Optimización Numérica sin restricciones

Tema 4: Gradiente Conjugado

Oscar Dalmau

Centro de Investigación en Matemáticas
CIMAT

20 de abril de 2021

Orden del Tema

1 Gradiente Conjugado Lineal
- Direcciones Conjugadas
- Solución mediante Direcciones Conjugadas
- Algoritmo Direcciones Conjugadas Iterativo
- Gradiente Conjugado-Versión Preliminar
- Forma práctica de GC

Problema de optimización

Se quiere resolver, mediante un algoritmo iterativo, el problema de optimización sin restricciones

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} \tag{1}$$

donde $\mathbf{Q}$ es una matriz simétrica y definida positiva.

Note que es equivalente a resolver el sistema de ecuaciones

$$\mathbf{Q}\boldsymbol{x} = \boldsymbol{b} \tag{2}$$

$$\boldsymbol{x}^* = \mathbf{Q}^{-1}\boldsymbol{b} \tag{3}$$

Ejemplo en 2D

Para ello, consideremos el caso particular en $\mathbb{R}^2$

$$\min_{(x,y)\in\mathbb{R}^2} f(x,y) = \frac{1}{2}\lambda_1(x-a)^2 + \frac{1}{2}\lambda_2(y-b)^2 \tag{4}$$

con $\lambda_1, \lambda_2 > 0$.

Note que en este caso:

$$f(x,y) = f_1(x) + f_2(y) \tag{5}$$

$$\nabla f(x,y) = [f_1'(x), f_2'(y)]^T \tag{6}$$

Además $(x,y)^* = (a,b)$

Ejemplo

Sea dado (x_k, y_k) y la actualización en la dirección de máximo descenso,

$$\begin{aligned} x_{k+1} &= x_k - \alpha_x f_1'(x_k) && (7) \\ y_{k+1} &= y_k - \alpha_y f_2'(y_k) && (8) \end{aligned}$$

donde α_x y α_y solo los tamaños de paso.

Ejemplo

Supongamos que $f_1'(x_k) \neq 0$ y $f_2'(y_k) \neq 0$. Si además, α_x y α_y son tamaños de paso exactos en cada dirección (se usa descenso coordenado)

$$\alpha_x = \frac{x_k - a}{f_1'(x_k)} \tag{9}$$

$$x_{k+1} = x_k - \alpha_x f_1'(x) = x_k - \frac{x_k - a}{f_1'(x_k)} f_1'(x_k) = a \tag{10}$$

similarmente $y_{k+1} = b$. Luego $(x_{k+1}, y_{k+1}) = (x, y)^* = (a, b)$, es decir, para cualquier punto (x_k, y_k) se llega al óptimo en una iteración!!

Observaciones

- El ejemplo anterior es una forma cuadrática orientada en los ejes
- Si se aplica máximo descenso coordenado con tamaño de paso exacto, se llega al óptimo, en una iteración.
- Notar que el problema es separable en cada variable!, ie

$$f(x,y) = \frac{1}{2}[x,y]\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} - [a,b]\begin{bmatrix} x \\ y \end{bmatrix} + cte$$

es decir, la matriz es diagonal

Gradiente Conjugado

Se podrán usar las ideas anteriores al caso general?

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} \tag{11}$$

Gradiente Conjugado

Basados en el ejemplo en $\mathbb{R}^2$, la idea sería diagonalizar $\mathbf{Q}$

$$f(\boldsymbol{y}) = \frac{1}{2}\boldsymbol{y}^T \boldsymbol{\Lambda} \boldsymbol{y} - \tilde{\boldsymbol{b}}^T \boldsymbol{y} \tag{12}$$

donde $\mathbf{Q} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$, $\boldsymbol{y} = \mathbf{U}^T \boldsymbol{x}$ y $\tilde{\boldsymbol{b}} = \boldsymbol{b}\mathbf{U}$, donde $\boldsymbol{\Lambda}$ es diagonal y $\mathbf{U}$ es ortogonal.

Luego, la solución es muy facil, ie $\boldsymbol{y} = \boldsymbol{\Lambda}^{-1}\tilde{\boldsymbol{b}}$ y $\boldsymbol{x} = \mathbf{U}\boldsymbol{y}$.

El problema es que habría que descomponer la matriz $\mathbf{Q}$, lo cual es computacionalmente costoso, se puede hacer algo mejor?

Orden del Tema

1. Gradiente Conjugado Lineal
   - Direcciones Conjugadas
   - Solución mediante Direcciones Conjugadas
   - Algoritmo Direcciones Conjugadas Iterativo
   - Gradiente Conjugado-Versión Preliminar
   - Forma práctica de GC

Direcciones Conjugadas

Sea $\mathbf{Q}$ una matriz simétrica y definida positiva. Dos vectores $\boldsymbol{d}_1, \boldsymbol{d}_2$ se dicen *conjugados con respecto a* $\mathbf{Q}$ o simplemente $\mathbf{Q}$-ortogonales si $\boldsymbol{d}_1^T \mathbf{Q} \boldsymbol{d}_2 = 0$.

Un conjunto de vectores $\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_k$ son mutuamente $\mathbf{Q}$-ortogonales si $\boldsymbol{d}_i^T \mathbf{Q} \boldsymbol{d}_j = 0$ para $i \neq j$.

Proposición Direcciones Conjugadas

Proposición: Sea $\mathbf{Q}$ una matriz simétrica definida positiva. Si el conjunto de vectores $\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_k$ son mutuamente $\mathbf{Q}$-ortogonales entonces son linealmente independientes.

Demostración

Supongamos que existen números reales α_i, $i = 0, 1, \cdots, k$ para los que se cumple

$$\sum_{j=0}^{k} \alpha_j \boldsymbol{d}_j = \mathbf{0}, \tag{13}$$

Premultiplicando por $\boldsymbol{d}_i^T \mathbf{Q}$ (con $i = 0, 1, \cdots, k$)

$$\sum_{j=0}^{k} \alpha_j \boldsymbol{d}_i^T \mathbf{Q} \boldsymbol{d}_j = \boldsymbol{d}_i^T \mathbf{Q} \mathbf{0}, \tag{14}$$

como para $i \neq j$ se cumple $\boldsymbol{d}_i^T \mathbf{Q} \boldsymbol{d}_j = 0$ entonces

$$\alpha_i \boldsymbol{d}_i^T \mathbf{Q} \boldsymbol{d}_i = 0, \tag{15}$$

y por tanto $\alpha_i = 0$ para $i = 0, 1, \cdots, k$. Con lo cual se concluye la demostración.

