

UNIVERSIDAD CENTROCCIDENTAL
“LISANDRO ALVARADO”

Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas



“ESTUDIO DE UN MÉTODO DE REGIÓN DE CONFIANZA
CON ÁNGULO PARA OPTIMIZACIÓN SIN EL USO DE
DERIVADA. ”

TRABAJO ESPECIAL DE GRADO PRESENTADO POR

BR MARY I. HERNÁNDEZ C.

COMO REQUISITO FINAL
PARA OBTENER EL TÍTULO DE LICENCIADO
EN CIENCIAS MATEMÁTICAS
ÁREA DE CONOCIMIENTO: MATEMÁTICA APLICADA.
TUTOR: M.Sc. ALÍ DUIN.

Barquisimeto, Venezuela. Febrero de 2013



Universidad Centroccidental
 "Lisandro Alvarado"
 Decanato de Ciencias y Tecnología
 Licenciatura en Ciencias Matemáticas



ACTA
 TRABAJO ESPECIAL DE GRADO

Los suscritos miembros del Jurado designado por el Jefe del Departamento de Matemáticas del Decanato de Ciencias y Tecnología de la Universidad Centroccidental "Lisandro Alvarado", para examinar y dictar el veredicto sobre el Trabajo Especial de Grado titulado:

“ESTUDIO DE UN MÉTODO DE REGIÓN DE CONFIANZA CON
 ÁNGULO PARA OPTIMIZACIÓN SIN EL USO DE DERIVADA.”

presentado por el ciudadano BR MARY I. HERNÁNDEZ C. titular de la Cédula de Identidad No. 17.944.833, con el propósito de cumplir con el requisito académico final para el otorgamiento del título de Licenciado en Ciencias Matemáticas.

Luego de realizada la Defensa y en los términos que imponen los Lineamientos para el Trabajo Especial de Grado de la Licenciatura en Ciencias Matemáticas, se procedió a discutirlo con el interesado habiéndose emitido el veredicto que a continuación se expresa:

¹ _____

Con una calificación de _____ puntos.

En fe de lo expuesto firmamos la presente Acta en la Ciudad de Barquisimeto a los ____ días del mes de _____ de _____.

 TUTOR

 FIRMA

 PRINCIPAL

 FIRMA

 PRINCIPAL

 FIRMA

OBSERVACIONES:

¹ Aprobado ó Reprobado

A mis padres Sheila y Florencio.

A mis hermanos y mis sobrinos.

A mi abuela.

AGRADECIMIENTOS

A Dios, por permitirme nacer en una familia maravillosa, por acompañarme cada día en los momentos buenos y malos, por brindarme la alegría de conocer en este trayecto de mi vida a excelentes personas.

A mis padres, hermanos, primos y demas familiares, gracias por su apoyo, amor y comprensión.

A mis amigas Dayana, Maria José y Ranghely por su apoyado incondicional, gracias por estar siempre ahí.

A la profesora M.Sc. Alí Duín por la valiosa colaboración y motivación que me ha brindado.

A la U.C.L.A, al Decanato de Ciencias y Tecnología, a sus profesores por la formación académica que me ofrecieron.

Resumen

RESUMEN DEL TRABAJO DE GRADO PRESENTADO PARA OPTAR AL
TÍTULO DE LICENCIADO EN CIENCIAS MATEMÁTICAS

ESTUDIO DE UN MÉTODO DE REGIÓN DE CONFIANZA CON ÁNGULO
PARA OPTIMIZACIÓN SIN EL USO DE DERIVADA.

por

Mary Ines Hernández C.

Universidad Centroccidental Lisandro Alvarado

U.C.L.A.

Barquisimeto, Febrero 2013

M.Sc. Alí Duin

Tutor

Se plantea en este trabajo el estudio de un método de región de confianza propuesto por [Marazzi y Nocedal]. Este método se basa en las estrategias del modelo de región de confianza sin hacer uso de las derivadas aún, cuando se supone que la función a minimizar es suave, generando un modelo que interpola la función objetivo con el conjunto de puntos de muestra y usando región de confianza para promover convergencia. Los puntos de muestra se actualizan usando un esquema que mejora la exactitud del modelo de interpolación cuando sea necesario. En el subproblema del modelo aproximado [Marazzi y Nocedal] imponen una restricción de ángulo con el propósito de asegurar que dicho subproblema sea resoluble y estable desde el punto

de vista numérico. Ellos proponen un modelo lineal y un modelo cuadrático y además afirman haber hecho pruebas para métodos establecidos de optimización libre de derivadas.

Palabras claves: Optimización sin derivada, Región de Confianza.

ÍNDICE

Agradecimientos	i
Índice de figuras	v
1. INTRODUCCIÓN	1
2. PRELIMINARES	3
2.1. Optimización Libre de Derivadas	8
3. ALGORITMO	10
4. MODELO LINEAL	15
5. MODELO CUADRÁTICO	19
6. RESULTADOS	26
Referencias	31

Índice de figuras

2.1. Radios de región de Confianza	5
2.2. Punto de Cauchy	8
4.1. Figura 1	16
4.2. Figura 2	17
5.1. Figura 3	23

CAPÍTULO 1

INTRODUCCIÓN

En este trabajo se estudia el método de Región de Confianza para Optimización Libre de Derivadas propuesto por [Marazzi y Nocedal]. Esto es, resolver el problema de minimizar una función suave f de varias variables cuyas derivadas no están disponibles. Formalmente

$$\min_{x \in \mathbb{R}^n} f(x)$$

Para resolver este tipo de problemas han sido propuestos varios métodos cuando las derivadas no están disponibles, entre ellos los métodos de región de confianza sobre la base de modelos de interpolación. En este trabajo el enfoque pertenece a la última clase: se forman un modelo lineal o cuadrático de la función objetivo y se hace uso de región de confianza para promover convergencia. A diferencia de los métodos de región de confianza que hacen uso de derivadas, éste método incluye una restricción en el subproblema de región de confianza que asegura que la posición de todos los puntos generados por el algoritmo es tal que ellos definen un modelo lineal o un modelo cuadrático adecuado. Además se crea un modelo que aproxima a la función alrededor del iterado actual, éste modelo es requerido para interpolar la función f en el iterado actual y en un conjunto de puntos de muestra.

El propósito de éste trabajo está en analizar detenidamente la estrategia expuesta por [Marazzi y Nocedal] detallando algunos aspectos y enmarcando el trabajo en el contexto general, tanto de optimización sin derivadas como el región de confianza.

No podemos dejar de mencionar algunos antecedentes importantes como lo es el trabajo realizado por [D. Winfield 1973], el cuál se considera pionero al utilizar el método de región de confianza para resolver problemas de optimización sin restricciones y sin el uso de derivada, quien propuso usar valores de la función objetivo disponibles

para construir un modelo cuadrático por interpolación y trabajó bajo la hipótesis de que el modelo es válido en un entorno del iterado actual, el entorno es descrito como una región de confianza cuyo radio es ajustado iterativamente. Esta idea fue retomada por [M. J. D. Powell 1994 a] quien propuso un método para optimización con restricciones en el cual la función objetivo y las restricciones son aproximadas por la interpolación multivariada lineal, nuevamente [M. J. D. Powell 1994 b] retoma esta idea pero ahora describe un algoritmo para optimización sin restricciones usando un modelo de interpolación cuadrática multivariable de la función objetivo en un marco de región de confianza, una aproximación muy similar a la de [D. Winfield 1973]. Por otra parte, tenemos a [A. R. Conn and Ph. L. Toint 1996] quien describe un esquema de interpolación cuadrática usando polinomio fundamental de Newton.

CAPÍTULO 2

REGIÓN DE CONFIANZA

En este capítulo se dará una descripción general de los métodos de región de confianza para minimizar funciones sin la presencia de restricciones y se darán algunos comentarios sobre la optimización de una función sin la presencia de restricciones y sin la ayuda de las derivadas de la función.

