.23}$$

Dependiendo de que forma se calcule β el algoritmo de gradiente conjugado respectivo recibe un nombre distinto dando crédito a quien primero lo utilizó de esa manera. Por ejemplo cuando se usa (3.22) se conoce como la versión de Polak y Ribiére y la versión que usa (3.23) es la de Fletcher y Reeves.

La terminación cuadrática del método de gradiente conjugado se cumplirá siempre que no haya errores de redondeo. Ilustremos el comportamiento de este método aplicándolo para determinar el mínimo del ejemplo de la sección anterior. En ese caso

$$F(x, y) = .01x^2 + .01y^2 + .007xy - 485x - 675y + 400,000.$$

El gradiente de F está dado por

$$\nabla F(x, y) = (.02x + .007y - 485,\ .007x + .02y - 675)$$

y la solución exacta con dos cifras decimales es $(14,173.789, 28,789.17)$. Supongamos que $\vec{x}_0 = (10000, 20000)$ entonces

Tabla 3.6

i	$\vec{x}_i$	$\|\|\vec{g}_i\|\|$	R_i
0	(10,000,20,000)	205.09	-
1	(15 450.5,27 706.6)	.2975	22
2	(14 173.79,28 789.17)	1×10^{-5}	5×10^{-2}
3	(14 173.79,28 789.17)	0.0	0.0

En este caso en dos iteraciones se alcanza el mínimo pero debido a la forma en que está el algorimo lo detecta hasta en la tercera iteración. ¿Qué sucede en el caso no lineal, se puede aplicar este método? La respuesta es afirmativa, pero hay que hacer notar que para funciones no lineales la terminación cuadrática se pierde, por lo que a las n iteraciones habremos agotado el conjunto de direcciones conjugadas, es decir cualquier otra será una combinación lineal de las anteriores. Para evitar este problema se reinicializa el proceso cada n iteraciones tomando como dirección inicial a $-\vec{g}_n$. Entonces las modificaciones que hay que hacerle al algoritmo para el caso no lineal son las siguientes:

3.5.2. Algoritmo gradiente conjugado: caso no lineal

1. Dado $\vec{x}_0 \in V_\delta(\vec{x}^*)$, y una tolerancia $rtol > 0$; $\vec{d}_0 = -\vec{g}_0$
2. Para $k = 0, 1, \ldots$
$$\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{d}_k,$$
con $\alpha_k \in \Re$ que satisface
$$F(\vec{x}_k + \alpha_k \vec{d}_k) \leq F(\vec{x}_k + \alpha \vec{d}_k), \quad \forall \alpha \in \Re.$$
3. La dirección $\vec{d}_{k+1}$ se calcula por
$$\vec{d}_{k+1} = -\vec{g}_{k+1} + \beta_k \vec{d}_k,$$
con
$$\beta_k = \frac{\vec{g}_{k+1}^t \; H_F(\vec{x}_k) \; \vec{d}_k}{\vec{d}_k^t \; H_F(\vec{x}_k) \; \vec{d}_k}, \text{si } k < n.$$
Si $k \geq n$, $\beta_k = 0$ y volver a iniciar a partir de $k = 0$.

4. Si $||\vec{g}_{k+1}|| \leq rtol$ y $\frac{||\vec{x}_{k+1}-\vec{x}_k||}{||\vec{x}_{k+1}||} \leq rtol$ entonces se toma a $\vec{x}^* \approx \vec{x}_{k+1}$

5. Si no se cumple lo anterior se regresa a 2 y se calcula $\vec{x}_{k+2}$.

Para que el algoritmo no entre en un ciclo infinito se requiere de acotar el número máximo de iteraciones que se lleven a cabo. Apliquemos este algoritmo a los ejemplos que se presentan en la sección 3.3.2.

1. Aplicar el método de gradiente conjugado para aproximar el mínimo de $F(x, y) = 3x^2 + y^2 - x^4 - 12$ tomando como $\vec{x}_0 = (1/3, 1)$.

