Fórmulas Broyden y Powell-Symmetric-Broyden (PSB)

Tarea

24 de febrero, 2015

Motivación para formular métodos basados en el método de la secante

Consideremos el problema de encontrar la raíz de una función $f : \mathbb{R} \to \mathbb{R}$ utilizando modelos locales de f. El modelo para obtener el iterando x_{k+1} se formula como

$$m_k(x) = f(x_k) + b_k(x - x_k),$$

en donde b_k juega el papel de la derivada de f en x_k. Notar que si $b_k = f'(x_k)$ tendríamos el método de Newton; sin embargo estamos interesados en métodos que NO utilicen derivadas de f. Es importante resaltar que el modelo satisface la condición

$$m_k(x_k) = f(x_k),$$

para cualquier valor de b_k. Para obtener un único valor de b_k recurrimos al iterando anterior x_{k-1} y forzamos al modelo a satisfacer la condición

$$m_k(x_{k-1}) = f(x_{k-1}) = f(x_k) + b_k(x_{k-1} - x_k),$$

de donde obtenemos

$$b_k(x_k - x_{k-1}) = f(x_k) - f(x_{k-1}).$$

La igualdad anterior es llamada *la condición de la secante*, de donde es inmediato obtener a b_k como:

$$b_k = \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}.$$

En el caso $F : \mathbb{R}^n \to \mathbb{R}^n$ el modelo adopta la forma

$$M_k(x) = F(x_k) + B_k(x - x_k),$$

de donde, por analogía con el caso unidimensional, es trivial obtener la condición de la secante

$$B_k(x_k - x_{k-1}) = F(x_k) - F(x_{k-1}).$$

Notar que de la condición anterior NO es posible determinar de manera única a B_k.

En el área de métodos cuasi-Newton se acostumbra definir las diferencias entre iterandos consecutivos y sus correspondientes valores funcionales como:

$$\begin{aligned} s_{k-1} &= x_k - x_{k-1} \\ y_{k-1} &= F(x_k) - F(x_{k-1}). \end{aligned}$$

Por lo tanto la condición de la secante queda definida como

$$B_k s_{k-1} = y_{k-1}.$$

Entonces B_k se obtiene a partir de B_{k-1} y la pareja de vectores (s_{k-1}, y_{k-1}).

Actualización de Broyden

Consideremos el siguiente problema de optimización estudiado en clase

$$\text{minimizar} \qquad \phi(m) = \frac{1}{2}\sum_{i,j}(m_{ij} - b_{ij})^2 \tag{1a}$$

$$\text{sujeta a} \qquad Am - y = 0, \tag{1b}$$

en donde

$$b = [b_{11}b_{12}\dots b_{1n}\,|\,b_{12}b_{22}\dots b_{2n}\,|\,\dots|b_{n1}b_{n2}\dots b_{nn}]^T$$

$$A = \left[\begin{array}{cccc|cccc|c|cccc} s_1 & s_2 & \dots & s_n & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & s_1 & s_2 & \dots & s_n & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & s_1 & s_2 & \dots & s_n \end{array}\right], \tag{2}$$

los escalares $s_i, \quad i = 1, 2, \dots n$, son componentes del vector s.

El problema de optimización es estrictamente convexo y las condiciones de primer orden de KKT son suficientes para asegurar la existencia y unicidad de un minimizador global. Tal mininimizador se puede obtener resolviendo el siguiente sistema de KKT

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} m \\ -\lambda \end{bmatrix} = \begin{bmatrix} Hb \\ y \end{bmatrix}, \tag{3}$$

en donde H es la matriz identidad.