Orden del Tema

1. Gradiente Conjugado Lineal
   - Direcciones Conjugadas
   - Solución mediante Direcciones Conjugadas
   - Algoritmo Direcciones Conjugadas Iterativo
   - Gradiente Conjugado-Versión Preliminar
   - Forma práctica de GC

Solución mediante Direcciones Conjugadas

- Sea $\mathcal{D} = \{\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_{n-1}\}$ un conjunto de direcciones conjugadas (*previamente conocidas o dadas*) con respecto a una matriz simétrica y definida positiva $\mathbf{Q}$.
- De acuerdo a la proposición anterior el conjunto $\mathcal{D}$ es linealmente independiente por lo tanto $\mathcal{D}$ es una base de $\mathbb{R}^n$.
- Supongamos adicionalmente que $\boldsymbol{x}^*$ es la solución del problema de optimización.
- Luego, podemos expresar $\boldsymbol{x}^*$, de forma única, como una combinación lineal usando la base $\mathcal{D}$

Algoritmo Básico de las Direcciones Conjugadas

$$\boldsymbol{x}^* = \sum_{j=0}^{n-1} \alpha_j \boldsymbol{d}_j \tag{16}$$

Premultiplicando por $\boldsymbol{d}_i^T \mathbf{Q}$, usando (2) y por el hecho de que $\boldsymbol{d}_i^T \mathbf{Q} \boldsymbol{d}_j = 0$ para $i \neq j$ (por ser direcciones conjugadas), obtenemos los coeficientes de la combinación lineal:

$$\boldsymbol{d}_i^T \mathbf{Q} \boldsymbol{x}^* = \sum_{j=0}^{n-1} \alpha_j \boldsymbol{d}_i^T \mathbf{Q} \boldsymbol{d}_j \tag{17}$$

$$\alpha_i \boldsymbol{d}_i^T \mathbf{Q} \boldsymbol{d}_i = \boldsymbol{d}_i^T \mathbf{Q} \boldsymbol{x}^* \tag{18}$$

$$\alpha_i = \frac{\boldsymbol{d}_i^T \boldsymbol{b}}{\boldsymbol{d}_i^T \mathbf{Q} \boldsymbol{d}_i} \tag{19}$$

Solución mediante Direcciones Conjugadas

Finalmente,

$$\boldsymbol{x}^* = \sum_{j=0}^{n-1} \frac{\boldsymbol{d}_j^T \boldsymbol{b}}{\boldsymbol{d}_j^T \mathbf{Q} \boldsymbol{d}_j} \boldsymbol{d}_j \tag{20}$$

Notar que

- El uso de direcciones $\mathbf{Q}$-conjugadas permite calcular los coeficientes, debido a que al premultiplicar adecuadamente por una de las direcciones, todos los términos de la derecha se anulan salvo un término.
- Al mismo tiempo, los coeficientes de la combinación lineal se pudieron expresar en términos de información conocida, ie, las direcciones conjugadas, $\mathbf{Q}$ y $\boldsymbol{b}$.

Orden del Tema

1. Gradiente Conjugado Lineal
   - Direcciones Conjugadas
   - Solución mediante Direcciones Conjugadas
   - Algoritmo Direcciones Conjugadas Iterativo
   - Gradiente Conjugado-Versión Preliminar
   - Forma práctica de GC

Sea dado un conjunto de direcciones conjugadas $\mathcal{D}$.
Definamos $\boldsymbol{g}_k \overset{def}{=} \nabla f(\boldsymbol{x}_k) = \mathbf{Q}\boldsymbol{x}_k - \boldsymbol{b}$.

Dado un punto inicial $\boldsymbol{x}_0$, generemos la secuencia

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k \tag{21}$$

donde $\alpha_k = -\frac{\boldsymbol{g}_k^T \boldsymbol{d}_k}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k}$ es el minimizador de $f(\cdot)$ a lo largo de la recta $\boldsymbol{x}_k + \alpha \boldsymbol{d}_k$

$$\alpha_k = \arg\min_{\alpha} f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) \tag{22}$$

De donde se obtiene

$$\begin{aligned}
\nabla^T f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)\boldsymbol{d}_k &= 0 \qquad (23)\\
(\mathbf{Q}(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) - \boldsymbol{b})^T \boldsymbol{d}_k &= 0 \qquad (24)\\
(\boldsymbol{g}_k + \alpha \mathbf{Q}\boldsymbol{d}_k)^T \boldsymbol{d}_k &= 0 \qquad (25)\\
\alpha_k &= -\frac{\boldsymbol{g}_k^T \boldsymbol{d}_k}{\boldsymbol{d}_k^T \mathbf{Q}\boldsymbol{d}_k} \qquad (26)
\end{aligned}$$

Luego $\boldsymbol{g}_{k+1}^T \boldsymbol{d}_k = 0$

Algoritmo Básico de las Direcciones Conjugadas

Require: $\boldsymbol{x}_0, \mathcal{D} = \{\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_{n-1}\}$
Ensure: $\boldsymbol{x}^*$

Hacer $k = 0$, $\boldsymbol{g}_k = \mathbf{Q}\boldsymbol{x}_k - \boldsymbol{b}$
while $k < n$ y $\|\boldsymbol{g}_k\| \neq 0$ (No conveja) **do**
  $\alpha_k = -\frac{\boldsymbol{g}_k^T \boldsymbol{d}_k}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k}$
  $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$
  $\boldsymbol{g}_{k+1} = \mathbf{Q}\boldsymbol{x}_{k+1} - \boldsymbol{b}$
  $k = k + 1$
end while

Conjugate Direction Theorem

Para cualquier $\boldsymbol{x}_0 \in \mathbb{R}^n$, la sucesión $\{\boldsymbol{x}_k\}$ generada usando el algoritmo anterior converge a la solución $\boldsymbol{x}^*$ en a lo sumo n pasos.

Demostración

Como $\mathcal{D}$ es un conjunto de vectores linealmente independiente, entonces podemos escribir

$$\boldsymbol{x}^* - \boldsymbol{x}_0 = \sum_{j=0}^{n-1} \sigma_j \boldsymbol{d}_j \tag{27}$$

de forma única para ciertos σ_j.