Tabla 3.6

i	$\vec{x}_i$	$\|\|\vec{g}_i\|\|$
0	(1/3,1.)	2.725
1	(-.18429,.33447)	1.3314
2	(-.06139,.19170)	.5297
3	(-.03273,.09295)	.270
4	(-..01518,.04768)	.1318
5	(-.00812,.023)	.067
6	(.0037,.011)	.032
7	(.00202,.0058)	.016

En la séptima iteración se tienen dos cifras significativas. Comparando con Newton observamos la convergencia es más lenta y que es lineal pero es mucho más rápida que la de descenso pronunciado. En ambos ejemplos se usa la expresión (3.16) para estimar el valor de α.

2. Usar Newton para determinar el mínimo de $F(x, y) = 16x^4 - 16x^3 + 6x^2 - x + \frac{1}{16} + 3x^2y^2$, con $\vec{x}_0 = (-1, 1)$.

Tabla 3.7

i	$\vec{x}_i$	$\|\|\vec{g}_i\|\|$	R_i
0	(1.,1.)	.687677	12.8097
1	(.7124,.9477)	10.57	24.67
2	(.8208,-.01419)	11.9064	1.17
3	(.6305,-.01327)	3.5279	.301
4	(.6305,-.0038))	3.5279	.014
5	(.5037,-.0034)	1.0453	.005
10	(.3251,-.00009)	.027	.115
13	(.2722,-.00001)	7.07×10^{-4}	.04

El error relativo en la treceava iteración es de .08. La convergencia es mejor que descenso pronunciado, pero menos buena que Newton. En este ejemplo el criterio de la sucesión es un mejor indicador de cuál es el error relativo que el gradiente.

.

Por último observemos que el método de gradiente conjugado es un método de descenso ya que por la igualdad (3.17)

$$\vec{d}_k = -\vec{g}_k + \beta_k \vec{d}_{k-1}.$$

Multipliquemos ambos lados de la igualdad por el vector $\vec{g}_k$

$$\vec{g}_k^t \ \vec{d}_k = -\vec{g}_k^t \ \vec{g}_k + \beta_k \vec{g}_k^t \ \vec{d}_{k-1} = -\vec{g}_k^t \ \vec{g}_k \leq 0$$

porque $\vec{g}_k^t \ \vec{d}_{k-1} = 0$ al haber búsqueda lineal exacta.

3.6. Ejercicios

1. ¿Qué tipo de convergencia tiene el siguiente algoritmo para determinar la raíz cuadrada de un número $a > 1$?

$$x_{k+1} = \frac{1}{2}(x_k + \frac{a}{x_k}).$$

2. Considere los siguientes problemas cuadráticos: Determine el mínimo de

   - $f(x, y) = x^2 + y^2 - 3x + 2y + 1$, con $\vec{x}_0 = (0, 0)$.
   - $f(x, y) = 2x^2 - xy + 15y^2 - 2x + 3y + 4$, con $\vec{x}_0 = (0, 0)$.

   Determine cuál es la rapidez de convergencia del método de descenso pronunciado cuando se aplica a los problemas anteriores. ¿Cuántas iteraciones se requieren para tener dos cifras significativas? Aplique el método para el primer caso y compare el número de iteraciones que se necesitan con búsqueda lineal exacta, para obtener la aproximación deseada. ¿Qué observa? Aplique el criterio de la sucesión y del gradiente y compare con el error relativo. ¿Qué sucede si aplica interpolación cuadrática en el paso de la búsqueda lineal si se usa como tolerancia a 10^{-2}?

3. Aplique descenso pronunciado con búsqueda lineal exacta para determinar el punto crítico de $F(x,y) = x^2 - 2xy - \frac{1}{2}(y^2 - 1)$. ¿Converge? ¿Por qué?

4. Aplique descenso pronunciado para determinar el mínimo de $F(x,y) = \frac{1}{2}x^2 - 2xy + 2y^2$ tomando como $x_0 = (-1,1)$. ¿Converge? ¿Por qué?