Esta exposición esta basada en una descripción informal de los métodos de región de confianza, esto es, en cada iteración se aproxima la función f que se pretende minimizar por una aproximación bien sea lineal o cuadrática limitando la búsqueda del mínimo de ésta función aproximada a una bola de tamaño “conveniente”.

Si un paso no es aceptable, se reduce el tamaño de la región y se encuentra un nuevo minimizador. En general, la dirección de paso cambia siempre que el tamaño de la región de confianza es modificado.

En la optimización sin restricciones que no usan región de confianza tradicionalmente se sigue una estrategia como ésta: se determina una dirección de búsqueda, la cual permite mejorar el iterado actual pero no directamente sino moviendose a lo largo de ella, entonces entra en juego la búsqueda lineal, la cual efectúa una minimización en una dimensión buscando a lo largo de la dirección previamente determinada. A diferencia de los métodos de región de confianza.

En los algoritmos prácticos, escogemos el tamaño de la región de acuerdo con el rendimiento del algoritmo durante las iteraciones previas. Si el modelo es en general seguro, produciendo buenos pasos y pronosticando el comportamiento de la función objetivo con exactitud a lo largo de estos pasos, el tamaño de la región de confianza es incrementado regularmente para permitir que pasos más largos y ambiciosos sean tomados.

Por otro lado, un paso fallido indica que nuestro modelo es una representación inadecuada de la función objetivo sobre la región de confianza en curso, así que reducimos el tamaño de la región otra vez.

Subproblema.

Para obtener cada paso en el método de Regiones de Confianza se pide una solución al subproblema:

$$\min_{s \in \mathbb{R}^n} m_k(s) = f_k + \nabla f_k^\top s + \frac{1}{2} s^\top B_k s \quad \text{Sujeto a: } \|s\| \leq \Delta_k \quad (2.1)$$

Donde $\Delta_k > 0$ es el radio de región de confianza.

Como se considera que no se tiene información de derivadas, el modelo m_c que se usa para aproximar a la función impone pocas restricciones sobre la matriz B_k como por ejemplo que B_k sea definida positiva y uniformemente acotada.

Los métodos de región de confianza proponen la solución de una secuencia de problemas del tipo (2.1), la solución a estos subproblemas no es tan obvia, pero generalmente se encuentra sin demasiado costo computacional. Necesitamos solamente una solución aproximada para obtener la convergencia y buen comportamiento.

Enfoque de la Región de Confianza

Uno de los componentes clave en el algoritmo de región de confianza, es la estrategia para elegir el radio de región de confianza Δ_k en cada iteración. Se basa estratégicamente en escoger el valor de un parámetro ρ_k , el cual se calcula en cada iteración luego de resolver el subproblema (2.1). De un modo aproximado, el parámetro ρ_k trata de “medir” la relación entre el progreso que se alcanza en m_k al resolverlo y el progreso que se alcanza en f con el mismo movimiento.

Dado el paso s_k definimos el radio

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{m_k(0) - m_k(s_k)} \quad (2.2)$$

El numerador es llamado Reducción Actual (**ared**) mientras que el denominador

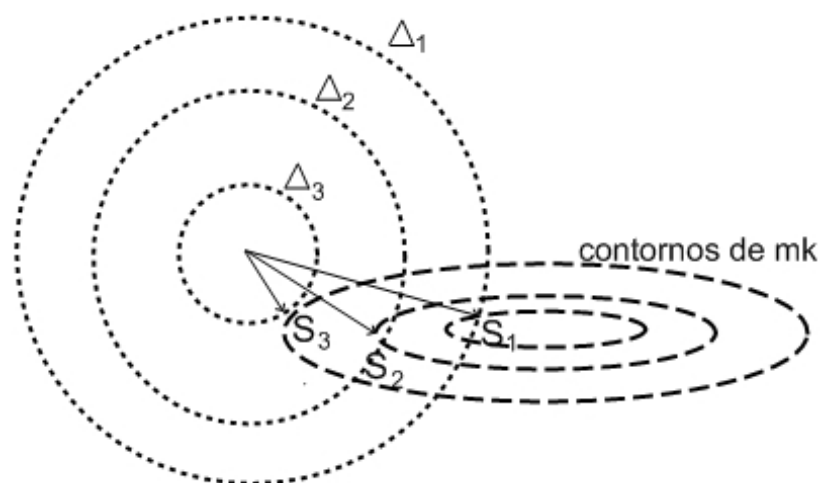


Figura 2.1: Solución del subproblema de Región de Confianza para Diferentes Radios.

es llamado Reducción Pronosticada (**pred**). Ya que el paso s_k es obtenido por minimización del modelo m_k sobre la región que incluye $s = 0$, la reducción pronosticada debería ser siempre no negativa.

Ahora bien, si ρ_k es negativo, el nuevo valor objetivo $f(x_k + s_k)$ es mayor que el valor actual $f(x_k)$, por lo que el paso debe ser rechazado.

Si ρ_k es cercano a 1, hay buen acuerdo entre el modelo m_k y la función f sobre este paso, así que es seguro ampliar la región de confianza para la próxima iteración.

Si ρ_k es positivo pero no cercano a 1, no modificamos la región.

Si ρ_k es cercano a 0 o negativo, escogemos una nueva región de confianza al reducir Δ_k en la siguiente iteración.

Algoritmo (Región de Confianza)

Dado $\hat{\Delta} > 0$, $\Delta_0 \in (0, \hat{\Delta})$, y $\eta \in [0, \frac{1}{4})$:

```

for  $k = 0, 1, 2, \dots$ 
  Obteniendo  $s_k$  (aproximadamente) resolviendo (2.1);
  Evaluar  $s_k$  en (2.2)
  if  $\rho_k < \frac{1}{4}$ 
     $\Delta_{k+1} = \frac{1}{4}\Delta_k$ 
  else
    if  $\rho_k > \frac{3}{4}$  y  $\|s_k\| = \Delta_k$ 
       $\Delta_{k+1} = \min(2\Delta_k, \hat{\Delta})$ 
    else
       $\Delta_{k+1} = \Delta_k$ ;
  if  $\rho_k > \eta$ 
     $x_{k+1} = x_k + s_k$ ;
  else
     $x_{k-1} = x_k$ ;
end(for)

```

Donde $\hat{\Delta}$ es una cota sobre toda longitud de paso. Notemos que el radio es incrementado solo si $\|s_k\|$ en realidad alcanza el límite de la región de confianza.

Punto de Cauchy

El punto de Cauchy no es más que el reductor de m_k a lo largo de la dirección de máximo descenso $-g_k$, sujeto a la región de confianza. Los métodos de búsqueda lineal pueden ser convergentes, incluso cuando la longitud del paso óptimo no se utiliza en cada iteración. La longitud del paso α_k sólo necesita satisfacer criterios. Aunque en un principio pedimos la solución óptima del subproblema (2.1), es suficiente para los propósitos de la convergencia global encontrar una solución aproximada p_k que está dentro de la región de confianza y da una suficiente reducción en el modelo. La suficiente reducción puede ser cuantificada en relación con el punto de Cauchy, que denotaremos por p_k^c y se define por el siguiente procedimiento:

Algoritmo. (Punto de Cauchy)

Encuentre el vector p_k^s tal que resuelva una versión lineal de (2.1), es decir,

$$p_k^s = \arg \min_{p \in \mathbb{R}^n} f_k + g_k^\top s \quad \text{Sujeto a: } \|s\| \leq \Delta_k \quad (2.3)$$

Calculamos el escalar

$$\tau_k = \arg \min_{\tau > 0} m_k(\tau p_k^s) \quad \text{Sujeto a: } \|\tau p_k^s\| \leq \Delta_k; \quad (2.4)$$

El conjunto $p_k^c = \tau_k p_k^s$.

Es fácil deducir que la solución de (2.3) es simplemente

$$p_k^s = -\frac{\Delta_k}{\|g_k\|} g_k$$

Para obtener τ_k explícitamente, consideramos los casos $g_k^\top B_k g_k \leq 0$ y $g_k^\top B_k g_k > 0$ por separado.

