Idea del método de gradiente conjugado (un tema del curso "Álgebra Lineal Numérica")

Egor Maximenko

Instituto Politécnico Nacional
Escuela Superior de Física y Matemáticas
México

18 de junio de 2021

Plan

Objetivo:

- entender la idea del método de gradiente conjugado;
- demostrar fórmulas eficientes para los vectores y escalares que aparecen en este método.

Prerrequisitos:

- métodos iterativos para resolver sistemas $Ax = b$ con $A > 0$;
- la búsqueda exacta sobre una recta;
- el método de direcciones conjugadas;
- la ortogonalización de Gram–Schmidt.

Plan

Del sistema $Ax = b$ a un problema de minimización

Proposición

Sea $A \in \mathcal{M}_n(\mathbb{R})$ tal que $A > 0$, y sea $b \in \mathbb{R}^n$. Entonces el punto

$$s := A^{-1}b$$

es el mínimo global estricto de la función $f : \mathbb{R}^n \to \mathbb{R}$,

$$f(x) := \frac{1}{2}x^\top Ax - b^\top x.$$

La minimización de f sobre una recta

Proposición

Sean A, b, f como en la proposición anterior, sean $y \in \mathbb{R}^n$, $p \in \mathbb{R}^n \setminus \{0_n\}$.
Definimos $g\colon \mathbb{R} \to \mathbb{R}$,

$$g(\alpha) := f(y + \alpha p).$$

Entonces g alcanza su mínimo global estricto en el punto

$$\alpha_{\min} = \frac{p^\top r}{p^\top A\, p},$$

donde

$$r := b - Ay.$$

El cambio del residuo al cambiar el vector

Proposición

Sean $A \in \mathcal{M}_n(\mathbb{R})$, $y \in \mathbb{R}^n$, $p \in \mathbb{R}^n \setminus \{0\}$, $\alpha \in \mathbb{R}$,

$$x := y + \alpha p.$$

Entonces

$$b - Ax = (b - Ay) - \alpha Ap.$$

Idea de métodos iterativos para $A > 0$

$$p_s := ???,$$

$$q_s := A p_s,$$

$$\alpha_s := \frac{p_s^\top r_s}{p_s^\top q_s},$$

$$x_{s+1} := x_s + \alpha_s p_s,$$

$$r_{s+1} := r_s - \alpha_s q_s.$$

```
while rr > tol * tol,
    p = ???;
    q = A * p;
    al = (p' * r) / (p' * q);
    x = x + al * p;
    r = r - al * q;
    rr = r' * r;
endwhile
```

La ortogonalidad del residuo a la dirección anterior, repaso

Proposición

En las suposiciones anteriores, para cada s,

$$r_{s+1} \perp p_s.$$

La ortogonalidad del residuo a la dirección anterior, repaso

Proposición

En las suposiciones anteriores, para cada s,

$$r_{s+1} \perp p_s.$$

Demostración:

$$p_s^\top r_{s+1} =$$

La ortogonalidad del residuo a la dirección anterior, repaso

Proposición

En las suposiciones anteriores, para cada s,

$$r_{s+1} \perp p_s.$$

Demostración:

$$p_s^\top r_{s+1} = p_s^\top (r_s - \alpha_s q_s) =$$

La ortogonalidad del residuo a la dirección anterior, repaso

Proposición

En las suposiciones anteriores, para cada s,

$$r_{s+1} \perp p_s.$$

Demostración:

$$p_s^\top r_{s+1} = p_s^\top (r_s - \alpha_s q_s) = p_s^\top r_s - \alpha_s p_s^\top q_s =$$

La ortogonalidad del residuo a la dirección anterior, repaso

Proposición

En las suposiciones anteriores, para cada s,

$$r_{s+1} \perp p_s.$$

Demostración:

$$p_s^\top r_{s+1} = p_s^\top (r_s - \alpha_s q_s) = p_s^\top r_s - \alpha_s p_s^\top q_s = p_s^\top r_s - \frac{p_s^\top r_s}{p_s^\top q_s} p_s^\top q_s =$$

La ortogonalidad del residuo a la dirección anterior, repaso

Proposición

En las suposiciones anteriores, para cada s,

$$r_{s+1} \perp p_s.$$

Demostración:

$$p_s^\top r_{s+1} = p_s^\top (r_s - \alpha_s q_s) = p_s^\top r_s - \alpha_s p_s^\top q_s = p_s^\top r_s - \frac{p_s^\top r_s}{p_s^\top q_s}\, p_s^\top q_s = 0.$$

El producto interno asociado a una matriz, repaso

Sea $A \in \mathcal{M}_n(\mathbb{R})$, $A > 0$.

El producto interno asociado a una matriz, repaso

Sea $A \in \mathcal{M}_n(\mathbb{R})$, $A > 0$.