5. Puede aplicarse Newton para aproximar la solución del ejercicio 3 (justifique su respuesta).

6. Aplique Newton para determinar el mínimo de la segunda función del ejercicio 1. ¿Tiene terminación cuadrática?

7. Considere la función $F(x,y) = 16x^4 - 16x^3 + 6x^2 - x + \frac{1}{16} + 3x^2y^2$. ¿Puede aplicar Newton iniciando en $\vec{x}_0 = (1/2, 1)$? Modifique el Hessiano a $H_F + \mu I$ para aplicar este algoritmo. ¿Que valor seleccionó de μ. ¿Por qué? Compruebe que también se puede usar los valores propios para determinar el valor de μ . Proponga un algoritmo que determine μ tomando en cuenta el valor de los valores propios.

8. Construya un conjunto de vectores conjugados respecto a la matriz A.

$$\begin{bmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 20 \end{bmatrix}.$$

9. Demuestre que el método de gradiente conjugado tiene convergencia lineal.

10. Justifique porque el método de gradiente conjugado puede aplicarse para aproximar la solución de un sistema lineal $A\vec{x} = \vec{b}$ con A una matriz n por n cualquiera que sea simétrica y positiva definida. ¿En cuántas iteraciones convergerá? ¿Por qué?

11. Aplicar el algoritmo de gradiente conjugado para aproximar la solución del sistema $Ax = b$

$$\begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix}$$

con ${}^t b = (1, 0, 0, 0, 0)$.

12. Aplicar gradiente conjugado a las funciones del ejercicio 1.

13. Sea $F(x, y) = 100\epsilon^{-xy} + x + 10$. Determine si tiene un punto mínimo en $A = \{(x, y)|\ x, y \geq 0\}$. Si se aplica descenso pronunciado, aproximadamente ¿cuántas iteraciones se requieren para tener un error menor a 10^{-1}?

Capítulo 4

Optimización con restricciones

4.1. 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

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

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

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

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 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 restricciones de igualdad es de la forma

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

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

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

Definición 4.1.1. *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 4.1.2. *Un punto $\vec{x}^*$ se dice que es un mínimo relativo 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 4.1.3. *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 $J_h = C$ con C 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

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

$N(\vec{x})$ 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

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

Definición 4.1.4. *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.

4.2. 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

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

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 4.2.1. *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*

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

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 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

$$\vec{x}(t) = \vec{x}^* + \vec{y}t + J_h^t(\vec{x}^*)\ \vec{u}(t) \tag{4.1}$$

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

$$\vec{h}(\vec{x}^* + \vec{y}t + J_h^t(\vec{x}^*)\vec{u}(t)) = 0. \tag{4.2}$$

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 (4.2) ¿Tiene solución este sistema para toda t en el intervalo $[0, t_0]$, para alguna t_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 única $\vec{u}(t)$ 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 4.2.2. *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*

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

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 4.2.3. *Si F y $\vec{h}$ son funciones continuamente diferenciables en un abierto que contenga a Ω y $\vec{x}^*$ un punto extremo local 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*

$$\nabla F(\vec{x}^*) + J_h(\vec{x}^*)^t\vec{\lambda} = 0. \tag{4.3}$$

Por el lema anterior $\nabla F(\vec{x}^*)$ es ortogonal a todo vector en el plano tangente a la superficie Ω y por el teorema 4.2.1 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

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

□

Observemos que, como en el caso lineal, la expresión (4.3) 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 4. 1 de esta introducción:

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

donde $T(x,y,z) = xz+y^2+600$ y $\Omega = \{(x,y,z) \mid h(x,y,z) = x^2+y^2+z^2-4 = 0\}$.

Apliquemos las condiciones de primer orden a este problema entonces existe $\lambda \in \Re$ 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 da 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 4.2.4. *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 $\vec{y} \in N(\vec{x}^*)$.