En el primer caso, la función $m_k(\tau p_k^s)$ decrece monótonicamente con τ cada vez que $g_k \neq 0$, así que τ_k es sólo el valor más grande que satisface la región de confianza acotada, esto es que, $\tau_k = 1$.

Para el segundo caso, $g_k^\top B_k g_k > 0$, $m_k(\tau p_k^s)$ es convexo cuadrático en τ , así que τ_k es cualquier minimizador irrestricto de esta función cuadrática, $\|g_k\|^3 / (\Delta_k g_k^\top B_k g_k)$, o el valor límite 1, lo que ocurra primero. En efecto, derivando $m_k(\tau p_k^s)$ con respecto a τ e igualando 0, tenemos

$$p_k^s g_k + \tau (p_k^s)^\top B_k p_k^s = 0 \implies \tau = \frac{-p_k^s g_k}{(p_k^s)^\top B_k p_k^s}$$

Sustituyendo $(p_k^s) = \frac{-\Delta_k g_k}{\|g_k\|}$

$$\begin{aligned} \tau &= \frac{\frac{\Delta_k g_k}{\|g_k\|} \cdot g_k}{\frac{-\Delta_k g_k^\top B_k g_k}{\|g_k\|} - \frac{-\Delta_k g_k}{\|g_k\|}} \\ &= \frac{\frac{\Delta_k}{\|g_k\|} \|g_k\|^2}{\frac{\Delta_k^2 g_k^\top B_k g_k}{\|g_k\|^2} - \frac{\|g_k\|^3}{\Delta_k g_k^\top B_k g_k}} \end{aligned}$$

En resumen, tenemos

$$p_k^c = -\tau_k \frac{\Delta_k}{\|g_k\|} g_k \quad (2.5)$$

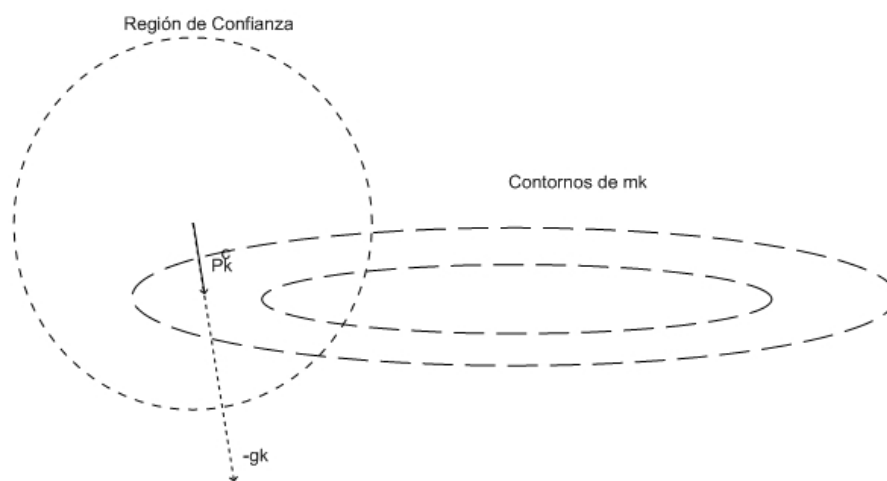


Figura 2.2: Punto de Cauchy.

Específicamente, un método de región de confianza será globalmente convergente si sus pasos dan una reducción s_k en el modelo m_k que es por lo menos algún múltiplo positivo fijo del decrecimiento conseguido por el paso de Cauchy.

Donde

$$\tau_k = \begin{cases} 1 & \text{si } g_k^\top B_k g_k \leq 0 \\ \text{mín}(\|g_k\|^3 / (\Delta_k g_k^\top B_k g_k), 1) & \text{en otro caso} \end{cases} \quad (2.6)$$

2.1. Optimización Libre de Derivadas

Los métodos de optimización libre de derivadas (DFO) son fundamentales en la optimización para cierto tipo de problemas. En particular, se ocupan de problemas donde la información de la derivada es de costo computacional no alcanzable o difícil de calcular. Este tipo de problemas se producen a menudo en las industrias.

Los algoritmos (DFO) difieren de otros algoritmos en la forma en que usan los valores de la función de trabajo para determinar el nuevo iterado. Una clase método construye un modelo lineal o cuadrático de la función objetivo y define la siguiente iteración intentando minimizar este modelo dentro de la región de confianza. Estos métodos no están tan bien desarrollados como los métodos basados en gradiente, los

algoritmos solo son efectivos para problemas pequeños.

CAPÍTULO 3

ALGORITMO

Se describe en esta sección la estructura general para el método de región de confianza con ángulo descrito por [Marazzi y Nocedal]. Para el iterado actual x_c se define el modelo

$$m_c(x_c + s) = f(x_c) + g_c^\top s + \frac{1}{2} s^\top G_c s \quad (3.1)$$

donde el vector $g_c \in \mathfrak{R}^n$ y G_c la matriz simétrica de orden $n \times n$ que deberá determinarse mediante la interpolación de f con los puntos de muestra. (Es claro que para el modelo lineal se define $G_c \equiv 0$)

El modelo (3.1) es minimizado con respecto a $s \in \mathfrak{R}^n$ sujeto a una restricción de la forma $\|s\| \leq \Delta_c$ para generar un paso $s_c \in \mathfrak{R}^n$ que conduce al punto de ensayo $x_+ = x_c + s_c$. Si x_+ reduce la función objetivo, éste es aceptado como el nuevo iterado y el radio de región de confianza Δ puede ser incrementado; en otro caso Δ es reducido y un nuevo paso de ensayo es calculado.

Para definir el modelo m_c de tal forma que quede determinado de manera única, se requiere del punto actual x_c y de un conjunto de m puntos los cuales se denominan satélites de x_c , y se denotan como

$$\Sigma_c = \{y^1, \dots, y^m\}$$

Es fácil ver que para un modelo lineal se debe tomar $m = n$ y para un modelo cuadrático m_c éste debe elegirse como

$$m = \frac{(n+1)(n+2)}{2} - 1$$

Se impone entonces las condiciones de interpolación

$$m_c(x_c) = f(x_c), \quad m_c(y^l) = f(y^l), \quad l = 1, \dots, m$$

Cuando el modelo m_c es determinado únicamente por éstas condiciones, se dice que el conjunto de interpolación $\{x_c\} \cup \Sigma_c$ es no degenerado.

Suponga que comenzamos la iteración actual con un conjunto de muestra no degenerado $\{x_c\} \cup \Sigma_c$. Entonces al calcular un nuevo punto de ensayo usando el modelo m_c , el satélite más lejano del iterado actual x_c , llámese $y^{l_{out}}$, es identificado como el punto que será removido de Σ_c . Esta elección promueve la conservación de puntos que suministran información local de f alrededor de x_c .

Como se pretende diseñar un método numérico, se debe tener cuidado con los posibles errores en operaciones sobre valores de punto flotante, evitando resolver sistemas de ecuaciones que están “Cerca ”de ser degenerado, por lo cual se requiere identificar aquellos puntos en \mathfrak{R}^n que esten “Cerca”de la región Tabú. A éste conjunto se le denomina conjunto de **Ángulo**

Se define ahora una “Región Tabú”como la región que contiene aquellos puntos de la forma $x_c + s$ tales que si alguno de ellos (cualquiera que se tome) y se colocase en lugar de $y^{l_{out}}$ en el conjunto Σ_c daría lugar a un conjunto de puntos degenerado. Además se define un “ángulo”, el cuál es un conjunto W_c que contiene τ_c y que es designado para evitar que los puntos estén muy cercanos a τ_c . La descripción de τ_c y W_c para el caso del modelo lineal y cuadrático serán dadas más adelante, donde se muestra que las representaciones apropiadas de estos conjuntos son fáciles de calcular.