Demostración

A partir de

$$\boldsymbol{x}^* - \boldsymbol{x}_0 = \sum_{j=0}^{n-1} \sigma_j \boldsymbol{d}_j \tag{28}$$

Premultiplicando por $\boldsymbol{d}_k^T \mathbf{Q}$, y por el hecho de que $\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_j = 0$ para $k \neq j$,

$$\boldsymbol{d}_k^T \mathbf{Q}(\boldsymbol{x}^* - \boldsymbol{x}_0) = \sum_{j=0}^{n-1} \sigma_j \boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_j \tag{29}$$

$$\sigma_k = \frac{\boldsymbol{d}_k^T \mathbf{Q}(\boldsymbol{x}^* - \boldsymbol{x}_0)}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k} \tag{30}$$

Demostración

$$
\begin{aligned}
\sigma_k &= \frac{\boldsymbol{d}_k^T \mathbf{Q}(\boldsymbol{x}^* - \boldsymbol{x}_0)}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k} \\
&= \frac{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{x}^* - \boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{x}_0}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k} \\
&= \frac{\boldsymbol{d}_k^T (\boldsymbol{b} - \mathbf{Q} \boldsymbol{x}_k)}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k}, \text{pues } \mathbf{Q}\boldsymbol{x}^* = \boldsymbol{b} \text{ y } \boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{x}_k = \boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{x}_0 \\
&= -\frac{\boldsymbol{g}_k^T \boldsymbol{d}_k}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k}, \text{pues } \boldsymbol{g}_k = \mathbf{Q} \boldsymbol{x}_k - \boldsymbol{b}
\end{aligned}
$$

Luego $\sigma_k = \alpha_k$ con lo que se concluye la demostración.

Demostración: Comentario

A partir de $\sigma_k = \alpha_k$, definiendo $\mathbf{D} := [\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_{n-1}]$

$$\begin{aligned} \boldsymbol{x}^* &= \boldsymbol{x}_0 + \sum_{k=0}^{n-1} \sigma_k \boldsymbol{d}_k = \boldsymbol{x}_0 + \mathbf{D}\boldsymbol{\sigma} \\ \boldsymbol{x}_n &= \boldsymbol{x}_0 + \sum_{k=0}^{n-1} \alpha_k \boldsymbol{d}_k = \boldsymbol{x}_0 + \mathbf{D}\boldsymbol{\alpha} \end{aligned}$$

se tiene $\boldsymbol{x}_n = \boldsymbol{x}^*$

Orden del Tema

1. Gradiente Conjugado Lineal
   - Direcciones Conjugadas
   - Solución mediante Direcciones Conjugadas
   - Algoritmo Direcciones Conjugadas Iterativo
   - Gradiente Conjugado-Versión Preliminar
   - Forma práctica de GC

Algoritmo Gradiente Conjugado

La idea del Algoritmo Gradiente Conjugado se basa en el Algoritmo de las direcciones conjugadas, pero sin conocer las direcciones conjugadas a priori.

Se comienza selecciónando la primera dirección $\boldsymbol{d}_0 = -\boldsymbol{g}_0$ y el resto

$$\boldsymbol{d}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_{k+1}\boldsymbol{d}_k$$

donde β_{k+1} se selecciόna de modo que $\boldsymbol{d}_k$ y $\boldsymbol{d}_{k+1}$ sean $\mathbf{Q}$-conjugados.

Algoritmo Gradiente Conjugado

Para calcular β_{k+1} basta premultiplicar por $\boldsymbol{d}_k^T\mathbf{Q}$ en la igualdad $\boldsymbol{d}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_{k+1}\boldsymbol{d}_k$

$$\begin{aligned}
\boldsymbol{d}_{k+1} &= -\boldsymbol{g}_{k+1} + \beta_{k+1}\boldsymbol{d}_k \\
\boldsymbol{d}_k^T\mathbf{Q}\boldsymbol{d}_{k+1} &= -\boldsymbol{d}_k^T\mathbf{Q}\boldsymbol{g}_{k+1} + \beta_{k+1}\boldsymbol{d}_k^T\mathbf{Q}\boldsymbol{d}_k \\
-\boldsymbol{d}_k^T\mathbf{Q}\boldsymbol{g}_{k+1} + \beta_{k+1}\boldsymbol{d}_k^T\mathbf{Q}\boldsymbol{d}_k &= 0 \\
\beta_{k+1} &= \frac{\boldsymbol{g}_{k+1}^T\mathbf{Q}\boldsymbol{d}_k}{\boldsymbol{d}_k^T\mathbf{Q}\boldsymbol{d}_k}
\end{aligned}$$

Ver algoritmo siguiente (versión preliminar)

Algoritmo Gradiente Conjugado

Algorithm 1 GC-Versión Preliminar

Require: $\boldsymbol{x}_0$

Ensure: x^*

1: Hacer $\boldsymbol{g}_0 = \mathbf{Q}\boldsymbol{x}_0 - \boldsymbol{b}$,$\boldsymbol{d}_0 = -\boldsymbol{g}_0$, $k = 0$

2: **while** $\|\boldsymbol{g}_k\| \neq 0$ (No conveja) **do**

3: $\alpha_k = -\dfrac{\boldsymbol{g}_k^T \boldsymbol{d}_k}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k}$

4: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$

5: $\boldsymbol{g}_{k+1} = \mathbf{Q}\boldsymbol{x}_{k+1} - \boldsymbol{b} = \nabla f(\boldsymbol{x}_{k+1})$

6: $\beta_{k+1} = \dfrac{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{g}_{k+1}}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k}$

7: $\boldsymbol{d}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_{k+1} \boldsymbol{d}_k$

8: $k = k + 1$

9: **end while**

Orden del Tema

1. Gradiente Conjugado Lineal
   - Direcciones Conjugadas
   - Solución mediante Direcciones Conjugadas
   - Algoritmo Direcciones Conjugadas Iterativo
   - Gradiente Conjugado-Versión Preliminar
   - Forma práctica de GC

Forma práctica de GC

Algunas relaciones

- $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$
- $\boldsymbol{g}_{k+1} = \mathbf{Q}\boldsymbol{x}_{k+1} - \boldsymbol{b}$
- $\boldsymbol{g}_{k+1} = \boldsymbol{g}_k + \alpha_k \mathbf{Q}\boldsymbol{d}_k$
- $\mathbf{Q}\boldsymbol{x}_{k+1} = \mathbf{Q}\boldsymbol{x}_k + \alpha_k \mathbf{Q}\boldsymbol{d}_k$
- $\mathbf{Q}\boldsymbol{x}_{k+1} - \boldsymbol{b} = \mathbf{Q}\boldsymbol{x}_k - \boldsymbol{b} + \alpha_k \mathbf{Q}\boldsymbol{d}_k$
- $\boldsymbol{g}_{k+1} = \boldsymbol{g}_k + \alpha_k \mathbf{Q}\boldsymbol{d}_k$

Forma práctica de GC

Algunas relaciones

- $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$
- $\boldsymbol{g}_{k+1} = \mathbf{Q}\boldsymbol{x}_{k+1} - \boldsymbol{b}$
- $\boldsymbol{g}_{k+1} = \boldsymbol{g}_k + \alpha_k \mathbf{Q}\boldsymbol{d}_k$
- $\mathbf{Q}\boldsymbol{x}_{k+1} = \mathbf{Q}\boldsymbol{x}_k + \alpha_k \mathbf{Q}\boldsymbol{d}_k$
- $\mathbf{Q}\boldsymbol{x}_{k+1} - \boldsymbol{b} = \mathbf{Q}\boldsymbol{x}_k - \boldsymbol{b} + \alpha_k \mathbf{Q}\boldsymbol{d}_k$
- $\boldsymbol{g}_{k+1} = \boldsymbol{g}_k + \alpha_k \mathbf{Q}\boldsymbol{d}_k$

