UNIVERSIDAD DE GUANAJUATO

DEPARTAMENTO DE MATEMÁTICAS
CAMPUS GUANAJUATO

An efficient gradient-free line search

INTEGRANTES:

Dan Heli Muñiz Sanchez
Bryan Calderón Rivera

PROFESOR:

Oscar S. Dalmau Cedeño

03 de junio del 2023

Introducción

Se describe un nuevo método de optimización para funciones continuamente diferenciables que utiliza una trayectoria de búsqueda suave y arbitraria. Este método utiliza evaluaciones de la función en lugar de gradientes al tratar de encontrar un paso optimo y garantiza la convergencia global a un punto estacionario para funciones acotadas por debajo y con gradientes continuos de Lipschitz. El método establece límites de complejidad y ofrece una alternativa interesante a otros métodos de optimización que dependen del cálculo de gradientes, especialmente en casos donde el cálculo de gradientes es costoso o impracticable.

Resumen

En el artículo presentado se analiza un nuevo método de búsqueda lineal llamado CLS, el cual es un método eficiente que no requiere de evaluaciones adicionales del gradiente. Se procede a buscar puntos $x(\alpha)$ en una curva direccional diferenciable p, hasta encontrar un tamaño de paso α tal que $f(x(\alpha))$ sea menor que $f(x)$.
Para este método utilizamos como condición de descenso

$$\mu(\alpha)|\mu(\alpha)-1| \geq \beta$$

con $\beta > 0$ fijo, donde $x(\alpha) = x + \alpha p$ y

$$\mu(\alpha) = \frac{f(x(\alpha)) - f(x)}{\alpha g(x)^T p} \text{ para } \alpha > 0$$

que es una variación de Goldstein pero mas fácil de satisfacer. Aun así, no admite valores que son muy pequeños o muy grandes, puesto que condición requiere que $\mu(\alpha)$ no este muy cerca de uno y que no sea muy pequeño a su vez.
CLS se comporta bien en regiones que no son fuertemente convexas y acepta un mayor rango de pasos significativos que las condiciones de Goldstein o Wolfe, además de garantizar la terminación para funciones cuadráticas estrictamente convexas después de dos iteraciones como máximo.

Desarrollo

Se plantea un nuevo método de busqueda en linea para resolver el problema de optimización

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

donde f es una función continuamente diferenciable, acotada por abajo y con gradiente continuo Lipschitz.

La teoria de convergencia estandar nos dice que para que un método de busqueda en linea sea eficiente en el sentido de que al momento de regresar un paso fijo α se cumpla que el cociente

$$(f(x) - f(x + \alpha p))\frac{||p||^2}{(g(x)^T p)^2}$$

esta acotado por un número positivo fijo. La mayoria de los metodos de optimizacion utiliza las condiciones de Wolfe para encontrar el tamaño de paso α, las condiciones de Wolfe se mantienen(para μ y ν fijos) si se satisface la condición de Armijo

$$f(x + \alpha p) \leq f(x) + \alpha \mu g(x)^T p, \text{ con } 0 < \mu < 1$$

y la condición de curvatura

$$g(x+\alpha p)^T p \geq \nu g(x)^T p, \text{ con } \mu \leq \nu < 1.$$

Pero para esto se requiere evaluar la función y el gradiente en el punto $x + \alpha p$ para la condición de Armijo y de curvatura respectivamente.
En cambio, las condiciones de Goldstein

$$f(x) + \alpha\mu_2 g(x)^T p \leq f(x+\alpha p) \leq f(x) + \alpha\mu_1 g(x)^T p$$

para valores fijos $0 < \mu_1 < \mu_2 < 1$, el cual no necesita una evaluación adicional del gradiente antes de determinar la proxima iteración. Estos metodos de busqueda garantizan tamaños de paso eficientes y por eso pueden ser utilizados para algoritmos de convergencia global. Sin embargo, tecnicas de backtracking y las condiciones de Goldstein a menudo funcionan mal en regiones fuertemente no convexas, un problema parcialmente superado por las condiciones de Wolfe.

Una vista al método

Comparado a los enfoques tradicionales, CLS muestra nuevas caracteristicas:

- CLS utiliza una nueva condición suficiente de descenso, mas tolerante que las busquedas en linea Goldstein o Wolfe.
- La condición de descenso suficiente es una variación de la condición de Goldstein. Es mas facil de satisfacer, pero aun asi no acepta valores que sean demasiado pequeños o demasiado grandes, y por tanto conduce a una reducción razonable del valor de la función.
- CLS es una busqueda en linea eficiente en el sentido formal sin requerir evaluaciones adicionales del gradiente.
- CLS se comporta bien en regiones no fuertemente convexas, acepta un rango mucho mayor de pasos significativos que las condiciones de Goldstein o Wolfe.
- Para funciones cuadraticas estrictamente convexas, la terminación despues de dos iteraciones esta garantizada.
- Una condición de descenso suficiente, junto a una condición debil en la busqueda de direccion es suficiente para la convergencia global.