Una vez que el ángulo W_c ha sido determinado, se calcula el paso de ensayo s_c por aproximación resolviendo

$$\min_s m_c(x_c + s) = f(x_c) + g_c^\top s + \frac{1}{2} s^\top G_c s \quad (3.2)$$

$$\text{sujeto a } \|s\| \leq \Delta_c \quad (3.3)$$

$$s \notin W_c \quad (3.4)$$

y se define $x_+ = x_c + s_c$. Si éste punto de ensayo x_+ reduce a f , entonces x_+ se convierte en el nuevo iterado y x_c pasa a ser un punto satélite, reemplazando y^{lout} . Si por el contrario x_+ no reduce f , el iterado actual no es actualizado y x_+ puede o no ser descartado, dependiendo de lo lejos que esté del iterado actual x_c comparado con y^{lout} . Este y otros aspectos del algoritmo son descritos a continuación.

Algoritmo 1

Se eligen los parámetros de región de confianza $\alpha, \beta \in (0, 1)$, un radio de región de confianza inicial $\Delta_c > 0$, y un valor inicial x_c . Se selecciona un conjunto de puntos satélites $\Sigma_c = \{y^1, y^2, \dots, y^m\}$ tales que $x_c \cup \Sigma_c$ es no degenerado. Acá $m = n$ para el modelo lineal y $m = \frac{1}{2}(n+1)(n+2) - 1$ para el modelo cuadrático. (Se supone que $f(x_c) \leq f(y) \forall y \in \Sigma_c$).

Repetir

1. Encuentre el satélite más lejano del iterado actual (romper los empates arbitrariamente):

$$y^{lout} = \arg \max_{y \in \Sigma_c} \|y - x_c\|$$

2. Forme un modelo $m_c(x_c + s)$ mediante la interpolación de $\{x_c\} \cup \Sigma_c$, y defina la restricción de ángulo W_c
3. Calcule s_c por aproximación resolviendo el subproblema (3.2) y evaluando $f(x_c + s_c)$.
4. Establezca $ared(s_c) \equiv f(x_c) - f(x_c + s_c)$ y $pred(s_c) \equiv m_c - m_c(x_c + s_c)$.

5. Actualice Δ_c .

If $ared(s_c) > \alpha pred(s_c)$, elija Δ_+ tal que $\Delta_+ \geq \Delta_c$,

Else elija $\Delta_+ = \beta \Delta_c$.

6. **If** $f(x_c + s_c) < f(x_c)$ (iteración exitosa)

Actualice el iterado actual, incluir x_c en el conjunto de satélites, descartando y^{lout} :

a) $x_+ = x_c + s_c$.

b) $\Sigma_+ = \{x_c\} \cup \Sigma_c \setminus \{y^{lout}\}$.

Else (iteración sin éxito)

El nuevo punto de ensayo x_+ no es más lejano de x_c que y^{lout} , entonces x_+ es admitido en el conjunto de satélites Σ_c descartando y^{lout} ; en otro caso descartar el nuevo punto de ensayo:

c) $x_+ = x_c$

d) **If** $\|y^{lout} - x_c\| \geq \|(x_c + s_c) - x_c\|$, $\Sigma_+ = \{x_c + s_c\} \cup \Sigma_c \setminus \{y^{lout}\}$

Else $\Sigma_c = \Sigma_+$

7. $x_c = x_+$, $\Sigma_c = \Sigma_+$, $\Delta_c = \Delta_+$

End

Más adelante se hablará de la definición del modelo m_c , el ángulo W_c y el procedimiento para resolver aproximadamente el subproblema de región de confianza.

El Algoritmo 1 es conceptualmente simple ya que se genera sólo un tipo de paso, es decir un paso de minimización que siempre intenta disminuir a f . Su novedad radica en el uso de la restricción de ángulo y en la aceptación de la estrategia en el paso 6. Esta estrategia asegura que el modelo es suficientemente preciso cuando es necesario, y por lo tanto contribuye significativamente a la robustez de la iteración, como se discutirá ahora.

Los métodos de región de confianza para optimización basado en gradiente garantizan que un paso exitoso se genera cuando la región de confianza es lo suficientemente pequeña (y asumiendo que x_c es un punto no estacionario de f). Con el fin de mantener esta propiedad importante en modelos basados en interpolación, desviándose de la práctica habitual de descartar puntos de ensayo que dan lugar a un aumento en la función objetivo, si estos puntos ayudan a mejorar la precisión del modelo en una vecindad de x_c . Más específicamente, supongase que el modelo m_c es pobre y que a consecuencia, una secuencia de puntos de ensayo sin éxito es calculada. Si estos puntos de ensayo fueran descartados, entonces el modelo de interpolación no cambiaría y los pasos subsiguientes todavía pueden ser pobres a pesar del hecho de que la región de confianza se ha reducido.

Para asegurar que la calidad del modelo mejora a medida que los pasos están siendo rechazados, se propone el mecanismo descrito en el paso 6 del algoritmo. Ya que un punto de ensayo no exitoso $x_c + s_c$ será retenido como un satélite si y solo si no es más lejano de x_c que y^{lout} (paso 6d), se promueve la conservación del punto de ensayo en la vecindad de x_c , y además evita la pérdida debido a evaluaciones costosas de la función en un punto de ensayo no exitoso $x_c + s_c$ siempre que sea posible.

Si una secuencia de puntos de ensayo no exitosa es generado y como resultado Δ decrece suficientemente, el punto de ensayo eventualmente será admitido como satélite de x_c (paso 6d). Estos satélites serán entonces no degenerados (debido a la acción del ángulo) y cada vez más agrupados entorno a x_c (ya que Δ disminuye). Por lo tanto, el modelo de interpolación se convertirá en una aproximación cada vez más precisa de f en una vecindad de x_c , de modo que, finalmente un paso exitoso será calculado. Este es el principal mecanismo que conduce la teoría de convergencia. Como en las pruebas de convergencia de los métodos de región de confianza con gradientes exactos, también es necesaria una cota inferior para la reducción $pred(s_c)$ alcanzada en cada paso por el modelo. Hasta ahora hemos obtenido este límite sólo para el caso cuando m_c es lineal (referencia).

En el siguiente capítulo se desarrollará la construcción del modelo lineal.

CAPÍTULO 4

MODELO LINEAL

Recordemos que los métodos de región de confianza proponen el uso de un modelo aproximado de la función f que sea “sencillo” para así minimizarlo “facilmente”. Se piensa inicialmente en modelos lineales (los más sencillos) y en modelos cuadráticos (que se amoldan mejor a las funciones no lineales).

El modelo lineal puede ser útil en optimización libre de derivada porque sólo requiere $n+1$ puntos de muestra, una característica útil cuando el número de variables no es muy pequeño. El modelo, en la iteración actual x_c tiene la forma

$$m_c(x_c + s) = f(x_c) + g_c^\top s \quad (4.1)$$

donde g_c es un vector en \mathbb{R}^n que será determinado. Como g_c tiene n componentes, se mantiene en adición a x_c , el conjunto de n satélites

$$\Sigma_c = \{y^1, \dots, y^n\}$$

Se imponen las condiciones de interpolación $m_c(y^l) = f(y^l)$, $l = 1, \dots, n$, que se pueden escribir como

$$g_c^\top s^l = f(y^l) - f(x_c) \quad l = 1, \dots, n \quad (4.2)$$

Donde s^l es el desplazamiento desde x_c hasta y^l , esto es

$$y^l = x_c + s^l \quad l = 1, \dots, n$$

Se sigue de (4.2) que el modelo lineal (4.1) es únicamente determinado si y solo si el conjunto de puntos de muestra $\{x_c\} \cup \Sigma_c$ es tal que el conjunto

$$\{s^l : l = 1, \dots, n\}$$

es linealmente independiente.