Entonces la siguiente función es un producto interno en $\mathbb{R}^n$:

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

El producto interno asociado a una matriz, repaso

Sea $A \in \mathcal{M}_n(\mathbb{R})$, $A > 0$.

Entonces la siguiente función es un producto interno en $\mathbb{R}^n$:

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

Denotamos por $\|\cdot\|_A$ la norma inducida por este producto interno:

$$\|v\|_A := \sqrt{\langle v, v \rangle_A} = \sqrt{\langle Av, v \rangle} = \sqrt{v^\top Av}.$$

El producto interno asociado a una matriz, repaso

Sea $A \in \mathcal{M}_n(\mathbb{R})$, $A > 0$.

Entonces la siguiente función es un producto interno en $\mathbb{R}^n$:

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

Denotamos por $\| \cdot \|_A$ la norma inducida por este producto interno:

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

Dos vectores $v, w \in \mathbb{R}^n$ se llaman *A-ortogonales* o *A-conjugados*, si $\langle v, w \rangle_A = 0$:

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

Métodos de direcciones conjugadas, repaso

Es una subfamilia de los métodos anteriores, cuando se pide una condición adicional: $p_0, p_1, p_2, \ldots$ son A-ortogonales a pares.

Métodos de direcciones conjugadas, repaso

Es una subfamilia de los métodos anteriores, cuando se pide una condición adicional: $p_0, p_1, p_2, \ldots$ son A-ortogonales a pares.

Teorema (ortogonalidad del residuo a las dir. anteriores en los métodos de dir. conj.)

Sean $p_0, p_1, p_2, \ldots \in \mathbb{R}^n$ vectores A-ortogonales.

Supongamos que $r_0 \in \mathbb{R}^n$ y $r_1, r_2, \ldots$ están definidos mediante la regla

$$r_{s+1} := r_s - \alpha_s A p_s, \qquad \text{donde} \qquad \alpha_s = \frac{\langle r_s, p_s \rangle}{\|p_s\|_A^2}.$$

Entonces

$$\forall s \in \mathbb{N} \qquad \forall k \in \{0, \ldots, s-1\} \qquad r_s \perp p_k.$$

Idea del método de gradiente conjugado

Construir p_s como el complemento ortogonal de r_s respecto a $p_0, \ldots, p_{s-1}$, en el sentido del producto interno asociado a la matriz A.

$$\beta_{s,k} := \frac{\langle r_s, p_k \rangle_A}{\langle p_k, p_k \rangle_A} = \frac{p_k^\top A r_s}{p_k^\top A p_k}, \qquad p_s := r_s - \sum_{k=0}^{s-1} \beta_{s,k} p_k.$$

Idea del método de gradiente conjugado

Construir p_s como el complemento ortogonal de r_s respecto a $p_0, \ldots, p_{s-1}$, en el sentido del producto interno asociado a la matriz A.

$$\beta_{s,k} := \frac{\langle r_s, p_k \rangle_A}{\langle p_k, p_k \rangle_A} = \frac{p_k^\top A r_s}{p_k^\top A p_k}, \qquad p_s := r_s - \sum_{k=0}^{s-1} \beta_{s,k} p_k.$$

Entonces, por supuesto, $p_0, p_1, p_2, \ldots$ son A-ortogonales.

Idea del método de gradiente conjugado

Construir p_s como el complemento ortogonal de r_s respecto a $p_0, \ldots, p_{s-1}$, en el sentido del producto interno asociado a la matriz A.

$$\beta_{s,k} := \frac{\langle r_s, p_k \rangle_A}{\langle p_k, p_k \rangle_A} = \frac{p_k^\top A r_s}{p_k^\top A p_k}, \qquad p_s := r_s - \sum_{k=0}^{s-1} \beta_{s,k} p_k.$$

Entonces, por supuesto, $p_0, p_1, p_2, \ldots$ son A-ortogonales.

Es decir, el método del gradiente conjugado es un ejemplo de métodos de direcciones conjugadas.

Idea del método de gradiente conjugado

Construir p_s como el complemento ortogonal de r_s respecto a $p_0, \ldots, p_{s-1}$, en el sentido del producto interno asociado a la matriz A.

$$\beta_{s,k} := \frac{\langle r_s, p_k \rangle_A}{\langle p_k, p_k \rangle_A} = \frac{p_k^\top A r_s}{p_k^\top A p_k}, \qquad p_s := r_s - \sum_{k=0}^{s-1} \beta_{s,k} p_k.$$

Entonces, por supuesto, $p_0, p_1, p_2, \ldots$ son A-ortogonales.

Es decir, el método del gradiente conjugado es un ejemplo de métodos de direcciones conjugadas.

A diferencia de la situación abstracta de métodos de direcciones conjugadas, aquí se propone una receta explícita para construir $p_0, p_1, p_2, \ldots$.