Cociente de Goldstein

Una buena y computacionalmente util medida del progreso de una busqueda en linea es el cociente de Goldstein

$$\mu(\alpha) = \frac{f(x(\alpha)) - f(x)}{\alpha g(x)^T p} \text{ para } \alpha > 0.$$

Condición de descenso suficiente

Al asumir que $g^T p > 0$, tenemos que $f(x(\alpha)) < f(x)$ ssi $\alpha > 0$ y $\mu(\alpha) > 0$. Estas restricciones en los valores del cociente de Goldstein definen regiones donde el descenso se puede alcanzar.

Para este método consideramos como condición de descenso suficiente(SDC)

$$\mu(\alpha)|\mu(\alpha) - 1| \geq \beta \quad (3,1,1)$$

para un valor fijo $\beta > 0$. Esta condición requiere cumplir ambas condiciones, $\mu(\alpha)$ no este muy cerca a 1, asi evadiendo tamaños de paso que sean muy cortos, y a su vez que sean suficientemente positivos, evadiendo pasos muy largos forzando $f(x(\alpha)) < f(x)$. Dicha condición es mas facil de satisfacer que la condición de Goldstein. Garantizar la condición de descenso suficiente garantiza un sensible decrecimiento de la función objetivo. La condición de Goldstein la cual al trabajar con gráficas concavas y muy planas permite solo un rango muy pequeño de tamaños de paso muy ineficientes, la SDC evita este defecto sin necesidad de evaluaciones adicionales del gradiente.

Satisfaciendo la condición de descenso suficiente

Si se toma $\mu = 1/2$ tiene una importancia especial. Cerca de un minimizador local, las funciones dos veces continuamente diferenciables están acotadas por debajo y, debido al teorema de Taylor, son casi cuadráticas. Para una trayectoria de búsqueda lineal y una función cuadrática estrictamente convexa,

$$\begin{aligned} f(x+\alpha p) &= f(x) + \alpha g(x)^T p + \frac{\alpha^2}{2} p^T G(x) p \\ &=: f + a\alpha + \alpha^2 b \\ &= f - \frac{a^2}{4b} + b(\alpha - \hat{\alpha}) \end{aligned}$$

y por las condiciones en las secciones anteriores , se tiene que

$$a < 0 < b, \quad \hat{\alpha} = -\frac{a}{2b}$$

y despues de desarrollar se obtiene que

$$\mu(\alpha) = 1 - \frac{\alpha}{2\hat{\alpha}}$$

y que si $\alpha > 0$,$\mu(\alpha) = 1/2$ y el minimizador $\hat{\alpha} = \frac{\alpha}{2(1-\mu(\alpha))}$

por lo anterior y porque f la funcion objetivo debe ser acotada, entonces podemos enunciar el siguiente teorema.

Teorema 1. *sea un $\beta \in (0, 1/4)$, $g(x)^T p < 0$ la ecuacion $\mu(\hat{\alpha}) = 1/2$ tiene una solución $\alpha > 0$ tal que α cumple la* ***Condicion de suficiente descenso***

Algoritmo

El algoritmo pide unos requerimientos $v = -g(x)^T x'(0) > 0$, α_{init}(tamano de paso inicial),$\alpha_{\text{máx}}$(tamano de paso maximo) tales que

$$0 < \alpha_{init} < \alpha_{\text{máx}} < \infty$$

y los valores con lo que se empieza es $\alpha = \alpha_{init}$, $\alpha_{\text{máx}} = \infty$, first $= 1$

Algorithm 1 CLS

Require: $x(\alpha)$ (camino de busqueda) , f_0 , v (Derivada direccional),
Ensure: α que cumple $\mu(\alpha)|\mu(\alpha) - 1| \geq \beta$

```
while TRUE do
  Calcular el cociente de Goldstein μ(α) = (f_0 − f(x(α)))/(αv)
  if μ(α)|μ(α) − 1| ≥ β then
    break
  end if
  if μ(α) > 1/2 then
    α_ = α        (α underlined)
  else if α = α_máx then
    break
  else
    ᾱ = α
  end if
  if first then
    first = 0
    if μ(α) < 1 then
      α = ½ α/(1 − μ(α))
    else
      α = αQ
    end if
  else
    if ᾱ = ∞ then
      α = αQ
    else if α_ = 0 then
      α = ½ α/(1 − μ(α))
    else
      α = √(ᾱ α_)
    end if
  end if
  Restringir mín(α, α_máx)
end while
return α
```

while TRUE **do**

  Calcular el cociente de Goldstein $\mu(\alpha) = (f_0 - f(x(\alpha)))/(\alpha v)$