Para calcular un nuevo iterado, primero se selecciona $y^{l_{out}}$, el satélite más lejano de x_c . La región tabú τ_c , que es la región que se quiere evitar cuando se calcula un nuevo punto de modo que el nuevo conjunto de muestra es no degenerado, es por lo tanto definido como el subespacio $(n - 1)$ dimensional generado por los vectores de desplazamiento

$$\{s^l : l = 1, \dots, n \quad l \neq l_{out}\} \quad (4.3)$$

correspondientes a los satélites que permanecerán en el conjunto de muestra, (Figura 1). Una representación más conveniente es

$$\tau_c = \{s \in \mathfrak{R}^n : b_c^\top s = 0\}$$

donde $b_c \in \mathfrak{R}^n$ es normal a los vectores de desplazamiento (4.3).

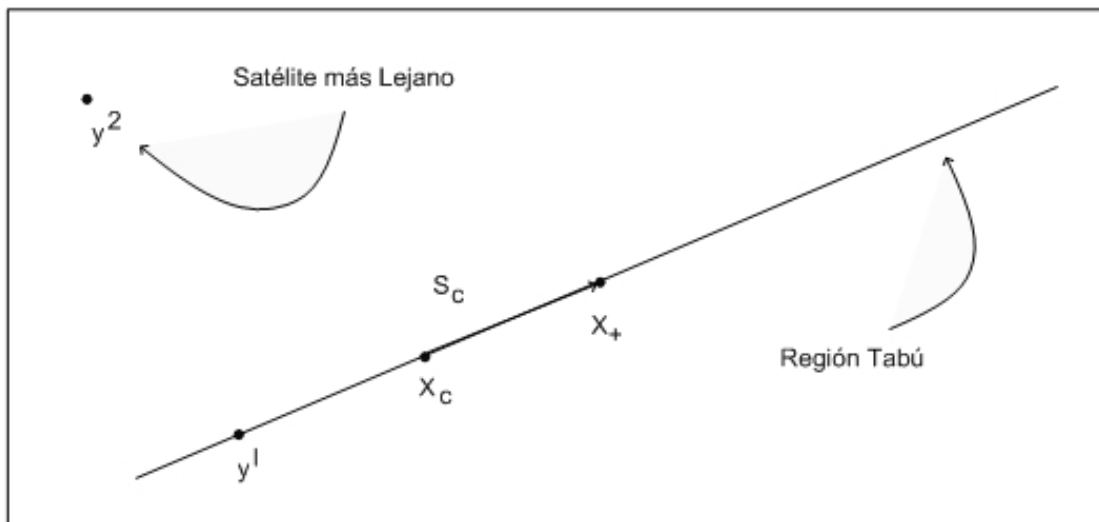


Figura 4.1: Figura 1

Como se ha mencionado, también se prefiere evitar los pasos s_c que estén muy cercanos a la región tabú, de modo que el sistema (4.2) no esté mal condicionado

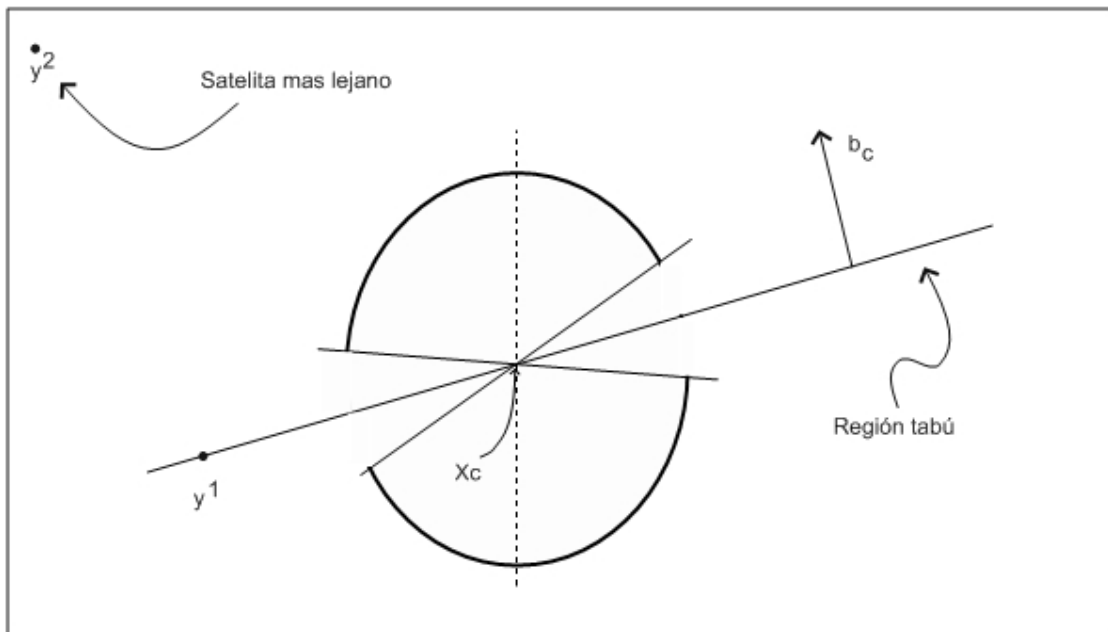


Figura 4.2: Figura 2

y para asegurar que los puntos de muestra están razonablemente espaciados. Para lograr esto, se pide que la magnitud del coseno del ángulo entre el paso s_c y la normal b_c no sea menor que una constante dada $\gamma \in (0, 1)$, esto es,

$$|b_c^\top s| \geq \gamma \|b_c\| \|s\| \quad (4.4)$$

Se llamará a la desigualdad (4.4) **la restricción de ángulo**, y el parámetro γ determina el “ancho” del ángulo.

Se calcula un paso de ensayo mediante la solución de

$$\min_s m_c(x_c + s) = f(x_c) + g_c^\top s \quad (4.5)$$

$$\text{sujeto a } \|s\| \leq \Delta_c \quad (4.6)$$

$$|b_c^\top s| \geq \gamma \|b_c\| \|s\| \quad (4.7)$$

Este problema puede ser solucionado fácilmente. Si se ignora la restricción de ángulo (3.2), la solución es

$$s_{tr} = - \left(\frac{\Delta_c}{\|g_c\|} \right) g_c.$$

Si s_{tr} satisface la restricción de ángulo (3.2) (esto es, si se encuentra fuera del ángulo), entonces $s_c = s_{tr}$ es la solución del subproblema (3.2). De otra forma, la restricción de ángulo está activa, y es fácil verificar que una solución óptima se encuentra en el espacio generado por g_c y b_c . Mediante la rotación de s_{tr} en el plano del $\text{span}\{s_{tr}, b_c\}$ nos encontramos con los dos puntos en este plano en el que la restricción de ángulo se satisface como igualdad, y elegimos uno con el objetivo de un modelo más bajo. Esto proporciona una solución global al subproblema (3.2), que es única, excepto en el caso cuando $b_c^\top g_c = 0$, cuando son exactamente dos soluciones globales.

CAPÍTULO 5

MODELO CUADRÁTICO

Con el fin de definir de manera única un modelo cuadrático

$$m_c(x_c + s) = f(x_c) + g_c^\top s + \frac{1}{2} s^\top G_c s \quad (5.1)$$

que interpole un conjunto de puntos, se necesita determinar los coeficientes $g_c \in \mathfrak{R}^n$, y la matriz simétrica G_c de orden $n \times n$, con un total de $m = \frac{1}{2}(n+1)(n+2) - 1$ escalares desconocidos, en efecto

n para el vector g

n para la primera fila de G

$n - 1$ para la segunda fila de G

$n - 2$ para la segunda fila de G

\vdots

i para la n -ésima fila de G

$$\frac{n(n+1)}{2} + n = \frac{n^2+n+2n}{2} = \frac{n^2+3n}{2} = \frac{1}{2}(n+1)(n+2) - 1$$