De el primer bloque de (3) tenemos

$$m = b + A^T\lambda. \tag{4}$$

Un ejercicio relativamente sencillo consiste en probar que el vector $A^T\lambda$ es precisamente la forma extendida de la matriz de rango uno

$$\lambda s^T.$$

Si organizamos (4) en notación matricial estándar, tenemos que B se actualiza como

$$M = B + \lambda s^T. \tag{5}$$

El multiplicador de Lagrange se puede calcular de la expresión

$$AA^T\lambda = y - Ab,$$

obtenida mediante la combinación de los bloques de (3), en donde es fácil verificar que $AA^T = (s^Ts)I$. Por lo tanto,

$$\lambda = \frac{y - Bs}{s^Ts}, \tag{6}$$

lo cual resulta en la fórmula de la actualización de Broyden reportada en la mayoría de los libros de texto

$$M = B + \frac{(y - Bs)s^T}{s^Ts}. \tag{7}$$

Fórmula PSB

Considera el siguiente problema de optimización convexa para construir la fórmula PSB.

$$\begin{aligned} \text{minimizar} \qquad & \phi(m) = \frac{1}{2}\sum_{i=1}^{n}(m_{ii} - b_{ii})^2 + \sum_{i<j}(m_{ij} - b_{ij})^2 & \text{(8a)} \\ \text{sujeta a} \qquad & Am - y = 0, & \text{(8b)} \end{aligned}$$

en donde es conveniente definir la forma extendida de una matriz simétrica B considerando la parte triangular superior de B, es decir:

$$b = [b_{11}b_{12}\ldots b_{1n}\,|\,b_{22}b_{23}\ldots b_{2n}\,|\,\ldots|b_{nn}]^T. \tag{9}$$

A es la matriz en bloques $n \times n(n+1)/2$

$$A = \left[\begin{array}{cccc|cccc|c|c} s_1 & s_2 & \ldots & s_n & 0 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & s_1 & \ldots & 0 & s_2 & s_3 & \ldots & s_n & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & s_2 & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & s_1 & 0 & 0 & \ldots & s_2 & \cdots & s_n \end{array}\right], \tag{10}$$

en donde los escalares $s_i, \quad i = 1, 2, \ldots n$ son las componentes del vector s.

El correspondiente sistema de Karush-Kuhn-Tucker es

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} b_+ \\ -\lambda \end{bmatrix} = \begin{bmatrix} Hb \\ y \end{bmatrix}, \tag{11}$$

en donde H es diagonal con entradas en el conjunto $\{1, 2\}$. Las entradas iguales a 1 corresponden a los elementos diagonales m_{ii}, mientras que las entradas iguales a 2 corresponden a elementos fuera de la diagonal m_{ij}.

El problema ahora consiste en resolver el sistema de KKT (11); del primer bloque de ecuaciones tenemos la expresión

$$b_+ = b + H^{-1}A^T\lambda, \tag{12}$$

que indica que la fórmula para actualizar b depende del vector de multiplicadores de Lagrange λ y de la matriz B. Es fácil probar que el vector $H^{-1}A^T\lambda$ es la forma extendida de la matrix de rango-2

$$\frac{1}{2}(s\lambda^T + \lambda s^T).$$

Por lo tanto, organizando (12) en la forma matricial estándar, tenemos que

$$B_+ = B + \frac{1}{2}(s\lambda^T + \lambda s^T). \tag{13}$$

El cálculo del vector de multiplicadores de Lagrange se hace combinando (12) con el segundo bloque de (11)

$$AH^{-1}A^T\lambda = y - Ab = y - Bs. \tag{14}$$

De la definición de A es fácil prober que el producto $AH^{-1}A^T$ se puede representar como el siguiente múltiplo de una corrección de rango-1 de la matriz identidad

$$\frac{1}{2}s^Ts\left(I + \frac{ss^T}{s^Ts}\right).$$

Por lo tanto la inversa de $AH^{-1}A^T$ existe y se puede obtainer mediate la fórmula de Sherman-Morrison.

$$(AH^{-1}A^T)^{-1} = \frac{2}{s^Ts}\left(I - \frac{ss^T}{2s^Ts}\right),$$

es decir

$$\lambda = \frac{2}{s^Ts}\left(I - \frac{ss^T}{2s^Ts}\right)(y - Bs).$$

La substitución of λ en (13) da el resultado deseado

$$B_+ = B + \frac{(y - Bs)s^T + s(y - Bs)^T}{s^Ts} - \frac{(y - Bs)^Ts}{(s^Ts)^2}ss^T. \tag{15}$$

Ejercicios. Justifica matemáticamente todos los pasos en la construcción anterior.