La primera parte se demuestra por el Teorema 4.2.3. 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

$$\vec{x}'(0)^t H_F(\vec{x}^*)\vec{x}'(0) + \nabla F(\vec{x}^*)\vec{x}'(0) \geq 0. \tag{4.4}$$

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

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

y al derivar dos veces esta expresión y al evaluarla en $t = 0$ se tiene que

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

Sumando respecto a i

$$\vec{x}'^t(0)\sum_{i=1}^{m}\lambda_i H_{h_i}(\vec{x}^*)\vec{x}'(0) + J^t_{\vec{h}}(\vec{x}^*)\vec{\lambda}\vec{x}''(0) = 0. \tag{4.5}$$

Recordemos que por el teorema 4.2.3

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

y sumando (4.4) y (4.5) 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 $N(\vec{x}^*)$. □

Teorema 4.2.5. *Supóngase que hay un punto $\vec{x}^*$ en Ω y una $\vec{\lambda} \in \Re^m$ tal que*

$$\nabla F(\vec{x}^*) + J^t_h(\vec{x}^*)\vec{\lambda} = 0. \tag{4.6}$$

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 local estricto entonces existe $\vec{y} \in V_\delta(\vec{x}^*)$ 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 $\vec{s}^* \in \Omega$. 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

$$0 = \sum_{i=1}^{m} \lambda_i h_i(\vec{y}_k) = \sum_{i=1}^{m} \lambda_i (h_i(\vec{x}^*) + \delta_k \ \nabla h_i^t(\vec{x}^*)\vec{s}_k + \frac{\delta_k^2}{2} \ \vec{s}_k^t H_{h_i}(\eta_i)\vec{s}_k). \quad (4.7)$$

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 (4.7) 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 (4.6) 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 $N(\vec{x}^*)$. □

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 (4.6) 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

$$\vec{y}^t L(-\frac{1}{2})\vec{y} = b^2 - 4a^2 < 0$$

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 raźonamiento 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

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

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

$$A\vec{x} + C\vec{\lambda} = \vec{b} \tag{4.8}$$
$$C^t\vec{x} = \vec{e} \tag{4.9}$$

Lemma 4.2.6. *El sistema (4.8) 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 (4.8) 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

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

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 como C es una matriz de rango completo, C^tC es una matriz no singular, por lo que la única solución es $\vec{\lambda} = \vec{0}$. □

La solución del sistema (4.8) se puede reducir a resolver el siguiente sistema de ecuaciones :

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

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 (4.8) 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 $\vec{\overline{r}}$ 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

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

La solución $\vec{w}^*$ está dada por

$$[\Sigma]\vec{w} = -\lambda_1 \vec{\overline{r}} - \lambda_2 \vec{1}$$

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 $N(\vec{w}^*)$.

4.3. Caso de restricciones de desigualdad

Consideremos el siguiente problema

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

En la Figura 4.1 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.

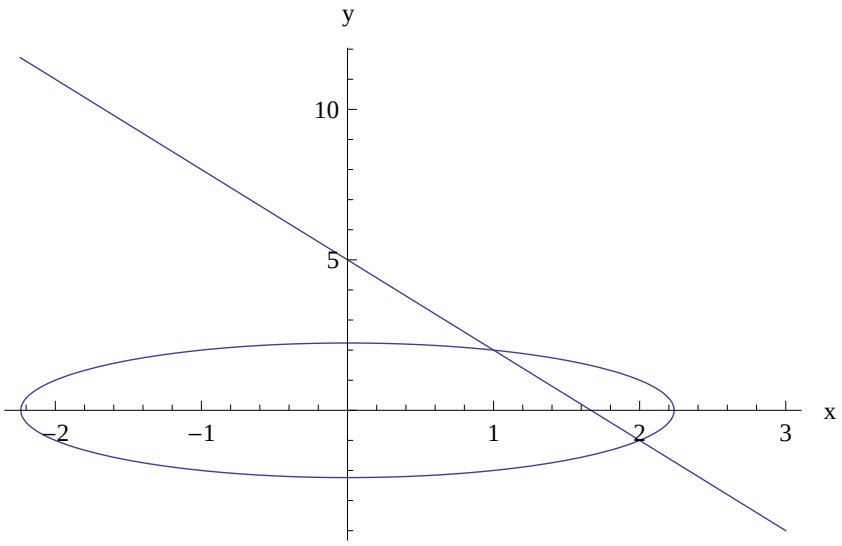

Figura 4.1: Región factible Ω.

