# Deducción de las fórmulas del método del gradiente conjugado

**Objetivos.** Demostrar el teorema sobre los subespacios de Krylov en el método del gradiente conjugado.

**Requisitos.** Subespacios generados por listas de vectores, definición axiomática de producto interno, proceso de ortogonalización de Gram–Schmidt, búsqueda exacta sobre una recta.

## Búsqueda exacta sobre una recta (repaso)

**1. Fórmulas para la búsqueda exacta sobre una recta.** Sean $A \in \mathcal{M}_n(\mathbb{R})$ una matriz cuadrada positiva definida, $b \in \mathbb{R}^n$,

$$f(y) \coloneqq \frac{1}{2}\langle y, Ay\rangle - \langle b, y\rangle.$$

Sean $x, p \in \mathbb{R}^n$ dos vectores fijos, $p \neq \mathbf{0}_n$. Minimizando la función

$$g(\lambda) = f(x + \lambda p)$$

y usando la notación $r = b - Ax$, $r_{\text{nuevo}} = b - Ax_{\text{nuevo}}$, obtenemos:

$$\lambda = \frac{p^\top r}{p^\top A\, p}, \qquad x_{\text{nuevo}} = x + \lambda p, \qquad r_{\text{nuevo}} = r - \lambda A p, \qquad p^\top r_{\text{nuevo}} = 0.$$

## Subespacios generados por listas de vectores (repaso)

**2. Definición (subespacio generado por una lista de vectores).** Sean $a_1, \ldots, a_m \in \mathbb{R}^n$. Denotemos por $\ell(a_1, \ldots, a_m)$ al conjunto de todas las combinaciones lineales de los vectores $a_1, \ldots, a_m$:

$$\ell(a_1, \ldots, a_m) \coloneqq \left\{ x \in \mathbb{R}^n \colon \quad \exists \lambda_1, \ldots, \lambda_m \in \mathbb{R} \quad x = \sum_{i=1}^{m} \lambda_i a_i \right\}.$$

Se sabe que $\ell(a_1, \ldots, a_m)$ es un subespacio de $\mathbb{R}^n$. Además el subespacio más pequeño que contiene a los vectores $a_1, \ldots, a_m$.

**3. Proposición (criterio de la contención de subespacios generados por listas de vectores).** Sean $a_1, \ldots, a_k, b_1, \ldots, b_m \in \mathbb{R}^n$. Entonces

$$\ell(a_1, \ldots, a_k) \subseteq \ell(b_1, \ldots, b_m) \qquad \Longleftrightarrow \qquad a_1, \ldots, a_k \in \ell(b_1, \ldots, b_m).$$

**4. Corolario.** Sean $a_1, \ldots, a_k, b_1, \ldots, b_k \in \mathbb{R}^n$ tales que

$$\ell(a_1, \ldots, a_{k-1}) = \ell(b_1, \ldots, b_{k-1}), \qquad a_k \in \ell(b_1, \ldots, b_k), \qquad b_k \in \ell(a_1, \ldots, a_k).$$

Entonces $\ell(a_1, \ldots, a_k) = \ell(b_1, \ldots, b_k)$.

## Vector ortogonal al subespacio generado por una lista de vectores (repaso)

**5. Definición (vector ortogonal a un subespacio).** Sea $S$ un subespacio de $\mathbb{R}^n$ y sea $x \in \mathbb{R}^n$. Se dice que $x$ es *ortogonal* a $S$ y se escribe $x \perp S$ si $x$ es ortogonal a todo elemento de $S$:

$$x \perp S \qquad \stackrel{\text{def}}{\Longleftrightarrow} \qquad \forall y \in S \qquad x \perp S.$$

**6.** Sean $a_1, \ldots, a_k \in \mathbb{R}^n$ y $S := \ell(a_1, \ldots, a_k)$. Entonces para todo $x \in \mathbb{R}^n$ las siguientes condiciones son equivalentes:

$$\forall i \in \{1, \ldots, k\} \quad x \perp a_k \qquad \Longleftrightarrow \qquad x \perp S.$$

## Un paso de la ortogonalización de Gram–Schmidt (repaso)

**7.** Sean $a_1, \ldots, a_k \in \mathbb{R}^n$ vectores no nulos y ortogonales por pares, y sea $b \in \mathbb{R}^n$. Definimos

$$c := b - \sum_{i=1}^{k} \frac{\langle a_i, b \rangle}{\langle a_i, a_i \rangle} a_i.$$

Entonces $c \perp \ell(a_1, \ldots, a_k)$.