Forma práctica de GC

Algunas relaciones

- $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$
- $\boldsymbol{g}_{k+1} = \mathbf{Q}\boldsymbol{x}_{k+1} - \boldsymbol{b}$
- $\boldsymbol{g}_{k+1} = \boldsymbol{g}_k + \alpha_k \mathbf{Q}\boldsymbol{d}_k$
- $\mathbf{Q}\boldsymbol{x}_{k+1} = \mathbf{Q}\boldsymbol{x}_k + \alpha_k \mathbf{Q}\boldsymbol{d}_k$
- $\mathbf{Q}\boldsymbol{x}_{k+1} - \boldsymbol{b} = \mathbf{Q}\boldsymbol{x}_k - \boldsymbol{b} + \alpha_k \mathbf{Q}\boldsymbol{d}_k$
- $\boldsymbol{g}_{k+1} = \boldsymbol{g}_k + \alpha_k \mathbf{Q}\boldsymbol{d}_k$

Forma práctica de GC

Algunas relaciones

- $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$
- $\boldsymbol{g}_{k+1} = \mathbf{Q}\boldsymbol{x}_{k+1} - \boldsymbol{b}$
- $\boldsymbol{g}_{k+1} = \boldsymbol{g}_k + \alpha_k \mathbf{Q}\boldsymbol{d}_k$
- $\mathbf{Q}\boldsymbol{x}_{k+1} = \mathbf{Q}\boldsymbol{x}_k + \alpha_k \mathbf{Q}\boldsymbol{d}_k$
- $\mathbf{Q}\boldsymbol{x}_{k+1} - \boldsymbol{b} = \mathbf{Q}\boldsymbol{x}_k - \boldsymbol{b} + \alpha_k \mathbf{Q}\boldsymbol{d}_k$
- $\boldsymbol{g}_{k+1} = \boldsymbol{g}_k + \alpha_k \mathbf{Q}\boldsymbol{d}_k$

Forma práctica de GC: Algunas relaciones

Otras relaciones

- $\boldsymbol{x}_k - \boldsymbol{x}_i \perp_{\mathbf{Q}} \boldsymbol{d}_j,\ j \in \{0, 1, \cdots, i-1\}$, ie, $\boldsymbol{x}_k^T \mathbf{Q} \boldsymbol{d}_j = \boldsymbol{x}_i^T \mathbf{Q} \boldsymbol{d}_j$
- $\boldsymbol{x}_k - \boldsymbol{x}_i \perp_{\mathbf{Q}} \boldsymbol{d}_k$, $0 \leq i \leq k$, ie, $\boldsymbol{x}_k^T \mathbf{Q} \boldsymbol{d}_k = \boldsymbol{x}_i^T \mathbf{Q} \boldsymbol{d}_k$
- $\boldsymbol{g}_k - \boldsymbol{g}_i \perp \boldsymbol{d}_j$, $j \in \{0, 1, \cdots, i-1\}$, i.e., $\boldsymbol{g}_k^T \boldsymbol{d}_j = \boldsymbol{g}_i^T \boldsymbol{d}_j$
- $\boldsymbol{g}_k - \boldsymbol{g}_i \perp \boldsymbol{d}_k$, $0 \leq i \leq k$, i.e., $\boldsymbol{g}_k^T \boldsymbol{d}_k = \boldsymbol{g}_i^T \boldsymbol{d}_k$

Luego

- $\boldsymbol{x}_k^T \mathbf{Q} \boldsymbol{d}_k = \boldsymbol{x}_0^T \mathbf{Q} \boldsymbol{d}_k$
- $\boldsymbol{g}_k^T \boldsymbol{d}_k = \boldsymbol{g}_0^T \boldsymbol{d}_k$

Forma práctica de GC: Algunas relaciones

$$\begin{aligned} \boldsymbol{x}_k &= \boldsymbol{x}_{k-1} + \alpha_{k-1}\boldsymbol{d}_{k-1} = \boldsymbol{x}_{k-2} + \alpha_{k-2}\boldsymbol{d}_{k-2} + \alpha_{k-1}\boldsymbol{d}_{k-1} \\ &= \boldsymbol{x}_i + \sum_{s=i}^{k-1} \alpha_s \boldsymbol{d}_s \end{aligned}$$

Luego

$$\begin{aligned} \boldsymbol{x}_k - \boldsymbol{x}_i &\perp_{\mathbf{Q}} \boldsymbol{d}_j,\ i.e.,\ \boldsymbol{x}_k^T \mathbf{Q} \boldsymbol{d}_j = \boldsymbol{x}_i^T \mathbf{Q} \boldsymbol{d}_j \\ j &\in \{0, 1, \cdots, i-1\} \\ \boldsymbol{x}_k - \boldsymbol{x}_i &\perp_{\mathbf{Q}} \boldsymbol{d}_k,\ i.e.,\ \boldsymbol{x}_k^T \mathbf{Q} \boldsymbol{d}_k = \boldsymbol{x}_i^T \mathbf{Q} \boldsymbol{d}_k \end{aligned}$$

Forma práctica de GC:Algunas relaciones

$$\begin{aligned} \boldsymbol{g}_k &= \boldsymbol{g}_{k-1} + \alpha_{k-1}\mathbf{Q}\boldsymbol{d}_{k-1} = \boldsymbol{g}_{k-2} + \alpha_{k-2}\mathbf{Q}\boldsymbol{d}_{k-2} + \alpha_{k-1}\mathbf{Q}\boldsymbol{d}_{k-1} \\ &= \boldsymbol{g}_i + \sum_{s=i}^{k-1} \alpha_s\mathbf{Q}\boldsymbol{d}_s \end{aligned}$$

Luego

$$\begin{aligned} \boldsymbol{g}_k - \boldsymbol{g}_i &\perp \boldsymbol{d}_j,\ i.e.,\ \boldsymbol{g}_k^T\boldsymbol{d}_j = \boldsymbol{g}_i^T\boldsymbol{d}_j \\ j &\in \{0, 1, \cdots, i-1\} \\ \boldsymbol{g}_k - \boldsymbol{g}_i &\perp \boldsymbol{d}_k,\ i.e.,\ \boldsymbol{g}_k^T\boldsymbol{d}_k = \boldsymbol{g}_i^T\boldsymbol{d}_k \end{aligned}$$