  if $\mu(\alpha)|\mu(\alpha) - 1| \geq \beta$ **then**

    break

  end if

  if $\mu(\alpha) > 1/2$ **then**

    $\underline{\alpha} = \alpha$

  else if $\alpha = \alpha_{\text{máx}}$ **then**

    break

  else

    $\overline{\alpha} = \alpha$

  end if

  if first **then**

    first $= 0$

    if $\mu(\alpha) < 1$ **then**

      $\alpha = \frac{1}{2}\alpha/(1 - \mu(\alpha))$

    else

      $\alpha = \alpha Q$

    end if

  else

    if $\overline{\alpha} = \infty$ **then**

      $\alpha = \alpha Q$

    else if $\underline{\alpha} = 0$ **then**

      $\alpha = \frac{1}{2}\alpha/(1 - \mu(\alpha))$

    else

      $\alpha = \sqrt{\overline{\alpha}\underline{\alpha}}$

    end if

  end if

  Restringir $\text{mín}(\alpha, \alpha_{\text{máx}})$

end while

return α

Analisis de Complejidad

Si tomamos parametros $0 < \kappa < \lambda < \infty$, fijos y ajustables, para que podamos reestringir los valores de α_{init} y $\alpha_{\text{máx}}$ para lograr que las siguientes desigualdades se cumplan.

$$\frac{\kappa v}{||p||^2} \leq \alpha_{init} \leq \alpha_{\text{máx}} \leq \frac{\lambda v}{||p||^2} \quad (5{,}1)$$

Para que se cumpla el siguiente resultado de complejidad

Teorema 2. *Sean* $0 < \kappa < \lambda < \infty$ *Si la desigualdad (5.1) es verdadera entonces el numero de evaluaciones de funciones en el algoritmo **CLS** esta acotado por una constante en funcion de los valores* $Q, \beta, \kappa, \lambda$ *y la constante de Lipschitz* γ*,*

1. *Si* $\mu(\alpha_{init}) > c_2$ *entonces el algoritmo **CLS** acaba a los mas en* $\overline{L_E} + \overline{M_E}$ *evaluaciones de funciones donde*
$$\overline{L_E} = \left\lceil \frac{\log \frac{Q\lambda}{\kappa}}{\log Q} \right\rceil, \quad \overline{M_E} = \left\lceil \log_2 \frac{\gamma\lambda \log Q}{2(c_2 - c_1)} \right\rceil$$
2. *Si* $\mu(\alpha_{init}) < c_1$ *entonces el algoritmo **CLS** acaba a los mas en* $\overline{L_Q} + \overline{M_Q}$ *evaluaciones de funciones donde*
$$\overline{L_Q} = \left\lceil \frac{\log \frac{\gamma\lambda}{\kappa}}{\log(2 - 2c_1)} \right\rceil, \quad \overline{M_Q} = \left\lceil \log_2 \frac{\gamma\lambda \log(\gamma\lambda)}{2(c_2 - c_1)} \right\rceil$$
3. *Solo una iteracion , con una sola evaluacionde la funcion.*

Este teorema nos lleva directamente a otro teorema de complejidad

Teorema 3. *Dados* $0 < \kappa < \lambda < \infty$ *supongamos que las direcciones de busqueda de metodo* (d_k) *cumplen que*

$$\frac{g_k^T d_k}{||g_k|| ||d_k||} < -\delta < 0$$

para algun $\delta > 0$ *con los mismos pasos iniciales. y se cumple la desigualdad (5.1) entonces*

1. *El numero de iteraciones requeridas para alcanzar el punto* x^* *es tal que* $||g(x^*)|| \leq \epsilon$ *es de* $O(\epsilon^{-2})$
2. *Si el conjunto de cuervas de nivel es acotado entonces, partiendo desde el punto inicial* x_0 *por una subsecuencia de numeros converge a* x^*
3. *Si* f *tiene un minimizador local fuerte* $\hat{x}$ *y no hay ningún otro punto estacionario, entonces el número de valores de función necesarios para alcanzar un punto* x *con tal que* $||g(x)|| \leq \varepsilon$ *es* $O(\log \epsilon^{-1})$.

