CAPITULO XXIV. TECNICAS DE DESCENSO MAS RAPIDO

1. INTRODUCCION Y METODO

La ventaja de los métodos de Newton y de los cuasi-Newton para la solución de sistemas no lineales de ecuaciones es su rapidez de convergencia, una vez que se conoce una aproximación suficientemente precisa. La debilidad de estos métodos es que frecuentemente es necesario tener una aproximación inicial precisa para asegurar la convergencia. El método considerado en este capítulo, conocido como **método del descenso más rápido**, convergerá a la solución generalmente sólo en forma lineal, pero se puede utilizar para encontrar una aproximación inicial suficientemente precisa como para ser usada en las técnicas de Newton, de la misma manera que el método de bisección se usa para una sola ecuación.

El método de descenso más rápido determina un mínimo local para una función multivariada (de valores múltiples) de la forma $G : \mathcal{R}^n \to \mathcal{R}$. La conexión entre la minimización de una función de $\mathcal{R}^n$ a $\mathcal{R}$ y la solución de un sistema de ecuaciones no lineales se debe al hecho de que un sistema de la forma

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= 0 \ , \\ f_2(x_1, x_2, \ldots, x_n) &= 0 \ , \\ \cdots\cdots\cdots \quad & \ldots \\ f_n(x_1, x_2, \ldots, x_n) &= 0 \ , \end{aligned}$$

tiene una solución en $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$ precisamente cuando la función G definida como

$$G(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} [f_i(x_1, x_2, \ldots, x_n)]^2$$

tiene el valor mínimo cero.

Antes de describir el método de descenso más rápido, necesitamos revisar algunos resultados del cálculo.

Definición. Si es $G : \mathcal{R}^n \to \mathcal{R}$, el **gradiente** de G en $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$ se denota por $\nabla G(\mathbf{x})$ y se define como

$$\nabla G(\mathbf{x}) = \Big(\frac{\partial G}{\partial x_1}(\mathbf{x}), \frac{\partial G}{\partial x_2}(\mathbf{x}), \ldots, \frac{\partial G}{\partial x_n}(\mathbf{x}) \Big)^t \ .$$

El gradiente de una función multivariada es análogo a la derivada de una función de una sola variable en el sentido de que una función multivariada diferenciable puede tener un mínimo en $\mathbf{x}$ sólo cuando el gradiente es cero.

El gradiente tiene otra propiedad importante relacionada con la minimización de funciones multivariadas.

Definición. Supongamos que $\mathbf{v} = (v_1, v_2, \ldots, v_n)^t$ es un vector unitario en $\mathcal{R}^n$ ($||\mathbf{v}||_2^2 = 1$). La **derivada direccional** de G en la dirección de $\mathbf{v}$ se define por

$$D_{\mathbf{v}} G(\mathbf{x}) = \lim_{h \to 0} \frac{G(\mathbf{x} + h\ \mathbf{v}) - G(\mathbf{x})}{h} \ .$$

La derivada direccional de G en $\mathbf{x}$ en la dirección $\mathbf{v}$ mide el cambio en el valor de la función G relativo al cambio de la variable en la dirección de $\mathbf{v}$.

Un resultado estándar del cálculo de funciones multivariadas establece que la dirección que produce el valor máximo de la derivada direccional se presenta cuando se escoge $\mathbf{v}$ de manera que sea paralelo a $\nabla G(\mathbf{x})$ siempre y cuando $\nabla G(\mathbf{x}) \neq 0$. Como consecuencia, la dirección en la que el valor de G en $\mathbf{x}$ decrece más es aquélla dada por $-\nabla G(\mathbf{x})$.

Volvamos ahora al método de descenso más rápido.

El método del descenso más rápido para encontrar un mínimo local de una función G de $\mathcal{R}^n$ a $\mathcal{R}$ puede describirse intuitivamente de la manera siguiente:

(i) evaluar G en una aproximación inicial $\mathbf{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})^t$;

(ii) encontrar una dirección a partir de $\mathbf{x}^{(0)}$ que resulte en un descenso en el valor de G;

(iii) decidir cuánto debe moverse en esta dirección y llamar a este nuevo valor $\mathbf{x}^{(1)}$;

(iv) repetir los pasos (i) a (iii) reemplezando a $\mathbf{x}^{(0)}$ por $\mathbf{x}^{(1)}$.