Forma práctica de GC

Otras relaciones

- $\boldsymbol{g}_{k+1}^T \boldsymbol{d}_k = 0$
- $\alpha_k = -\frac{\boldsymbol{g}_k^T \boldsymbol{d}_k}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k}$
- $\beta_{k+1} = \frac{\boldsymbol{g}_{k+1}^T \mathbf{Q} \boldsymbol{d}_k}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k}$
- $\boldsymbol{d}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_{k+1} \boldsymbol{d}_k$

Forma práctica de GC

Proposicion

Si $\{\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_k\}$ son $\mathbf{Q}$-ortogonales entonces,

$$\begin{aligned} \boldsymbol{g}_{k+1}^T \boldsymbol{d}_i &= 0 \\ \boldsymbol{g}_{k+1}^T \boldsymbol{g}_i &= 0 \end{aligned}$$

para $i = 0, 1, \cdots, k$

Forma práctica de GC

Prueba de la Primera relación

Si $i = k$ es claro que se cumple

$$\boldsymbol{g}_{k+1}^T \boldsymbol{d}_i = 0$$

pues si α_k es el tamaño de paso exacto

$$0 = \phi'(\alpha_k) = \nabla f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k)^T \boldsymbol{d}_k = \boldsymbol{g}_{k+1}^T \boldsymbol{d}_k$$

Forma práctica de GC

Prueba de la Primera relación

Si $i < k$ y usando la $\mathbf{Q}$-ortogonalidad, la relación de la lamina anterior y la relacion

$$\boldsymbol{g}_{k+1} = \boldsymbol{g}_{i+1} + \sum_{s=i+1}^{k} \alpha_s \mathbf{Q} \boldsymbol{d}_s$$

Se concluye que

$$\boldsymbol{g}_{k+1}^T \boldsymbol{d}_i = 0$$

Forma práctica de GC

Prueba de la Segunda relación

Para ello, podemos usar la relacion para $0 \leq i \leq k$

$$\begin{aligned} \boldsymbol{d}_i &= -\boldsymbol{g}_i + \beta_i \boldsymbol{d}_{i-1} \\ \boldsymbol{g}_i &= -\boldsymbol{d}_i + \beta_i \boldsymbol{d}_{i-1} \end{aligned}$$

y con la primera parte de la proposición

$$\boldsymbol{g}_{k+1}^T \boldsymbol{g}_i = \boldsymbol{g}_{k+1}^T (-\boldsymbol{d}_i + \beta_i \boldsymbol{d}_{i-1}) = 0$$

Recalculando el tamaño de paso α_k

- $\alpha_k = -\frac{\boldsymbol{g}_k^T \boldsymbol{d}_k}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k}$
- $\boldsymbol{g}_k^T \boldsymbol{d}_i = 0$, $i = 0, 1, \cdots, k-1$
- $\boldsymbol{d}_k = -\boldsymbol{g}_k + \beta_k \boldsymbol{d}_{k-1}$

Luego

- $\alpha_k = \frac{\boldsymbol{g}_k^T \boldsymbol{g}_k}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k}$

Recalculando el β_{k+1}

- $\beta_{k+1} = \frac{\boldsymbol{g}_{k+1}^T \mathbf{Q} \boldsymbol{d}_k}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k}$
- $\boldsymbol{g}_k^T \boldsymbol{d}_i = 0,\, i = 0, 1, \cdots, k-1$
- $\boldsymbol{d}_k = -\boldsymbol{g}_k + \beta_k \boldsymbol{d}_{k-1}$
- $\alpha_k \mathbf{Q} \boldsymbol{d}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$

Luego

- $\beta_{k+1} = \frac{\boldsymbol{g}_{k+1}^T \boldsymbol{g}_{k+1}}{\boldsymbol{g}_k^T \boldsymbol{g}_k}$

Algoritmo Gradiente Conjugado

Algorithm 2 GC (Estandar)

Require: $\boldsymbol{x}_0$

Ensure: x^*

1: Hacer $\boldsymbol{g}_0 = \mathbf{Q}\boldsymbol{x}_0 - \boldsymbol{b}$,$\boldsymbol{d}_0 = -\boldsymbol{g}_0$, $k = 0$

2: **while** $\|\boldsymbol{g}_k\| \neq 0$ (No conveja) **do**

3: $\quad \alpha_k = \frac{\boldsymbol{g}_k^T \boldsymbol{g}_k}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k} = -\frac{\boldsymbol{g}_k^T \boldsymbol{d}_k}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k}$

4: $\quad \boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$

5: $\quad \boldsymbol{g}_{k+1} = \mathbf{Q}\boldsymbol{x}_{k+1} - \boldsymbol{b} = \nabla f(\boldsymbol{x}_{k+1})$

6: $\quad \beta_{k+1} = \frac{\boldsymbol{g}_{k+1}^T \boldsymbol{g}_{k+1}}{\boldsymbol{g}_k^T \boldsymbol{g}_k} = \frac{\boldsymbol{g}_{k+1}^T \mathbf{Q} \boldsymbol{d}_k}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k}$

7: $\quad \boldsymbol{d}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_{k+1} \boldsymbol{d}_k$

8: $\quad k = k + 1$

9: **end while**

Expanding Subspace Theorem

Sea $\boldsymbol{x}_0 \in \mathbb{R}^n$ el punto inicial de la sucesión $\{\boldsymbol{x}_k\}$ generada usando GC. Entonces $\boldsymbol{x}_k$ es el minimizador de $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \mathbf{Q}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$ en el conjunto:

$$\mathcal{H}_k = \{\boldsymbol{x} | \boldsymbol{x} = \boldsymbol{x}_0 + \text{gen}\{\mathcal{D}_k\}\} \tag{31}$$

donde $\text{gen}\{\mathcal{D}_k\}$ representa el subespacio generado por $\mathcal{D}_k = \{\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_{k-1}\}$, i.e., el conjunto de todas las combinaciones lineales de los vectores $\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_{k-1}$.

Expanding Subspace Theorem

La idea se basa en mostrar que $\boldsymbol{x}(\hat{\boldsymbol{\alpha}}) = \boldsymbol{x}(\boldsymbol{\alpha}^*)$ donde $\boldsymbol{x}(\hat{\boldsymbol{\alpha}})$ es solución obtenida usando GC y $\boldsymbol{x}(\boldsymbol{\alpha}^*)$ es el óptimo en $\mathcal{H}_k$, es decir

$$\boldsymbol{x}(\hat{\boldsymbol{\alpha}}) = \boldsymbol{x}_0 + \sum_{i=0}^{k-1} \hat{\alpha}_i \boldsymbol{d}_i; \; \hat{\alpha}_i = -\frac{\boldsymbol{d}_i^T \boldsymbol{g}_i}{\boldsymbol{d}_i^T \mathbf{Q} \boldsymbol{d}_i} \tag{32}$$

$$\boldsymbol{x}(\boldsymbol{\alpha}^*) = \boldsymbol{x}_0 + \sum_{i=0}^{k-1} \alpha_i^* \boldsymbol{d}_i \tag{33}$$

es decir basta probar que $\hat{\alpha}_i = \alpha_i^*$

Expanding Subspace Theorem

Sea $\boldsymbol{x}(\boldsymbol{\alpha}) \in \mathcal{H}_k$ entonces, si $\mathbf{D} = [\boldsymbol{d}_0, \cdots, \boldsymbol{d}_{k-1}]$

$$\boldsymbol{x}(\boldsymbol{\alpha}) = \boldsymbol{x}_0 + \sum_{i=0}^{k-1} \alpha_i \boldsymbol{d}_i = \boldsymbol{x}_0 + \mathbf{D}\boldsymbol{\alpha}$$

entonces es suficiente encontrar los coeficientes

$$\boldsymbol{\alpha}^* = \arg\min_{\boldsymbol{\alpha}} f(\boldsymbol{x}(\boldsymbol{\alpha}))$$