Así, además del punto x_c , el cual es la aproximación a la solución x^* en la iteración actual, se requieren de otros m puntos que se denotan como y^1, y^2, \dots, y^m y se denominan puntos satélites. La relación entre estos puntos con el iterado actual x_c viene dado por

$$y^l = x_c + s^l \quad l = 1, \dots, m$$

El modelo cuadrático (5.1) puede ser expresado como (se ignora el subíndice c de los elementos de G_c para mantener la notación simple)

$$\begin{aligned} m_c(x_c + s) &= f(x_c) + g_c^\top s + \sum_{i < j} G_{ij} s_i s_j + \frac{1}{2} \sum_i G_{ii} s_i^2 \\ &\equiv f(x_c) + \hat{g}_c^\top \hat{s} \end{aligned} \quad (5.2)$$

Donde se tienen reagrupados los elementos de g_c y G_c en un vector de m -incógnitas

$$\hat{g}_c \equiv \left(g_c^\top, \{G_{ij}\}_{i<j}, \left\{ \frac{1}{\sqrt{2}} G_{ii} \right\} \right)^\top \quad (5.3)$$

y se define el m -vector

$$\hat{s} \equiv \left(s^\top, \{s_i s_j\}_{i<j}, \left\{ \frac{1}{\sqrt{2}} s_i^2 \right\} \right)^\top$$

Dado que el modelo (5.2) tiene la misma forma que (4.1), la determinación del vector de coeficientes desconocidos \hat{g} será dado como en el caso lineal. Se deduce de (5.2) que las condiciones de interpolación toman la forma

$$(\hat{s}^l)^\top \hat{g}_c = f(y^l) - f(x_c) \quad l = 1, \dots, m \quad (5.4)$$

Donde

$$\hat{s}^l \equiv \left((s^l)^\top, \{s_i^l s_j^l\}_{i<j}, \left\{ \frac{1}{\sqrt{2}} (s_i^l)^2 \right\} \right)^\top \quad l = 1, \dots, m$$

El modelo (5.1) será unívocamente determinado si y sólo si el sistema (5.4) tiene una solución única, o equivalentemente si y sólo si el conjunto

$$\{\hat{s}^l : l = 1, \dots, m\} \quad (5.5)$$

es linealmente independiente. Esta es la condición que el conjunto de puntos de muestra debe satisfacer para ser no-degenerado. La región tabú es definida como

$$\text{span}\{\hat{s}^l : l = 1, \dots, m; \quad l \neq l_{out}\} \quad (5.6)$$

y puede ser expresado como

$$\tau_c = \{\hat{s} \in \mathfrak{R}^m : \hat{b}_c^\top \hat{s} = 0\} \quad (5.7)$$

donde $\hat{b}_c \in \mathfrak{R}^m$ es perpendicular al subespacio (5.6).

Como en el caso lineal, se define una región que contiene a τ_c exigiendo que la magnitud del coseno del ángulo entre \hat{s} y la normal \hat{b}_c sea mayor o igual a un escalar dado $\gamma \in (0, 1)$, esto es

$$|\hat{b}_c^\top \hat{s}| \geq \gamma \|\hat{b}_c\| \|\hat{s}\| \quad (5.8)$$

Todo lo que queda por hacer es expresar esta condición y la región tabú (5.7) en términos de las variables $s \in \mathfrak{R}^n$ del problema original. Escribimos a \hat{b}_c en la forma (5.3) tenemos

$$\hat{b}_c \equiv \left(b_c^\top, \{B_{ij}\}_{i < j}, \left\{ \frac{1}{\sqrt{2}} B_{ii} \right\} \right)^\top \quad (5.9)$$

donde $b_c \in \mathfrak{R}^n$. Ahora se elije B_c como una matriz simétrica de orden $n \times n$ con los elementos de la triangular superior dados por $\{B_{ij}\}_{i < j}$.

La región tabú (5.7) está dada por

$$\tau_c = \left\{ s \in \mathfrak{R}^n : b_c^\top s + \frac{1}{2} s^\top B_c s = 0 \right\} \quad (5.10)$$

y es por lo tanto el conjunto de soluciones para la ecuación cuadrática.

Suponga sin perder generalidad que el vector \hat{b}_c es normalizado y además que

$$\|\hat{b}_c\| = \sqrt{\|b_c\|^2 + \frac{1}{2} \|B_c\|_F^2} = 1$$

Luego la condición de ángulo (5.8) puede ser escrita como

$$|b_c^\top s + \frac{1}{2} s^\top B_c s| \geq \gamma \sqrt{\|s\|^2 + \frac{1}{2} \|ss^\top\|_F^2} \quad (5.11)$$

Dado que

$$\|ss^\top\|_F = \|s\|^2$$

La desigualdad (5.11) puede ser escrita como

$$|b_c^\top s + \frac{1}{2} s^\top B_c s| \geq \gamma \|s\| \sqrt{1 + \frac{1}{2} \|s\|^2}$$

Esto define la restricción de ángulo cuando el modelo es cuadrático.

Para determinar los coeficientes g_c , G_c , b_c y B_c se calcula la factorización QR de la matriz cuyas columnas son los vectores $\{s^l\}$ en (5.5). Usando ésta factorización podemos resolver para \hat{g}_c el sistema (5.4), obteniendo g_c y G_c a través de (5.3). La factorización QR también nos da el vector \hat{b}_c en (5.7) y a través de (5.9), el vector b_c y la matriz B_c .

El paso s_c es entonces definido como una solución aproximada de

$$\min_s m_c(x_c + s) = f(x_c) + g_c^\top s + \frac{1}{2} s^\top G_c s \quad (5.12)$$

$$\text{sujeto a } \|s\| \leq \Delta_c \quad (5.13)$$

$$|b_c^\top s + \frac{1}{2} s^\top B_c s| \geq \gamma \|s\| \sqrt{1 + \frac{1}{2} \|s\|^2} \quad (5.14)$$

El paso s_c es así dependiente del parámetro γ , cuya elección puede tener un impacto en la eficiencia del algoritmo de ángulo. Si γ es demasiado grande, la restricción de ángulo puede descartar pasos que conforman un progreso significativo hacia la solución. Para evitar ésta ineficiencia, se incluye en el algoritmo un procedimiento para hacer decrecer γ si es necesario. La actualización de γ se realizará mientras calcula una solución aproximada para (5.12), como será descrita a continuación.

Cálculo del Paso

Es difícil calcular una solución óptima al subproblema (5.12) dado que G_c puede ser indefinida y la región factible es usualmente no-convexa (ver figura). Sin embargo, se propone la estrategia de concentrarse en encontrar una solución aproximada de (5.12), y se va hacer a través de un procedimiento análogo al empleado en el caso lineal. Mientras se resuelve el subproblema también se determinará si el valor del parámetro de ángulo γ necesita ser reducido.

Primero usamos la técnica descrita por Moré y Sorensen [8] para calcular una solución s_{tr} del problema de región de confianza (5.12), ignorando la restricción de ángulo.

Si s_{tr} satisface (5.12), entonces la solución del subproblema (5.12) será dada por $s_c = s_{tr}$. En otro caso, se inicia desde s_{tr} y se calcula una secuencia de pasos de ensayo s^i , $i = 1, \dots$ con $\|s^i\| = \|s_{tr}\|$, que se mueven lejos de la región tabú a lo largo de una dirección en la cual la violación de la restricción de ángulo (5.12) inicialmente decrece. Generamos pasos de ensayo hasta que la restricción de ángulo se satisface o hasta que el valor del modelo cuadrático (5.12) ha aumentado demasiado respecto a su valor en s_{tr} , en tal caso γ se disminuye.