## Producto interno asociado a una matriz real simétrica positiva definida (repaso)

**8. Producto interno asociado a la matriz $\boldsymbol{A}$.** Sea $A$ una matriz real simétrica positiva definida: $A^\top = A$, $A > 0$. Entonces la función definida mediante la siguiente fórmula es un producto interno en $\mathbb{R}^n$:

$$\langle v, w \rangle_A := \langle v, Aw \rangle = v^\top A w.$$

La norma inducida por este producto interno se denota por $\| \cdot \|_A$:

$$\|v\|_A := \langle v, v \rangle_A = \langle v, Av \rangle = v^\top A v.$$

Dos vectores $v, w \in \mathbb{R}^n$ se llaman $A$-*conjugados* (o $A$-*ortogonales*) si son ortogonales con respecto a este producto interno:

$$v \perp_A w \qquad \Longleftrightarrow \qquad \langle v, w \rangle_A = 0.$$

Notemos que muchos de los conceptos anteriores, incluso la ortogonalidad de un vector a un subespacio y el proceso de la ortogonalización de Gram–Schmidt, se pueden generalizar al caso del producto interno asociado a una matriz simétrica positiva definida.

## Subespacios de Krylov

**9. Definición (subespacio de Krylov asociado a una matriz y un vector).** Sean $A \in \mathcal{M}_n(\mathbb{R})$ y $v \in \mathbb{R}^n$.

$$K_j(A, v) := \ell\big(v, Av, A^2 v, \ldots, A^{j-1} v\big).$$

**10. Lema (propiedad de los subespacios de Krylov).** Sean $A \in \mathcal{M}_n(\mathbb{R})$, $v \in \mathbb{R}^n$ y $x \in K_j(A, v)$. Entonces $Ax \in K_{j+1}(A, v)$.

## Teoremas sobre los subespacios y la ortogonalidad en el método del gradiente conjugado

**11. Idea del método del gradiente conjugado.** Es un caso particular del método de direcciones conjugadas, donde cada vector $p^{(s+1)}$ se obtiene al restar del vector $r^{(s+1)}$ su proyección ortogonal, con respecto al producto interno $\langle \cdot, \cdot \rangle_A$, al subespacio generado por los vectores $p^{(j)}$, $k \leq s$:

$$p^{(s+1)} := r^{(s+1)} - \sum_{j=0}^{s} \beta_{s+1,j} p^{(j)}.$$

Es la misma fórmula que en el método de Gram–Schmidt, y esta fórmula garantiza que $p^{(s+1)} \perp_A p^{(j)}$ para $j \leq s$.

**12. Lema (sobre el coeficiente $\alpha$ en el método de gradiente conjugado).**

$$\alpha_s = \frac{\|r^{(s)}\|^2}{\|p^{(s)}\|_A^2}.$$

*Demostración.* Recordamos que

$$p^{(s)} := r^{(s)} - \sum_{j=0}^{s-1} \beta_{s,j} p^{(j)}$$

y obtenemos

$$\langle p^{(s)}, r^{(s)} \rangle = \|r^{(s)}\|^2 - \sum_{j=0}^{s-1} \beta_{s,j} \langle p^{(j)}, r^{(s)} \rangle,$$

pero los últimos productos internos son cero por la propiedad de residuos y direcciones en métodos de direcciones conjugadas. □

**13. Teorema (sobre la igualdad de subespacios en el método de gradiente conjugado).** Sea $A \in \mathcal{M}_n(\mathbb{R})$ una matriz simétrica definida positiva y sea $p^{(0)} = r^{(0)} \in \mathbb{R}^n$ un vector no nulo. Definimos $r^{(1)}, r^{(2)}, \ldots$ y $p^{(1)}, p^{(2)}, \ldots$ mediante las siguientes fórmulas (mientras $\alpha_s \neq 0$):

$$\alpha_s \coloneqq \frac{\langle p^{(s)}, r^{(s)} \rangle}{\langle p^{(s)}, p^{(s)} \rangle_A}, \qquad \beta_{s+1,j} \coloneqq \frac{\langle p^{(j)}, r^{(s+1)} \rangle_A}{\langle p^{(j)}, p^{(j)} \rangle_A},$$

$$r^{(s+1)} \coloneqq r^{(s)} - \alpha_s A p^{(s)}, \qquad p^{(s+1)} \coloneqq r^{(s+1)} - \sum_{j=0}^{s} \beta_{s+1,j} p^{(j)}.$$

Entonces para cada $s$ tenemos:

$$\ell(p^{(0)}, \ldots, p^{(s)}) = \ell(r^{(0)}, \ldots, r^{(s)}) = K_{s+1}(A, r^{(0)}). \tag{1}$$

*Demostración.* Primero se demuestra la igualdad $\ell(p^{(0)}, \ldots, p^{(s)}) = \ell(r^{(0)}, \ldots, r^{(s)})$ por inducción sobre $s$, usando las siguientes ideas:

$$p^{(s+1)} = r^{(s+1)} - \sum_{j=0}^{s} \beta_{s+1,j} p^{(j)} \overset{\text{hip. ind.}}{\in} \ell(r^{(0)}, \ldots, r^{(s+1)}),$$

$$r^{(s+1)} = p^{(s+1)} + \sum_{j=0}^{s} \beta_{s+1,j} p^{(j)} \in \ell(p^{(0)}, \ldots, p^{(s+1)}).$$