Al graficar la región admisible junto con las curvas de nivel de F, ver la Figura 4.3 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 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

$$\Omega = \{\vec{y} \in \Re^n |\ h_j(\vec{y}) = 0,\ j = 1, \ldots, m,\ f_j(\vec{y}) \leq 0,\ j = 1, \ldots, s\} \qquad (4.10)$$

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

Dado un punto $x \in \Omega$ se definen las restricciones activas en este punto como aquellas para las cuales se satisface la igualdad. Denotemos como $I(\vec{x})$ a los índices asociados a las restricciones activas en $\vec{x}$. Entonces el espacio $N(\vec{x})$ se define para el caso e las desigualdades

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

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.

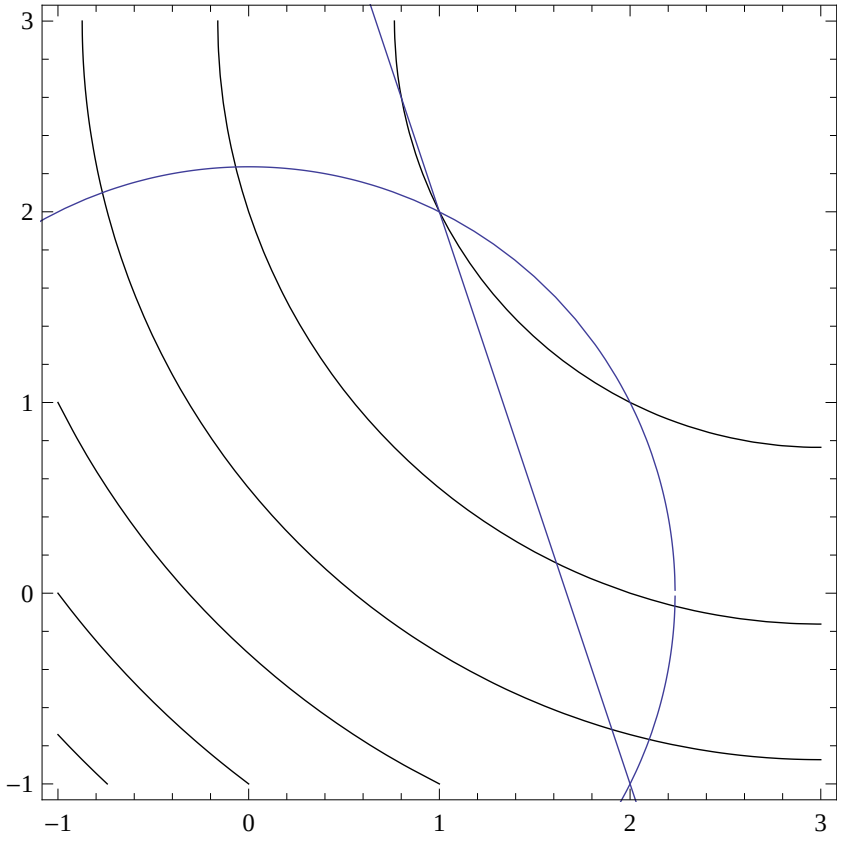

Figura 4.2: Región factible Ω y las curvas de nivel de F.

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

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

con $\Omega \neq \emptyset$ definido por

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

con h_j una función no lineal.