Como la dirección en la que el valor de G en $\mathbf{x}$ decrece más es aquélla dada por $-\nabla G(\mathbf{x})$, entonces, una elección apropiada para $\mathbf{x}$ es

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - \alpha\ \nabla G(\mathbf{x}^{(0)}) \qquad (XXIV.1)$$

para alguna constante $\alpha > 0$.

El problema se reduce ahora a escoger α tal que $G(\mathbf{x}^{(1)})$ sea significativamente menor que $G(\mathbf{x}^{(0)})$. Para encontrar cómo eligir el valor de α de una forma apropiada, consideramos la función de una variable

$$h(\alpha) = G(\mathbf{x}^{(0)} - \alpha\ \nabla G(\mathbf{x}^{(0)})) \ .$$

El valor de α que minimiza a h es el valor que se necesita en la ecuación $(XXIV.1)$.

Para encontrar el valor mínimo de h directamente se necesitaría derivar h y encontrar los puntos críticos; éste es un procedimiento muy costoso generalmente. En su lugar podemos empezar con tres aproximaciones no negativas α_1, α_2 y α_3 a α, el número que produce el valor mínimo de h. Después encontramos el polinomio cuadrático que interpola a h en α_1, α_2 y α_3. Se define α como el número que minimiza a este polinomio cuadrático. Se usa ese valor para determinar la nueva iteración para aproximar el valor mínimo de G

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - \alpha\ \nabla G(\mathbf{x}^{(0)}) \ .$$

2. ALGORITMO Y EJEMPLOS

El siguiente algoritmo usa el método del descenso más rápido para encontrar una aproximación al valor mínimo de $G(\mathbf{x})$. Para emplear este método en la aproximación de la solución al sistema

$$
\begin{aligned}
f_1(x_1, x_2, \ldots, x_n) &= 0 \ , \\
f_2(x_1, x_2, \ldots, x_n) &= 0 \ , \\
\ldots\ldots\ldots \quad & \ \ldots \\
f_n(x_1, x_2, \ldots, x_n) &= 0 \ ,
\end{aligned}
$$

simplemente se reemplaza la función G por $\sum\limits_{i=1}^{n} [f_i(x_1, x_2, \ldots, x_n)]^2$.

Algoritmo del descenso más rápido.

==

Para aproximar una solución $\mathbf{p}$ al problema de minimización $G(\mathbf{p}) = \min\limits_{\mathbf{x}\in\mathcal{R}^n} G(\mathbf{x})$, dada una aproximación inicial $\mathbf{x}$.

Entrada: número n de variables; aproximación inicial $\mathbf{x} = (x_1,\ x_2,\ \ldots,\ x_n)^t$; tolerancia TOL; número máximo de iteraciones N_0;

Salida: solución aproximada $\mathbf{x} = (x_1,\ x_2,\ \ldots,\ x_n)^t$ o mensaje de fracaso.

Paso 1: tomar $k = 0$;

Paso 2: mientras que $k \le N_0$ seguir pasos 3–15;

Paso 3: sean

$$g_1 = G(x_1, x_2, \ldots, x_n) \ ; \qquad (Nota: \quad g_1 = G(\mathbf{x}^{(k)}) \ ;)$$

$$\mathbf{z} = -\nabla G(x_1, x_2, \ldots, x_n) \ ; \qquad (Nota: \quad \mathbf{z} = -\nabla G(\mathbf{x}^{(k)}) \ ;)$$

$$z_0 = ||\mathbf{z}||_2 \ ;$$

Paso 4: si $z_0 = 0$ entonces SALIDA ('GRADIENTE CERO');
SALIDA $(x_1, x_2, \ldots, x_n, g_1)$;
(procedimiento completado, puede tener un mínimo; verificar más aún) PARAR;