Expanding Subspace Theorem

$$
\begin{aligned}
f(\boldsymbol{x}(\boldsymbol{\alpha})) &= \frac{1}{2}[\boldsymbol{x}_0 + \mathbf{D}\boldsymbol{\alpha}]^T \mathbf{Q}[\boldsymbol{x}_0 + \mathbf{D}\boldsymbol{\alpha}] - \boldsymbol{b}^T[\boldsymbol{x}_0 + \mathbf{D}\boldsymbol{\alpha}] \\
&= \frac{1}{2}\boldsymbol{\alpha}^T \mathbf{D}^T \mathbf{Q}\mathbf{D}\boldsymbol{\alpha} + \boldsymbol{x}_0^T \mathbf{Q}\mathbf{D}\boldsymbol{\alpha} - \boldsymbol{b}^T \mathbf{D}\boldsymbol{\alpha} \\
&\quad + \frac{1}{2}\boldsymbol{x}_0^T \mathbf{Q}\boldsymbol{x}_0 - \boldsymbol{b}^T \boldsymbol{x}_0 \\
&= \frac{1}{2}\boldsymbol{\alpha}^T \mathbf{D}^T \mathbf{Q}\mathbf{D}\boldsymbol{\alpha} + (\mathbf{Q}\boldsymbol{x}_0 - \boldsymbol{b})^T \mathbf{D}\boldsymbol{\alpha} + f(\boldsymbol{x}_0) \\
\nabla_\alpha f(\boldsymbol{x}(\boldsymbol{\alpha})) &= \mathbf{D}^T \mathbf{Q}\mathbf{D}\boldsymbol{\alpha} + \mathbf{D}^T[\mathbf{Q}\boldsymbol{x}_0 - \boldsymbol{b}] \\
&= \mathbf{D}^T \mathbf{Q}\mathbf{D}\boldsymbol{\alpha} + \mathbf{D}^T \boldsymbol{g}_0
\end{aligned}
$$

Expanding Subspace Theorem

$$\nabla_{\boldsymbol{\alpha}} f(\boldsymbol{x}(\boldsymbol{\alpha})) = \mathbf{D}^T \mathbf{Q} \mathbf{D} \boldsymbol{\alpha} + \mathbf{D}^T \boldsymbol{g}_0 = 0$$

Luego

$$\boldsymbol{\alpha}^* = -(\mathbf{D}^T \mathbf{Q} \mathbf{D})^{-1} \mathbf{D}^T \boldsymbol{g}_0 \equiv \hat{\boldsymbol{\alpha}}$$

Pues $\mathbf{D}^T\mathbf{Q}\mathbf{D}$ es una matriz diagonal (por la $\mathbf{Q}$-ortogonalidad) con entradas $\boldsymbol{d}_i^T \mathbf{Q} \boldsymbol{d}_i$ y las entradas de $\mathbf{D}^T \boldsymbol{g}_0$ son $\boldsymbol{d}_i^T \boldsymbol{g}_0 \equiv \boldsymbol{d}_i^T \boldsymbol{g}_i$ (Ver próxima lámina), es decir, $\alpha_i^* = \hat{\alpha}_i$

Expanding Subspace Theorem

$$\mathbf{D}^T\mathbf{Q}\mathbf{D} = \begin{bmatrix} \boldsymbol{d}_0^T \\ \boldsymbol{d}_1^T \\ \cdot \\ \boldsymbol{d}_{k-1}^T \end{bmatrix} \mathbf{Q}\left[\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots \boldsymbol{d}_{k-1}\right] = [\boldsymbol{d}_i^T\mathbf{Q}\boldsymbol{d}_j]_{i,j=1,\cdots,k-1}$$

donde $\boldsymbol{d}_i^T\mathbf{Q}\boldsymbol{d}_j = 0$ para $i \neq j$

$$\mathbf{D}^T\boldsymbol{g}_0 = [\boldsymbol{d}_i^T\boldsymbol{g}_0]_{i=1,\cdots,k-1} = [\boldsymbol{d}_i^T\boldsymbol{g}_i]_{i=1,\cdots,k-1}$$

pues $\boldsymbol{d}_i^T\boldsymbol{g}_i = \boldsymbol{d}_i\boldsymbol{g}_0$ ó $\boldsymbol{d}_i^T(\boldsymbol{g}_i - \boldsymbol{g}_0) = 0$ ó $\boldsymbol{g}_i - \boldsymbol{g}_0 \perp \boldsymbol{d}_i$

Teorema

Suponga que hasta la iteración k ésima el Algoritmo Gradiente Conjugado no ha encontrado la solución óptima $\boldsymbol{x}^*$. Entonces se cumplen las siguientes propiedades

1. $$\begin{aligned} \text{gen}\{\boldsymbol{g}_0, \boldsymbol{g}_1, \cdots, \boldsymbol{g}_k\} &= \text{gen}\{\boldsymbol{g}_0, \mathbf{Q}\boldsymbol{g}_0, \cdots, \mathbf{Q}^k\boldsymbol{g}_0\} \\ \text{gen}\{\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_k\} &= \text{gen}\{\boldsymbol{g}_0, \mathbf{Q}\boldsymbol{g}_0, \cdots, \mathbf{Q}^k\boldsymbol{g}_0\} \end{aligned}$$

2. Las direcciones $\{\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_k\}$ del Algoritmo Gradiente Conjugado son $\mathbf{Q}$-ortogonales, ie, para $i = 0, 1, \cdots, k$

$$\boldsymbol{d}_{k+1}^T \mathbf{Q} \boldsymbol{d}_i = 0$$

Prueba Teorema: Por inducción

Parte 1:

$$\begin{aligned} \text{gen}\{\boldsymbol{g}_0, \boldsymbol{g}_1, \cdots, \boldsymbol{g}_k\} &= \text{gen}\{\boldsymbol{g}_0, \mathrm{Q}\boldsymbol{g}_0, \cdots, \mathrm{Q}^k\boldsymbol{g}_0\} \\ \text{gen}\{\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_k\} &= \text{gen}\{\boldsymbol{g}_0, \mathrm{Q}\boldsymbol{g}_0, \cdots, \mathrm{Q}^k\boldsymbol{g}_0\} \end{aligned}$$