Teorema 4.3.1. *Si $\vec{x}^*$ es un punto extremo 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*

$$\nabla F(\vec{x}^*) + \sum_{j=1}^{s} \mu_j \nabla h_j(\vec{x}^*) = 0 \tag{4.11}$$

y

$$\mu_j[h_j(\vec{x}^*)] = 0 \quad \forall\ j = 1, \ldots, s. \tag{4.12}$$

La demostración de este teorema es la siguiente: Sea

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

entonces si $\vec{x}^*$ es el mínimo de F en Ω entonces también lo es en S; dado que las restricciones que definen s S son restricciones de igualdad por lo que al tener puras restricciones de igualdad, en $\vec{x}^*$ se deben satisfacer las condiciones de primer orden para restricciones de igualdad vistas en la sección anterior, por lo tanto existen μ_j para $j \in I(\vec{x}^*)$ tal que

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

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

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

además se cumple que

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

Basta ahora 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 hay alguna $\mu_k < 0$ 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 $\vec{x}(0) = \vec{x}^*$ y que $\vec{x}'(0) = \vec{y}$ y que para alguna $\delta > 0$, $\vec{x}(t) \in S_k$ para $t \in < 0, \delta >$. Entonces

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

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

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

Teorema 4.3.2. *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 $N(\vec{x}^*) = \{\vec{y} \in \Re^m | {}^t\nabla\vec{h}(\vec{x}^*)\vec{y} = 0 \ \forall j \in I(\vec{x}^*)\}$ 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 4.3.3. *(Condiciones suficientes) Supóngase que hay un punto $\vec{x}^*$ en Ω y una $\vec{\mu} \in \Re^s$ tal que*

$$\nabla F(\vec{x}^*) + \sum_{j=1}^{s} \mu_j \nabla h_j(\vec{x}^*) = 0. \tag{4.13}$$

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 local estricto de F en Ω.*

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 el primer caso 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

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

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

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

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 $\mu_2 = 18/5$ por lo tanto no satisface las condiciones de Kuhn y Tucker. Entonces $(1,2)$ es candidato a ser el mínimo. 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 de Kuhn y Tucker (KT) correspondientes son: existen λ, $\mu \in \Re$ y $\nu_i \geq 0$, con $i = 1, \ldots, n$ tal que

$$\Sigma w + \lambda \overrightarrow{r} + \mu \overrightarrow{1} - \sum_{i=1}^{n} \nu_i \overrightarrow{e}_i = 0, \tag{4.14}$$

$$w^t \overrightarrow{r} = r^*, \tag{4.15}$$

$$\overrightarrow{1}^t w = 1, \tag{4.16}$$

$$\nu_i w_i = 0, \quad i = 1, \ldots, n \tag{4.17}$$

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

$$\Sigma w + \lambda \overrightarrow{r} + \mu \overrightarrow{1} = 0, \tag{4.18}$$

$$w^t \overrightarrow{r} = r^*, \tag{4.19}$$

$$\overrightarrow{1}^t w = 1. \tag{4.20}$$

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

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

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

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

En este caso el conjunto admisible es

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

En la Figura 4.3 se presenta la solución gráfica de este problema. Observemos que el punto de Ω que esta en la curva de nivel de menor valor es el que solo se cumple que $h_1 = 0$.

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

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

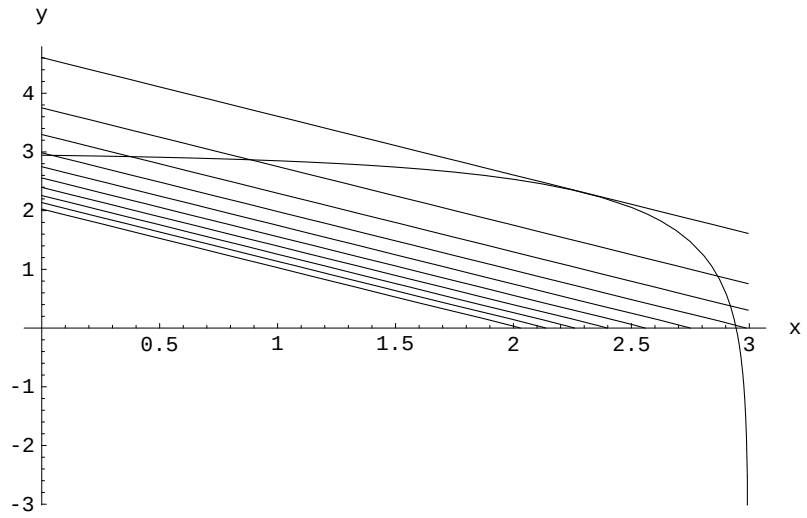

Figura 4.3: Región factible Ω con las curvas de nivel de F.