Paso 5: tomar

$$\mathbf{z} = \mathbf{z}/z_0 \ ; \qquad (Normalizar \ \ \mathbf{z} \ ;)$$

$$\alpha_1 = 0, \quad \alpha_3 = 1, \quad g_3 = G(\mathbf{x} + \alpha_3 \ \mathbf{z}) \ ;$$

Paso 6: mientras que $|g_3| \ge |g_1|$ seguir pasos 7 y 8;

Paso 7: tomar

$$\alpha_3 = \alpha_3/2, \quad g_3 = G(\mathbf{x} + \alpha_3 \ \mathbf{z}) \ ;$$

Paso 8: si $\alpha_3 < TOL \ / \ 2$ entonces SALIDA ('NO HUBO MEJORA');
SALIDA $(x_1, x_2, \ldots, x_n, g_1)$; *(procedimiento completado, puede tener un mínimo; verificar más aún)* PARAR;

Paso 9: tomar

$$\alpha_2 = \alpha_3/2, \quad g_2 = G(\mathbf{x} + \alpha_2 \ \mathbf{z}) \ ;$$

Paso 10: tomar

$$h_1 = (g_2 - g_1)/(\alpha_2 - \alpha_1) \ ;$$

$$h_2 = (g_3 - g_2)/(\alpha_3 - \alpha_2) \ ;$$

$$h_3 = (h_2 - h_1)/(\alpha_3 - \alpha_1) \ ;$$

$(Nota:$ *el polinomio cuadrático* $P(\alpha) = g_1 + h_1\ \alpha + h_3\ \alpha\ (\alpha - \alpha_2)$ *interpola a* $h(\alpha)$ *en* $\alpha = 0, \alpha = \alpha_2, \alpha = \alpha_3$ $)$

Paso 11: tomar $(Nota:$ *en* α_0 *hay un punto crítico de* $P)$

$$\alpha_0 = 0.5\ (\alpha_2 - h_1/h_3) \ ;$$

$$g_0 = G(\mathbf{x} + \alpha_0\ \mathbf{z}) \ ;$$

Paso 12: encontrar α de $\{\alpha_0, \alpha_1, \alpha_2, \alpha_3\}$ tal que

$$g = G(\mathbf{x} + \alpha\ \mathbf{z}) = \min\{g_0, g_1, g_2, g_3\} \ ;$$

Paso 13: tomar

$$\mathbf{x} = \mathbf{x} + \alpha\ \mathbf{z} \ ;$$

Paso 14: si $|g - g_1| < TOL$ entonces SALIDA $((x_1, x_2, \ldots, x_n, g)$; *(procedimiento completado satisfactoriamente)* PARAR;

Paso 15: tomar $k = k + 1$;

Paso 16: SALIDA $('$*Número máximo* N_0 *de iteraciones excedido,* $N_0 = \ ', N_0)$; PARAR.

==

Ejemplo 1.

Para encontrar una aproximación inicial razonable a la solución del sistema no lineal

$$f_1(x_1, x_2, x_3) = 3\ x_1 - \cos(x_2 x_3) - \frac{1}{2} = 0 \ ,$$
$$f_2(x_1, x_2, x_3) = x_1^2 - 81\ (x_2 + 0.1)^2 + \sin x_3 + 1.06 = 0 \ ,$$
$$f_3(x_1, x_2, x_3) = e^{-x_1 x_2} + 20\ x_3 + \frac{10\pi - 3}{3} = 0 \ ,$$

usamos el algoritmo del descenso más rápido con $TOL = 0.005$, $N_0 = 10$ y $\mathbf{x}^{(0)} = (0.5, 0.5, 0.5)^t$.