Inicio $i = 0$:

$$\begin{aligned} \text{gen}\{\boldsymbol{g}_0\} &= \text{gen}\{\boldsymbol{g}_0\} \\ \text{gen}\{\boldsymbol{d}_0\} &= \text{gen}\{\boldsymbol{g}_0\} \end{aligned}$$

es trivial pues $\boldsymbol{d}_0 = -\boldsymbol{g}_0$

Prueba Teorema: Por inducción

- $\text{gen}\{\boldsymbol{a}, \boldsymbol{b}\} \subset \text{gen}\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$
- $\text{gen}\{\boldsymbol{a}, \boldsymbol{b}\} = \text{gen}\{\boldsymbol{a}, \alpha\boldsymbol{a} + \beta\boldsymbol{b}\}$, $\beta \neq 0$
- Sea $\mathcal{A}$ un espacio lineal, $\text{gen}\{\boldsymbol{a}, \boldsymbol{b}\} \subset \mathcal{A}$ y $\boldsymbol{c} \in \mathcal{A}$ entonces $\text{gen}\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\} \subset \mathcal{A}$

Prueba Teorema: Por inducción

Parte 1: Definamos

$$\begin{aligned}\mathcal{G}_k &= \{\boldsymbol{g}_0, \boldsymbol{g}_1, \cdots, \boldsymbol{g}_k\}\\ \mathcal{D}_k &= \{\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_k\}\\ \mathcal{H}_k &= \{\boldsymbol{g}_0, \mathbf{Q}\boldsymbol{g}_0, \cdots, \mathbf{Q}^k\boldsymbol{g}_0\}\end{aligned}$$

Hipotesis: Si $\begin{aligned}\text{gen}\{\mathcal{G}_k\} &= \text{gen}\{\mathcal{H}_k\}\\ \text{gen}\{\mathcal{D}_k\} &= \text{gen}\{\mathcal{H}_k\}\end{aligned}$ Entonces

$$\begin{aligned}\text{gen}\{\mathcal{G}_{k+1}\} &= \text{gen}\{\mathcal{H}_{k+1}\}\\ \text{gen}\{\mathcal{D}_{k+1}\} &= \text{gen}\{\mathcal{H}_{k+1}\}\end{aligned}$$

Prueba Teorema: Por inducción

Parte 1: Para probar que gen$\{\mathcal{G}_{k+1}\}$ = gen$\{\mathcal{H}_{k+1}\}$ se puede comprobar que gen$\{\mathcal{G}_{k+1}\} \subset$ gen$\{\mathcal{H}_{k+1}\}$ y gen$\{\mathcal{H}_{k+1}\} \subset$ gen$\{\mathcal{G}_{k+1}\}$

Caso gen$\{\mathcal{G}_{k+1}\} \subset$ gen$\{\mathcal{H}_{k+1}\}$:

$$\begin{aligned}
\text{gen}\{\mathcal{G}_{k+1}\} &= \text{gen}\{\mathcal{G}_k, \boldsymbol{g}_{k+1}\} \\
\text{gen}\{\mathcal{G}_k\} &= \text{gen}\{\mathcal{H}_k\} \subset \text{gen}\{\mathcal{H}_{k+1}\} \\
\boldsymbol{g}_{k+1} &= \boldsymbol{g}_k + \alpha_k \mathbf{Q}\boldsymbol{d}_k
\end{aligned}$$

entonces gen$\{\mathcal{G}_{k+1}\}$ = gen$\{\mathcal{G}_k, \mathbf{Q}\boldsymbol{d}_k\}$ pues $\boldsymbol{g}_k \in \mathcal{G}_k$.

Prueba Teorema: Por inducción

Caso gen$\{\mathcal{G}_{k+1}\} \subset$ gen$\{\mathcal{H}_{k+1}\}$: $\boldsymbol{d}_k \in$ gen$\{\mathcal{D}_k\} =$ gen$\{\mathcal{H}_k\}$ entonces $\boldsymbol{d}_k = \sum_{i=0}^{k} \sigma_i \boldsymbol{h}_i$, $\boldsymbol{h}_i = \mathbf{Q}^i \boldsymbol{g}_0$. Luego

$$\mathbf{Q}\boldsymbol{d}_k = \sum_{i=0}^{k} \sigma_i \mathbf{Q}\boldsymbol{h}_i = \sum_{i=0}^{k} \sigma_i \mathbf{Q}^{i+1} \boldsymbol{g}_0 = \sum_{i=0}^{k+1} \sigma_{i-1} \mathbf{Q}^i \boldsymbol{g}_0$$

con $\sigma_{-1} = 0$. Luego $\mathbf{Q}\boldsymbol{d}_k \in$ gen$\{\mathcal{H}_{k+1}\}$ y como gen$\{\mathcal{G}_k\} \subset$ gen$\{\mathcal{H}_{k+1}\}$ entonces gen$\{\mathcal{G}_{k+1}\} =$ gen$\{\mathcal{G}_k, \mathbf{Q}\boldsymbol{d}_k\} \subset$ gen$\{\mathcal{H}_{k+1}\}$.

Prueba Teorema: Por inducción

Caso gen$\{\mathcal{H}_{k+1}\} \subset$ gen$\{\mathcal{G}_{k+1}\}$:

$$\begin{aligned} \text{gen}\{\mathcal{H}_{k+1}\} &= \text{gen}\{\mathcal{H}_k, \mathbf{Q}^{k+1}\boldsymbol{g}_0\} \\ \text{gen}\{\mathcal{H}_k\} &= \text{gen}\{\mathcal{G}_k\} \subset \text{gen}\{\mathcal{G}_{k+1}\} \end{aligned}$$

$$\mathbf{Q}^k\boldsymbol{g}_0 \in \text{gen}\{\mathcal{H}_k\} = \text{gen}\{\mathcal{D}_k\}, \text{ por hipotesis}$$

Luego $\mathbf{Q}^k\boldsymbol{g}_0 = \sum_i \sigma_i \boldsymbol{d}_i$ y $\mathbf{Q}^{k+1}\boldsymbol{g}_0 = \sum_i \sigma_i \mathbf{Q}\boldsymbol{d}_i$ por lo que

$$\begin{aligned} \mathbf{Q}\mathbf{Q}^k\boldsymbol{g}_0 = \mathbf{Q}^{k+1}\boldsymbol{g}_0 &\in \text{gen}\{\mathbf{Q}\mathcal{D}_k\} \\ &= \text{gen}\{\mathbf{Q}\boldsymbol{d}_0, \mathbf{Q}\boldsymbol{d}_1, \cdots, \mathbf{Q}\boldsymbol{d}_k\} \end{aligned}$$