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.

4.4. 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 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 periødico. 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 % la venta de anuncios. 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 $\$50,000$ nuevos pesos a la semana?

2. Resuelva analíticamente el siguiente problema

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

Bosqueje el conjunto Ω admisible.

3. Resuelva analíticamente

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

   Grafique la región admisible.

4. 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.

5. 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.

6. 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?

7. 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 $\sum_{i=1}^{N} x_i^2 = 1$.

8. 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.

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^*, sin ventas en corto y la suma de las W_i debe ser igual a 1.

10. Determine transformando el siguiente problema a un problema de programación lineal: 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 $x, y \geq 0$.

11. Determine la solución del problema anterior en forma directa.

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

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

15. Determine la proyección al rango de A: $P_{R(A)}$.

16. ¿Cómo determinar un punto admisible en el caso de restricciones lineales de igualdad? Usando la factorización QR para obtener un punto $\mathbf{x} \in \Re^n$ tal que $C^t\vec{x} = \vec{e}$. Primero se resuelve

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

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

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

Aplicar este procedimiento para resolver los siguientes sistemas

$$A\vec{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

y

$$B^t\vec{x} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

con A y B del ejercicio 12.

Capítulo 5

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. En la primera sección se aplicarán para el caso de restricciones de igualdad. 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.

5.1. 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:

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

5.1.1. 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

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

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 $A = QR$.

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 está 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

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

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 d_k en $EN(A^t)$ para que la sucesión generada por el método de descenso permanezca en Ω. A estas direcciones se le conocen 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

$$\nabla F(\vec{x}^*) + A\vec{\lambda} = 0, \tag{5.1}$$
$$A^t\vec{x}^* = \vec{e}, \tag{5.2}$$

que satisface

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

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

Con objeto de desacoplar las ecuaciones (5.1), usaremos el Lema 4.2.2. 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

$$\nabla F(\vec{x}^*)^t\ y = 0 \quad \forall\ \vec{y} \in EN(A^t) \tag{5.3}$$

y posteriormente

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

5.1.2. Método de Newton

Para determinar numéricamente un punto $\vec{x}^* \in \Omega$ que satisfaga la ecuación (5.2) y (5.3) usaremos el método de descenso; en cada iteración $k+1$ se genera un punto $\vec{x}_{k+1}$ que satisfaga (5.3) para la aproximación cuadrática que

se obtiene al aproximar $\nabla F(\vec{x}_{k+1})$ por medio de 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

$$0 = \vec{y}^t \ \nabla F(\vec{x}_{k+1}) \approx \vec{y}^t \ \nabla F(\vec{x}_k) + \vec{y}^t \ H_F(\vec{x}_k) \ (\vec{x}_{k+1} - \vec{x}_k) \tag{5.4}$$

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

Para garantizar que (5.4) 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

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

$\Rightarrow$

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

5.1.3. Algoritmo de Newton

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

$$Z^t H_F(\vec{x}_k) Z \ \vec{b_k} = -Z^t \nabla F(\vec{x}_k), \tag{5.5}$$

$$\vec{d_k} = Z\vec{b_k}. \tag{5.6}$$

   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 $\frac{\|\vec{x}_{k+1} - \vec{x}_k\|}{\|\vec{x}_k\|} \leq rtol \Rightarrow$

$$A^t A \lambda_{k+1} = -A^t \nabla F(\vec{x}_{k+1}). \tag{5.7}$$

   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.