Sea $G(x_1, x_2, x_3) = [f_1(x_1, x_2, x_3)]^2 + [f_2(x_1, x_2, x_3)]^2 + [f_3(x_1, x_2, x_3)]^2$; entonces

$$\begin{aligned} \nabla G(x_1, x_2, x_3) = \nabla G(\mathbf{x}) = \\ &= \Big(2\ f_1(\mathbf{x})\ \frac{\partial f_1}{\partial x_1}(\mathbf{x}) + 2\ f_2(\mathbf{x})\ \frac{\partial f_2}{\partial x_1}(\mathbf{x}) + 2\ f_3(\mathbf{x})\ \frac{\partial f_3}{\partial x_1}(\mathbf{x}), \\ &\quad 2\ f_1(\mathbf{x})\ \frac{\partial f_1}{\partial x_2}(\mathbf{x}) + 2\ f_2(\mathbf{x})\ \frac{\partial f_2}{\partial x_2}(\mathbf{x}) + 2\ f_3(\mathbf{x})\ \frac{\partial f_3}{\partial x_2}(\mathbf{x}), \\ &\quad 2\ f_1(\mathbf{x})\ \frac{\partial f_1}{\partial x_3}(\mathbf{x}) + 2\ f_2(\mathbf{x})\ \frac{\partial f_2}{\partial x_3}(\mathbf{x}) + 2\ f_3(\mathbf{x})\ \frac{\partial f_3}{\partial x_3}(\mathbf{x})\Big)^t \\ &= 2\ J(\mathbf{x})^t\ \mathbf{F}(\mathbf{x}) \ . \end{aligned}$$

Para $\mathbf{x}^{(0)} = (0.5, 0.5, 0.5)^t$,

$$G(\mathbf{x}^{(0)}) = 1159.24\ ,$$

$$\mathbf{z} = (-0.0131206, 0.9897569, 0.1421648)^t\ ,$$

$$z_0 = ||\mathbf{z}||_2 = 5359.964\ ,$$

$$\alpha_1 = 0.00, \quad g_1 = 1159.24,$$

$$\alpha_2 = 0.25, \quad g_2 = 454.8059, \quad h_1 = -2817.738,$$

$$\alpha_3 = 0.50, \quad g_3 = 363.5173, \quad h_2 = -365.1542, \quad h_3 = 4905.167,$$

así que $P(\alpha) = 1159.24 - 2817.173\ \alpha + 4905.167\ \alpha\ (\alpha - 0.25)$. Por lo que,

$$\alpha_0 = 0.4122214, \quad g_0 = 372.2808.$$

Entonces $\alpha = 0.5$ y

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - 0.5\ \mathbf{z} = (0.5065602, 0.005121529, 0.4289176)^t\ ,$$

$$G(\mathbf{x}^{(1)}) = 363.5173\ .$$

La siguiente tabla contiene los restantes resultados.

Tabla 1

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$G(x_1^{(k)}, x_2^{(k)}, x_3^{(k)})$
0	0.5	0.5	0.5	1159.240
1	0.5065602	0.005121529	0.4289176	363.5173
2	0.5045557	0.06421983	−0.5155808	1.874157
3	0.5068673	0.001808132	−0.5179668	0.01322149
4	0.5067566	0.001208410	−0.5235992	0.0005774418

En el ejemplo 2 del capítulo XXVIII se dio como solución al sistema no lineal $(0.5, 0, -0.5235988)^t$, así que los resultados obtenidos aquí pueden servir como una aproximación inicial adecuada para los métodos de Newton y de Broyden.

EJERCICIOS.

1. Usar el método del descenso más rápido con $TOL = 0.005$ para aproximar las soluciones a los sistemas no lineales dados en los dos capítulos anteriores.

2. Usar el método de descenso más rápido para aproximar los mínimos de las funciones siguientes con una precisión de 0.005:

$$
\begin{aligned}
(a)\quad & G(x_1, x_2) = \cos(x_1 + x_2) + \sin x_1 + \cos x_2 \ , \\
(b)\quad & G(x_1, x_2) = 100\ (x_1^2 - x_2)^2 + (1 - x_1)^2 \ , \\
(c)\quad & G(x_1, x_2, x_3) = x_1^2 + 2\ x_2^2 + x_3^2 - 2\ x_1\ x_2 + 2\ x_1 - 2.5\ x_2 - x_3 + 2 \ .
\end{aligned}
$$