Luego se demuestra por inducción sobre $s$ que el subespacio $K_{s+1}(A, r^{(0)})$ coincide con el subespacio $\ell(p^{(0)}, \ldots, p^{(s)})$. Se usa la igualdad ya demostrada y las siguientes ideas:

$$A p^{(s)} = \frac{r^{(s)} - r^{(s+1)}}{\alpha_s} \in \ell(r^{(0)}, \ldots, r^{(s+1)}), \tag{2}$$

$$r^{(s+1)} = \underbrace{r^{(s)}}_{\substack{\cap \\ \in K_{s+1}(A, r^{(0)})}} - \alpha_s A \underbrace{p^{(s)}}_{\substack{\cap \\ K_{s+1}(A, r^{(0)})}} \in K_{s+2}(A, r^{(0)}),$$

$$A \ell(p^{(0)}, \ldots, p^{(s)}) = \ell(A p^{(0)}, \ldots, A p^{(s)}) \overset{(2)}{\subseteq} \ell(r^{(0)}, \ldots, r^{(s+1)}),$$

y finalmente

$$A^{s+1} r^{(0)} = A(A^s r^{(0)}) \overset{\text{hip. ind.}}{\in} A \ell(p^{(0)}, \ldots, p^{(s)}) \subseteq \ell(r^{(0)}, \ldots, r^{(s+1)}). \qquad \square$$

**14. Corolario.** Para cada $j$,

$$p^{(j)} \perp_A K_j(A, r^{(0)}), \tag{3}$$

y

$$r^{(j)} \perp K_j(A, r^{(0)}). \tag{4}$$

Además los vectores $r^{(j)}$ son ortogonales por pares.

*Demostración.* El teorema dice que $K_j(A, r^{(0)}) = \ell(p^{(0)}, \ldots, p^{(j-1)})$. Por lo tanto, la conclusión (3) es la $A$-ortogonalidad de los vectores $p^{(j)}$, y la conclusión (4) es el hecho que ya vimos para el método de direcciones conjugadas: cada residuo es ortogonal a las direcciones anteriores. La última afirmación se obtiene de la igualdad $\ell(r^{(0)}, \ldots, r^{(j-1)}) = K_{j-1}(A, r^{(0)})$ y de (4). $\square$

**15. Corolario.** Las fórmulas para calcular los coeficientes $\beta_{s+1,j}$ se simplifican de la siguiente manera:

$$\beta_{s+1,0} = \ldots = \beta_{s+1,j} = 0, \qquad \beta_{s+1,s} = -\frac{\|r^{(s+1)}\|^2}{\|r^{(s)}\|^2}.$$

*Demostración.* Para cualquier $j \in \{0, \ldots, s-1\}$ tenemos

$$\langle p^{(j)}, r^{(s+1)} \rangle_A = \langle Ap^{(j)}, r^{(s+1)} \rangle = \frac{1}{\alpha_j} \langle r^{(j)} - r^{(j+1)}, r^{(s+1)} \rangle = 0.$$

Esto significa que

$$p^{(s+1)} = r^{(s+1)} - \beta_{s+1,s} p^{(s)}.$$

Ahora calculamos el denominador de $\beta_{s+1,s}$:

$$\|p^{(s)}\|_A^2 = \langle p^{(s)}, Ap^{(s)} \rangle = \frac{1}{\alpha_s} \langle p^{(s)}, r^{(s)} - r^{(s+1)} \rangle = \frac{1}{\alpha_s} \langle p^{(s)}, r^{(s)} \rangle = \frac{1}{\alpha_s} \|r^{(s)}\|^2.$$

El numerador de $\beta_{s+1,s}$:

$$\langle p^{(s)}, r^{(s+1)} \rangle_A = \langle Ap^{(s)}, r^{(s+1)} \rangle = \frac{1}{\alpha_s} \langle r^{(s)} - r^{(s+1)}, r^{(s+1)} \rangle = -\frac{1}{\alpha_s} \|r^{(s+1)}\|^2. \qquad \square$$

# Algoritmo (método del gradiente conjugado)

**Entrada:**

$A$ una matriz real simétrica positiva definida, denotamos su orden por $n$;

$b$ un vector real de longitud $n$;

$x0$ una aproximación inicial a la solución incógnita del sistema $Ax = b$;

$\varepsilon$ un número positivo que determina la "tolerancia" del residuo $b - Ax$;

$s_{\max}$ un número entero positivo, el número máximo permitido de iteraciones.

**Salida:**

$x$ es una aproximación a la solución del sistema $Ax = b$;

$s$ es el número de las iteraciones hechas.

La salida de la función satisface la siguiente condición:

$$s = s_{\max} \qquad \vee \qquad \left|b - Ax\right| < \varepsilon.$$

**Código en el lenguaje MATLAB:**

```
function [x, s] = conjugate_gradient_method(A, b, x0, tol, smax)
   x = x0;
   r = b - A * x;
   p = r;
   rr = r' * r;
   s = 0;
   tol2 = tol * tol;
   while (s < smax) && (rr >= tol2),
      q = A * p;
      la = (p' * r) / (p' * q);
      x = x + la * p;
      r = r - la * q;
      rrold = rr;
      rr = r' * r;
      be = - rr / rrold;
      p = r - be * p;
      s = s + 1;
   end
end
```