Observaciones:

1. Para comprobar en cada iteración que las condiciones de segundo orden se cumplen, el sistema (5.5) 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 $\vec{x}_{k+1}$
$$A^t A\lambda_{k+1} = -A^t[\nabla F(\vec{x}_{k+1}) + H_F(\vec{x}_{k+1})(\vec{x}_{k+1} - \vec{x}_k)]. \qquad (5.8)$$

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 $\frac{\|\vec{x}_{k+1} - \vec{x}_k\|}{\|\vec{x}_k\|} \leq rtol \Rightarrow$
$$R\lambda_{k+1} = Q_1^t \nabla F(\vec{x}_{k+1}). \qquad (5.9)$$
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

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

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

$$\vec{d_0} = Z b_0 = \begin{pmatrix} 0.904761905 \\ -0.226190476 \\ 1.130952381 \end{pmatrix}.$$

Entonces

$$\vec{x^*} = \vec{x}_0 + \vec{d_0} = (0.904761905, 0.523809524, -0.619047619).$$

Por otro lado,

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

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

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

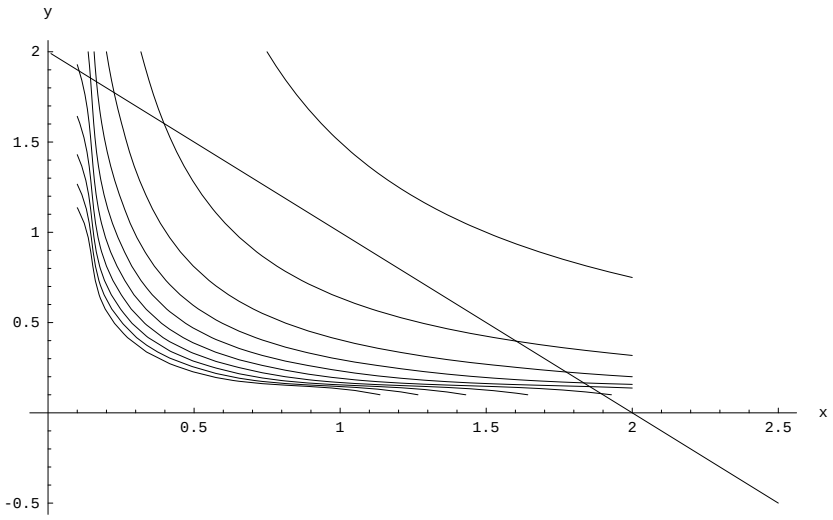

Figura 5.1: Región factible Ω con las curvas de nivel de F.

En la Figura 5.1 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 curvas de nivel que cortan al segmento de recta corresponden a valores mayores.

5.2. 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 pues en el la 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 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 $\vec{g}_{k+1} = \vec{g}_k$.

Si $\beta_t \neq 0$, pero

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

$$\begin{aligned} Q^{2\,t}_{q,k} H_F(\vec{x}_k) Q^2_{q,k}\ \vec{b}_k &= -Q^{2\,t}_{q,k} \nabla F(\vec{x}_k), \\ \vec{d_k} &= Q^2_{q,k} \vec{b}_k. \end{aligned}$$

3. Para seleccionar α_k se hace lo siguiente: sea

$$\beta_t = \min_{i \notin I(x_k)} \{\beta_i = -\frac{h_j(\vec{x}_k)}{\vec{A}^t_j d_k}\},$$

   $\Rightarrow \alpha_k = \beta_t$ 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 $\vec{g}_{k+1} = \nabla F(\vec{x}_{k+1})$. Si $F_{k+1} < F_k \Rightarrow$ ir al Paso 7 para incluir en las restricciones activas la t-ésima restricción. Si $F_{k+1} \geq F_k$ entonces determinar alguna ....

4. Si $\|Q^{2,t}_{k+1} \nabla F(\vec{x}_{k+1})\| < rtol$ y $\frac{\|\vec{x}_{k+1} - \vec{x}_k\|}{\|\vec{x}_k\|} \leq rtol \Rightarrow$

$$R_{k+1} \vec{\lambda}_{k+1} = Q^{1,t}_{k+1} \nabla F(\vec{x}_{k+1}) \tag{5.10}$$

   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

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

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

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

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. Consideremos un problema con función objetivo no lineal, pero con restricciones de desigualdad lineales.

5.3. Método de Wolfe

5.4. Ejercicios

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.

[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.