Resultados

Por la teoría antes mencionada, esta algoritmo funciona específicamente para funciones acotadas por abajo, asi pues podemos trabajar con funciones ya previamente trabajadas como lo son

1. **Función Wood**

$$f(\mathbf{x}) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2 + (x_3 - 1)^2 + 90(x_3^2 - x_4)^2 + 10{,}1((x_2 - 1)^2 + (x_4 - 1)^2) + 19{,}8(x_2 - 1)(x_4 - 1)$$

2. **Función Rosembrook**

$$f(\mathbf{x}) = \sum_{i=1}^{N-1} 100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2$$

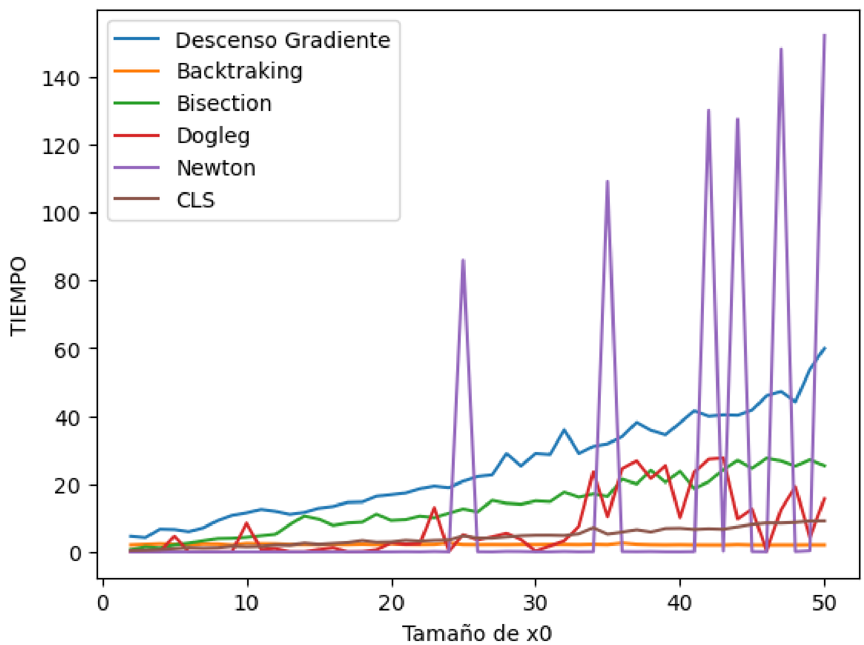
Si generamos puntos iniciales aleatorios $\mathbf{x_0} = (x_1, x_2, \cdots, x_n)$ tal que

$$x_i \sim \mathcal{N}(0,1) \quad i = 1, 2, \cdots, n$$

Obtenemos la siguiente tabla del promedio de iteraciones y de tiempo para encontrar el minimo de la funcion **Wood**

Metodo	Promedio de tiempo(s)	Promedio Num. Iteraciones
Descenso de Gradiente	4.29	159,681.3
Backtracking	1.12	14,353.6
Bisection	0.87	7720
Dogleg	0.02	21
Newton	0.0013	14.6
CLS	0.25	2666.6

Ahora al poner el ojo sobre la función **Rosembrook** es de interes cuando se tarda cada metodo a medida que aumentamos la dimension de $\mathbf{x_0} \in^n$ cuando variamos $n = 2, 3, \cdots, 50$, y al plantear ese experimento estos son los resultados de cada algoritmo.



Lo cual evidencia que el algoritmo **CLS** no solo se le puede decir que es *eficiente* si no que también bastante estable, puesto que no hay picos de aumento de tiempos como se ven en otros métodos.

Conclusiones

Basándonos en los resultados obtenidos al ejecutar los diferentes métodos de optimización en las funciones de Wood y Rosenbrock, podemos concluir que nuestro algoritmo CLS ha demostrado ser altamente efectivo al converger al mínimo global en un número considerablemente menor de iteraciones en comparación con los algoritmos de backtrack, bisección y descenso del gradiente.

Un aspecto destacado de nuestro algoritmo CLS es que no requiere evaluaciones adicionales del gradiente, lo que contribuye a reducir el tiempo necesario para cada iteración. Esto significa que nuestro algoritmo puede alcanzar resultados óptimos de manera más eficiente en términos de tiempo de ejecución.

Además, el hecho de que la condición de descenso suficiente sea más fácil de cumplir en nuestro algoritmo CLS en comparación con las condiciones de Wolfe también tiene un impacto positivo en el tiempo requerido para la convergencia. Esto asegura una mejora significativa en el tiempo total necesario para alcanzar el mínimo global de la función objetivo.

Además de los resultados efectivos y eficientes en la convergencia al mínimo global, otra ventaja significativa de nuestro algoritmo CLS es que su complejidad está acotada en función de los valores iniciales utilizados.

Esta característica es especialmente valiosa, ya que nos brinda un control adicional sobre el rendimiento y la eficiencia del algoritmo. Al elegir cuidadosamente los valores iniciales, podemos influir en la convergencia y garantizar una mayor rapidez en el proceso de optimización.

Bibliografia

1. Neumaier, A. (2004). An efficient gradient-free line search. Fakultät für Mathematik, Universität Wien, Oskar-Morgenstern-Platz 1, A-1090 Wien, Austria.

2. Nocedal, J., & Wright, S. J. (2006). Numerical optimization. Springer.