Idea del método de gradiente conjugado

$$\beta_{s,k} := \frac{\langle r_s, p_k \rangle_A}{\langle p_k, p_k \rangle_A} = \frac{p_k^\top A r_s}{p_k^\top A p_k}, \qquad p_s := r_s - \sum_{k=0}^{s-1} \beta_{s,k} p_k.$$

Idea del método de gradiente conjugado

$$\beta_{s,k} := \frac{\langle r_s, p_k \rangle_A}{\langle p_k, p_k \rangle_A} = \frac{p_k^\top A r_s}{p_k^\top A p_k}, \qquad p_s := r_s - \sum_{k=0}^{s-1} \beta_{s,k} p_k.$$

Esta receta parece muy ineficiente, pero sucede un milagro:

$$\forall k \leq s-2 \qquad r_s \perp_A p_k.$$

Idea del método de gradiente conjugado

$$\beta_{s,k} := \frac{\langle r_s, p_k \rangle_A}{\langle p_k, p_k \rangle_A} = \frac{p_k^\top A r_s}{p_k^\top A p_k}, \qquad p_s := r_s - \sum_{k=0}^{s-1} \beta_{s,k} p_k.$$

Esta receta parece muy ineficiente, pero sucede un milagro:

$$\forall k \leq s-2 \qquad r_s \perp_A p_k.$$

Por eso

$$p_s = r_s - \frac{\langle r_s, p_{s-1} \rangle_A}{\|p_{s-1}\|_A^2}\, p_{s-1}.$$

Para estudiar las propiedades de p_k y r_k y explicar el milagro, usaremos los subespacios de Krylov.

Plan

Subespacio generado por una lista de vectores

Definición

Sean $a_1, \ldots, a_m \in \mathbb{R}^n$.

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

Subespacio generado por una lista de vectores

Definición

Sean $a_1, \ldots, a_m \in \mathbb{R}^n$.

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

Es fácil demostrar que

- $\ell(a_1, \ldots, a_m)$ es un subespacio de $\mathbb{R}^n$;
- $a_1, \ldots, a_m \in \ell(a_1, \ldots, a_m)$;
- $\ell(a_1, \ldots, a_m)$ es el mínimo entre los subespacios de $\mathbb{R}^n$ que contienen a $a_1, \ldots, a_m$.

Criterio de la contención de subespacios generados por listas de vectores

Proposición

Sean $a_1, \ldots, a_k, b_1, \ldots, b_m \in \mathbb{R}^n$. Entonces

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

Criterio de la contención de subespacios generados por listas de vectores

Proposición

Sean $a_1, \ldots, a_k, b_1, \ldots, b_m \in \mathbb{R}^n$. Entonces

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

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

Definición

Sea S un subespacio de $\mathbb{R}^n$ y sea $x \in \mathbb{R}^n$.

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

Proposición

Sean $a_1, \ldots, a_k \in \mathbb{R}^n$, $S := \ell(a_1, \ldots, a_k)$.

Entonces para cada x en $\mathbb{R}^n$ las siguientes condiciones son equivalentes:

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

Subespacio de Krylov asociado a una matriz y un vector

Definición

Sean $A \in \mathcal{M}_n(\mathbb{R})$, $v \in \mathbb{R}^n$.

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

Subespacio de Krylov asociado a una matriz y un vector

Definición

Sean $A \in \mathcal{M}_n(\mathbb{R})$, $v \in \mathbb{R}^n$.

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

En la terminología algebraica estándar,

las potencias $I, A, A^2, \dots$ forman un semigrupo,

Subespacio de Krylov asociado a una matriz y un vector

Definición

Sean $A \in \mathcal{M}_n(\mathbb{R})$, $v \in \mathbb{R}^n$.

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

En la terminología algebraica estándar,
las potencias $I, A, A^2, \dots$ forman un semigrupo,
este semigrupo actúa de manera natural a los vectores de $\mathbb{R}^n$,

Subespacio de Krylov asociado a una matriz y un vector

Definición

Sean $A \in \mathcal{M}_n(\mathbb{R})$, $v \in \mathbb{R}^n$.

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

En la terminología algebraica estándar,
las potencias $I, A, A^2, \dots$ forman un semigrupo,
este semigrupo actúa de manera natural a los vectores de $\mathbb{R}^n$,
la sucesión $(A^k v)_{k=0}^{\infty}$ se conoce como la órbita de de v bajo esta acción,

Subespacio de Krylov asociado a una matriz y un vector

Definición

Sean $A \in \mathcal{M}_n(\mathbb{R})$, $v \in \mathbb{R}^n$.