Prueba Teorema: Por inducción

Caso gen$\{\mathcal{H}_{k+1}\} \subset$ gen$\{\mathcal{G}_{k+1}\}$: Como

$$\boldsymbol{g}_{i+1} = \boldsymbol{g}_i + \alpha_i \mathbf{Q} \boldsymbol{d}_i$$

para $i = 0, 1, \cdots, k$ entonces

$$\mathbf{Q}\boldsymbol{d}_i = \frac{1}{\alpha_i}\left(\boldsymbol{g}_{i+1} - \boldsymbol{g}_i\right) \in \text{gen}\{\mathcal{G}_{k+1}\}$$

Por lo que

$$\mathbf{Q}^{k+1}\boldsymbol{g}_0 \in \text{gen}\{\mathbf{Q}\boldsymbol{d}_0, \mathbf{Q}\boldsymbol{d}_1, \cdots, \mathbf{Q}\boldsymbol{d}_k\} \subset \text{gen}\{\mathcal{G}_{k+1}\}$$

y $\mathbf{Q}^{k+1}\boldsymbol{g}_0 \in$ gen$\{\mathcal{G}_{k+1}\}$

Prueba Teorema: Por inducción

Caso gen$\{\mathcal{H}_{k+1}\} \subset$ gen$\{\mathcal{G}_{k+1}\}$: Finalmente, como gen$\{\mathcal{H}_k\} \subset$ gen$\{\mathcal{G}_{k+1}\}$ y $\mathbf{Q}^{k+1}\boldsymbol{g}_0 \in$ gen$\{\mathcal{G}_{k+1}\}$ entonces

$$\text{gen}\{\mathcal{H}_{k+1}\} = \text{gen}\{\mathcal{H}_k, \mathbf{Q}^{k+1}\boldsymbol{g}_0\} \subset \text{gen}\{\mathcal{G}_{k+1}\}$$

Se concluye que gen$\{\mathcal{G}_{k+1}\} =$ gen$\{\mathcal{H}_{k+1}\}$

Prueba Teorema: Por inducción

Parte 1: Probemos que gen$\{\mathcal{D}_{k+1}\}$ = gen$\{\mathcal{H}_{k+1}\}$

$$\text{gen}\{\mathcal{D}_{k+1}\} = \text{gen}\{\mathcal{D}_k, \boldsymbol{d}_{k+1}\}$$

$\boldsymbol{d}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_{k+1}\boldsymbol{d}_k$ y como $\boldsymbol{d}_k \in$ gen$\{\mathcal{D}_k\}$ entonces

$$\begin{aligned}
\text{gen}\{\mathcal{D}_{k+1}\} &= \text{gen}\{\mathcal{D}_k, \boldsymbol{g}_{k+1}\} \\
&= \text{gen}\{\mathcal{G}_k, \boldsymbol{g}_{k+1}\} \text{ hipotesis} \\
&= \text{gen}\{\mathcal{G}_{k+1}\} \\
&= \text{gen}\{\mathcal{H}_{k+1}\} \text{ primera parte}
\end{aligned}$$

Prueba Teorema: Por inducción

Parte 2: $\boldsymbol{d}_{k+1}^T \mathbf{Q} \boldsymbol{d}_i = 0$

Inicio $i = 1$: $\boldsymbol{d}_1^T \mathbf{Q} \boldsymbol{d}_0 = 0$, Es trivial, pues es la condición de gradiente Conjugado, ie, que dos direcciones consecutivas son $\mathbf{Q}$-conjugadas

Hipotesis: Si $\mathcal{D}_k = \{\boldsymbol{d}_0, \cdots, \boldsymbol{d}_k\}$ son $\mathbf{Q}$-ortogonales entonces $\mathcal{D}_{k+1} = \{\boldsymbol{d}_0, \cdots, \boldsymbol{d}_k, \boldsymbol{d}_{k+1}\}$ tambien lo es

Prueba Teorema: Hipotesis

Como $\mathcal{D}_{k+1} = \{\mathcal{D}_k, \boldsymbol{d}_{k+1}\}$ y por hipotesis $\mathcal{D}_k$ es $\mathbf{Q}$-ortogonal, es suficiente mostrar que $\boldsymbol{d}_{k+1} \perp_{\mathbf{Q}} \mathcal{D}_k$

Caso $i = k$: $\boldsymbol{d}_{k+1}^T \mathbf{Q} \boldsymbol{d}_k = 0$, Es trivial, pues es la condición de gradiente Conjugado, ie, que dos direcciones consecutivas sean $\mathbf{Q}$-conjugadas

Caso $i < k$: $\boldsymbol{d}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_{k+1} \boldsymbol{d}_k$ y como $\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_i = 0$, pues $\boldsymbol{d}_k, \boldsymbol{d}_i \in \mathcal{D}_k$, entonces

$$
\begin{aligned}
\boldsymbol{d}_{k+1}^T \mathbf{Q} \boldsymbol{d}_i &= (-\boldsymbol{g}_{k+1} + \beta_{k+1} \boldsymbol{d}_k)^T \mathbf{Q} \boldsymbol{d}_i \\
&= -\boldsymbol{g}_{k+1}^T \mathbf{Q} \boldsymbol{d}_i + \beta_{k+1} \boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_i \\
&= -\boldsymbol{g}_{k+1}^T \mathbf{Q} \boldsymbol{d}_i
\end{aligned}
$$

Prueba Teorema: Hipotesis

Como $\mathcal{D}_{k+1} = \{\mathcal{D}_k, \boldsymbol{d}_{k+1}\}$ es suficiente mostrar que $\boldsymbol{d}_{k+1} \perp_{\mathbf{Q}} \mathcal{D}_k$

Caso $i < k$: Como $\mathbf{Q}\boldsymbol{d}_i = \frac{1}{\alpha_i}(\boldsymbol{g}_{i+1} - \boldsymbol{g}_i)$ y $\boldsymbol{g}_{k+1}^T \boldsymbol{g}_j = 0$ para $j < k+1$

$$\begin{aligned} \boldsymbol{d}_{k+1}^T \mathbf{Q} \boldsymbol{d}_i &= -\boldsymbol{g}_{k+1}^T \mathbf{Q} \boldsymbol{d}_i \\ &= -\frac{1}{\alpha_i} \boldsymbol{g}_{k+1}^T (\boldsymbol{g}_{i+1} - \boldsymbol{g}_i) = 0 \end{aligned}$$

y se concluye que $\mathcal{D}_{k+1}$ es $\mathbf{Q}$-ortogonal.