Más precisamente, se define la parte izquierda en (5.12) por $\phi(s)$, esto es

$$\phi(s) = |b_c^\top s + \frac{1}{2} s^\top B_c s|$$

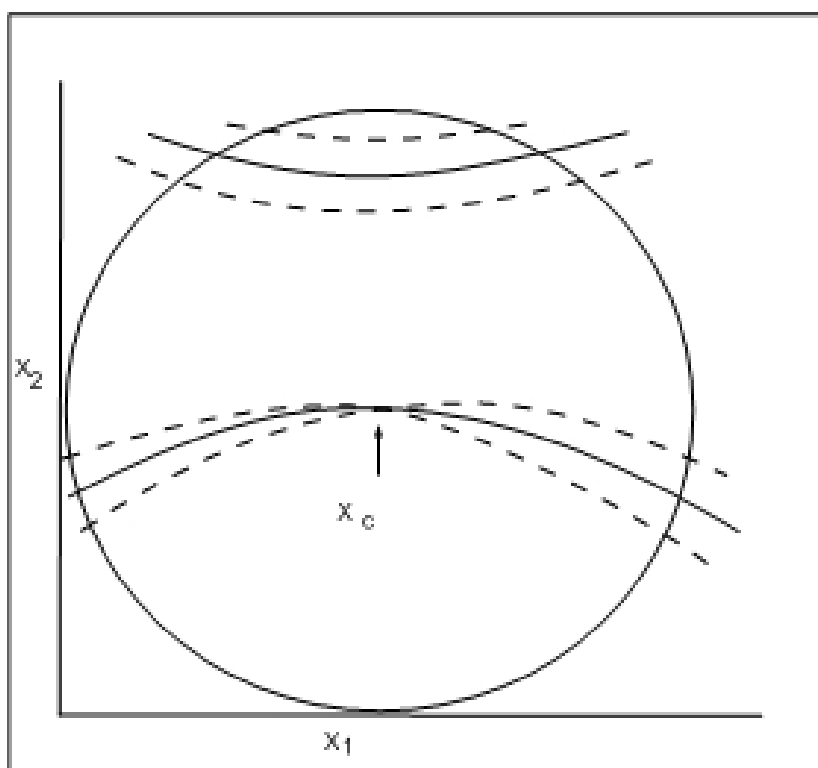


Figura 5.1: Figura 3

El gradiente es dado por

$$\nabla\phi(s) = \text{sign}(b_c^\top s + \frac{1}{2}s^\top B_c s)(B_c s + b_c)$$

Probar que $\phi(s) \neq 0$.

Se calculan los pasos de ensayo s^i , $i = 1, \dots$, por rotación incremental de s_{tr} en el plano del espacio $\text{span}\{s_{tr}, B_c s_{tr} + b_c\}$, y en la dirección de $\nabla\phi(s_{tr})$. Note que $B_c s_{tr} + b_c$ es normal a la región tabú en s_{tr} , y moviendonos a lo largo de esta normal, nos alejamos de la región tabú. (El gradiente de ϕ indica si debemos movernos a lo largo de la normal o el negativo de la misma.)

Por medio de esta rotación, la violación de la restricción de ángulo inicialmente decrece, como ϕ inicialmente decrece y el lado derecho de la restricción (5.12) se mantiene, se cumple la igualdad $\|s^i\| = \|s_{tr}\|$ para todo i .

Como se generan pasos de ensayo s^i , se controla el valor del modelo cuadrático (5.12); continuando con la rotación hasta un paso de ensayo s^i que satisfaga (5.12) o hasta que el valor de m_c ha aumentado demasiado, en el sentido de que

$$\text{pred}(s^i) \geq \frac{1}{2}\text{pred}(s_{tr}) \quad (5.15)$$

donde $\text{pred}(s) = m_c(x_c) - m_c(x_c + s)$ con $s \in \mathfrak{R}^n$. Se toma ésto como una indicación de que el parámetro de ángulo γ es demasiado grande. Entonces el conjunto $s_c = s^i$ y “cerramos el ángulo”, esto es, se reduce γ , de modo que el paso de ensayo más reciente s^i satisface (5.12) como una igualdad

$$\gamma = |b_c^\top s + \frac{1}{2}s^\top B_c s_c| / \left(\sqrt{\|s_c\|(1 + \frac{1}{2}\|s_c\|^2)} \right) \quad (5.16)$$

Esta nueva elección de γ se pasará a la siguiente iteración, de modo que los valores de γ forman una secuencia no creciente.

Este procedimiento de cálculo de pasos es descrito a continuación. Aquí $s(\theta)$ denota la rotación de s_{tr} , por un ángulo θ , en el espacio de $\text{span}\{s_{tr}, B_c s_{tr} + b_c\}$ en la dirección de $\nabla\phi(s_{tr})$.

Procedimiento Paso Cuadrático

Los parámetros de entrada son g_c , G_c , B_c , γ , Δ_c y $\delta\theta$

Resuelva el subproblema de región de confianza (12a)-(12b) para obtener s_{tr} .

If s_{tr} satisface (12c), se rota $s_c = s_{tr}$, y γ ; **Stop**

Else

El conjunto $\theta = \delta\theta$

Repeat Hasta que $s(\theta)$ satisfaga (12c) ó $pred(s(\theta)) \geq \frac{1}{2}pred(s_{tr})$

$\theta \leftarrow \theta + \delta\theta$

End Repeat

El conjunto $s_c = s(\theta)$

If s_c viola (12c), se define γ por (5.12)

End If

Regresar s_c y γ ; **Stop**

CAPÍTULO 6

RESULTADOS NUMÉRICOS

Para evaluar la robustez y eficiencia del algoritmo de ángulo, se realizó una comparación con otros tres métodos para optimización libre de derivadas en una selección de problemas de la colección CUTE. En esta sección se denota la k_{th} iteración por x_k ; el subíndice k será además usado en todas las cantidades asociadas a x_k ($f_k = f(x_k)$, Δ_k , etc).

Los parámetros de región de confianza en el paso 5 del algoritmo 1 se establecieron para $\alpha = 0$ y $\beta = \frac{1}{2}$, y la estrategia de actualización de región de confianza como la siguiente:

```
If  $ared(s_k) \leq 0$ 
 $\Delta_{k+1} = \frac{1}{2} \|s_k\|$ 
Else
If  $\|s_k\| = \Delta_k$ ,  $\Delta_{k+1} = 2\Delta_k$ 
Else  $\Delta_{k+1} = \Delta_k$ 
End Else
```

Otras estrategias son posibles, pero el que se describe acá parece funcionar bien en la práctica, ya que permite que el radio de región de confianza Δ se incremente rápidamente, permitiendo pasos más grandes.

Primero se prueba la versión lineal del algoritmo de ángulo (WEDlin) y COBYLA, un método de región de confianza que usa modelo de interpolación lineal. El punto inicial x_0 en este y todos los resultados reportados, fue suministrado por CUTE. Los $n+1$ satélites iniciales requeridos por la primera iteración de WEDlin se definen como

$$y^i = x_0 \pm \Delta_0 e_i, \quad i= 1, \dots, n \quad (6.1)$$

donde el radio de región de confianza inicial se estableció en $\Delta_0 = 1$, e_i denota el i -ésimo vector canónico elegido al azar. Para tener en cuenta la aleatoriedad introducida en la selección de los puntos de muestra iniciales en ambas versiones, lineal y cuadrática del algoritmo de ángulo, cada problema de prueba se corrió cinco veces. Los resultados de la mediana, en términos de evaluación de las funciones son reportados.

El parámetro de ángulo en WEDlin se eligió como $\gamma = 0,4$, y se mantuvo constante a lo largo de la iteración (sólo el algoritmo que usa modelos cuadráticos reduce el valor de γ). El *rhoend* (tamaño del simplex en la terminación) en COBYLA se estableció para $10 \times \text{macheps}$, donde *macheps* denota redondeo de unidades de doble precisión. En forma similar, WEDlin concluye si

$$\Delta_k \leq 10 \times \text{macheps} \quad (6.2)$$

Se toma como una indicación de que ningún progreso se puede hacer.

Las pruebas de parada para WEDlin y COBYLA fueron de la siguiente manera. En primer lugar resolver cada subproblema usando el paquete de programas KNITRO, que para problemas sin restricciones equivale a un Método de Newton usando segundas derivadas exactas. KNITRO se detuvo cuando $\|\nabla f(x_k)\| < 10^{-6}$ y su último valor de la función f se registró. WEDlin y COBYLA se detienen cuando

$$\frac{f_k - f_\Lambda}{f_0 - f_\Lambda} \leq \epsilon, \text{ o equivalentemente } f_0 - f_k \geq (1 - \epsilon)(f_0 - f_\Lambda) \quad (6.3)$$

Donde $\epsilon = 10^{-6}$ y f_0 es el valor objetivo en el punto inicial. Por lo tanto, se requiere que la disminución $f_0 - f_\Lambda$ obtenida por el algoritmo es al menos $(1 - \epsilon)$ veces la disminución obtenida por el KNITRO. Un límite de 8000 evaluaciones de la función se impuso en todas las corridas.

Los resultados reportan que el número de evaluaciones de la función y el valor final de la función objetivo obtenida por cada algoritmo. Además se reporta (% wed act) el porcentaje de iteraciones en WEDlin donde la restricción de ángulo está activa.

Aunque COBYLA fue diseñada principalmente para optimización con restricciones, estos resultados sugieren que WEDlin es competitivo con respecto a COBYLA en problemas sin restricciones. Es interesante observar además que la restricción de ángulo está activa en una fracción significativa de las iteraciones, mostrando que ésta restricción juega un papel importante en el método.

Comentaremos ahora sobre las terminaciones anormales de COBYLA, el cuál está escrito en precisión simple.

1. En el problema AKIVA, la versión de precisión simple de CUTE da un valor de la función inicial $f(x_0) = 2,70488$, que es menor que el valor objetivo óptimo obtenido por KNITRO, $f_\Lambda = 6,16604$. Por lo tanto, COBYLA satisface (6.3), terminando así en el punto de partida, sin realizar una sola iteración. Sin embargo, cuando corremos WEDlin en doble precisión, obtenemos el valor de la función inicial $f(x_0) = 14,5561$, que es muy superior al valor de parada.
2. COBYLA se bloquea cuando se trata de evaluar f en un punto donde no está definida.
3. COBYLA se detuvo porque el tamaño de la simplex (rhoend) es menor que ó igual a $10 \times macheps$; éste es un sistema incorporado en la detención de prueba.

A continuación se compara la versión cuadrática del algoritmo de ángulo (WEDquad) con el Cuasi Newton código L-BFGS-B usando diferencias finitas para aproximar el gradiente (QNfinDiff) y todo sus valores predeterminados.

El valor inicial del parámetro de ángulo en WEDquad fue elegido como $\gamma = 0,4$, y como fue explicado en el modelo cuadrático, se le permite cambiar con las iteraciones. El valor de $\delta\theta = \frac{\pi}{600}$ se utilizó en el procedimiento de paso cuadrático. Los $\frac{1}{2}(n+1)(n+2)$ puntos de muestra requeridos para definir el modelo de interpolación cuadrática inicial puede ser elegido como los vértices y los puntos medios de un

simplex. Esto se hizo en WEDquad, usando el simplex definido por x_0 y los puntos (6.1).

Además del número de evaluaciones de la función, el valor de la función final y el porcentaje de iteraciones (% wed act) en el que la restricción de ángulo era activa, también informamos sobre el valor final del parámetro de ángulo γ . Como se mencionó anteriormente, WEDquad se ejecutó cinco veces y la mediana de los resultados son reportados. La prueba de parada fue (6.3).

Los resultados muestran que WEDquad es a veces, pero no siempre, más eficiente que el algoritmo de diferencia finita Cuasi-Newton. Sin embargo WEDquad parece ser más confiable, con una falla en comparación con cinco fallas de QNfinDiff.

En el problema AKIVA, QNfinDiff devuelve el valor de la función $-\infty$ en la segunda iteración y terminado. La mayor parte de las fallas de QNfinDiff se atribuyen a la falta de progreso en la línea de búsqueda debido a errores en la aproximación por diferencias finitas a los gradientes (el código usa diferencias en adelante).

Finalmente, comparamos WEDquad con el código DFO que implementa método de región de confianza usando modelos de interpolación cuadrática. Al momento de redactar DFO no está disponible para el público, y las comparaciones descritas a continuación se basan en los resultados presentados en [5] más los resultados acerca de los problemas adicionales suministrados por Katya Scheinberg [13]. Por esta razón, WEDquad se detuvo cuando

$$f_k \leq f_{DFO} \tag{6.4}$$

donde f_{DFO} denota el valor objetivo final obtenido por DFO. Ambos algoritmos se ejecutaron cinco veces para cada problema y la mediana de los resultados fueron reportados. Los satélites iniciales y todos los otros parámetros de WEDquad se eligieron como en la prueba anterior.

Es difícil diseñar una prueba de parada adecuada para la comparación de métodos libre de derivadas, y ésta última comparación es un ejemplo de ello. El código DFO puede requerir un significativo número de evaluaciones para reconocer conver-

gencia en su mejor valor de f [3, 13] (que ocurre por defecto cuando $\Delta_k < 10^{-4}$). Como resultado, utilizando la prueba de parada (6.4) puede resultar adverso a DFO. Por otro lado WEDquad no emplea ningún mecanismo especial de terminación, su prueba de parada por defecto es (6.2), y en este sentido es completamente análoga al método de región de confianza basado en gradiente.

Se pudo observar que en la mayoría de los problemas el parámetro de ángulo γ es reducido significativamente a partir de su valor inicial de 0,4, y que al mismo tiempo la restricción de ángulo está activa con cierta frecuencia. Por lo tanto, el algoritmo parece haber logrado un buen balance entre la condición geométrica requerida y el deseo de permitir todos los pasos de región de confianza cuando sea posible.

Se concluye señalando que la elección del satélite inicial en WEDquad con los vértices y puntos medios de un simplex no es una estrategia eficiente puesto que requiere $O(n^2)$ evaluaciones de la función para iniciar el algoritmo. Sería más eficiente usar, por ejemplo modelos lineal (o indeterminado cuadrático) durante las primeras iteraciones, pero todavía no se ha comprobado esta opción. Entre otros refinamientos que puedan mejorar el rendimiento se incluye la opción de incrementar el parámetro ángulo en ciertas iteraciones, esto se puede hacer ya sea directamente (esto es, cuando s_{TR} en el procedimiento de paso cuadrático satisface la restricción de ángulo); o indirectamente relajando el (muy exigente) requisito de que la solución aproximada del subproblema cuadrático satisfaga la prueba (5.15). Una prueba alternativa podría ser $pred(s^i) \geq \frac{1}{2}pred(s_{CA})$, donde s_{CA} es el punto de Cauchy [10] para el subproblema (modelo cuadrático).

REFERENCIAS

- [*Marazzi y Nocedal*] Marcelo Marazzi, Jorge Nocedal, Wedge trust region methods for derivative free optimization, Math. Program., Ser. A 91: 289-305 (2002)
- [*Jorge Nocedal y Stephen J. Wright*] Numerical Optimization, Second Edition, 2000.
- [*Andrew R. Conn, Nicholas J. M. Gould y Phillippe L. Toint*] Trust Region Methods. MPS-SIAM series on optimization, 2000.
- [*D. Winfield 1973*] D. Winfield, Function Minimization by interpolation in data tables, Journal of the Institute of Mathematics and its Applications, 12, pp. 339-347.
- [*M. J. D. Powell 1994 a*] M. J. D. Powell, A direct search optimization methods that models the objective and constraint functions by interpolation. In S. Gomez, and J. P. Hennart, eds. Advances in optimization and Numerical Analysis, Oaxaca, Mexico. Vol. 275, pp. 51-67, Kluwer Academic Publishers, Dordrecht, the Netherlands
- [*M. J. D. Powell 1994 b*] M. J. D. Powell, A direct search optimization methods that models the objective by quadratic interpolation. Presentation at the 5th Stockholm Optimization Days, Stockholm, Sweden
- [*A. R. Conn and Ph. L. Toint 1996*] . An algorithm using quadratic interpolation for unconstrained derivative free optimization. In G. Di Pillo and F. Gianessi, eds., Nonlinear Optimization and Applications, pp. 27-47, Plenum, New York.