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

En la terminología algebraica estándar,
las potencias $I, A, A^2, \ldots$ forman un semigrupo,
este semigrupo actúa de manera natural a los vectores de $\mathbb{R}^n$,
la sucesión $(A^k v)_{k=0}^{\infty}$ se conoce como la órbita de de v bajo esta acción,
y el subespacio $K_j(A, v)$ está generado por los primeros j elementos de esta órbita.

Una propiedad de los subespacios de Krylov

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

Una propiedad de los subespacios de Krylov

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

Proposición

Sean $A \in \mathcal{M}_n(\mathbb{R})$, $v \in \mathbb{R}^n$, $x \in K_j(A, v)$. Entonces $Ax \in K_{j+1}(A, v)$.

Demostración

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

La suposición $x \in K_j(A, v)$ significa que

Demostración

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

La suposición $x \in K_j(A, v)$ significa que existen $\lambda_0, \dots, \lambda_{j-1} \in \mathbb{R}$ tales que

$$x = \sum_{k=0}^{j-1} \lambda_k A^k v.$$

Demostración

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

La suposición $x \in K_j(A, v)$ significa que existen $\lambda_0, \ldots, \lambda_{j-1} \in \mathbb{R}$ tales que

$$x = \sum_{k=0}^{j-1} \lambda_k A^k v.$$

Luego

$$Ax =$$

Demostración

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

La suposición $x \in K_j(A, v)$ significa que existen $\lambda_0, \ldots, \lambda_{j-1} \in \mathbb{R}$ tales que

$$x = \sum_{k=0}^{j-1} \lambda_k A^k v.$$

Luego

$$Ax = \sum_{k=0}^{j-1} \lambda_k A^{k+1} v \in$$

Demostración

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

La suposición $x \in K_j(A, v)$ significa que existen $\lambda_0, \dots, \lambda_{j-1} \in \mathbb{R}$ tales que

$$x = \sum_{k=0}^{j-1} \lambda_k A^k v.$$

Luego

$$Ax = \sum_{k=0}^{j-1} \lambda_k A^{k+1} v \in \ell(Av, \dots, A^j v)$$

Demostración

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

La suposición $x \in K_j(A, v)$ significa que existen $\lambda_0, \ldots, \lambda_{j-1} \in \mathbb{R}$ tales que

$$x = \sum_{k=0}^{j-1} \lambda_k A^k v.$$

Luego

$$Ax = \sum_{k=0}^{j-1} \lambda_k A^{k+1} v \in \ell(Av, \ldots, A^j v) \subseteq$$

Demostración

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

La suposición $x \in K_j(A,v)$ significa que existen $\lambda_0, \dots, \lambda_{j-1} \in \mathbb{R}$ tales que

$$x = \sum_{k=0}^{j-1} \lambda_k A^k v.$$

Luego

$$Ax = \sum_{k=0}^{j-1} \lambda_k A^{k+1} v \in \ell(Av, \dots, A^j v) \subseteq K_{j+1}(A,v).$$

Plan

Idea del método del gradiente conjugado

- $r_1, r_2, r_3, \ldots$ se construyen de $p_0, p_1, p_2, \ldots$, como en los métodos de direcciones conjugadas;
- $p_0, p_1, p_2, \ldots$ se construyen de $r_0, r_1, r_2, \ldots$, como en la ortogonalización de Gram–Schmidt, respecto al producto $\langle \cdot, \cdot \rangle_A$.

El método del gradiente conjugado, la forma teórica

Empezamos con $r_0 = b - Ax_0$, $p_0 = r_0$.

En cada paso s, como en otros métodos iterativos para $A > 0$,

$$\alpha_s := \frac{\langle r_s, p_s \rangle}{\langle Ap_s, p_s \rangle} = \frac{\langle r_s, p_s \rangle}{\|p_s\|_A^2}, \qquad x_{s+1} := x_s - \alpha_s p_s, \qquad r_{s+1} := r_s - \alpha_s A p_s,$$

$$\beta_{s+1,k} := \frac{\langle r_{s+1}, p_k \rangle_A}{\langle p_k, p_k \rangle_A}, \qquad p_{s+1} := r_{s+1} - \sum_{k=0}^{s} \beta_{s+1,k} p_k.$$

Vamos a demostrar que la fórmula para p_{s+1} se simplifica mucho.

Propiedades “gratuitas” del método de gradiente conjugado

(que salen de la teoría anterior)

- Los vectores $p_0, p_1, p_2, \ldots$ se obtienen por las fórmulas de Gram-Schmidt a partir de $r_0, r_1, r_2, \ldots$, por eso

Propiedades “gratuitas” del método de gradiente conjugado

(que salen de la teoría anterior)

- Los vectores $p_0, p_1, p_2, \ldots$ se obtienen por las fórmulas de Gram-Schmidt a partir de $r_0, r_1, r_2, \ldots$, por eso
$$\forall k < s \qquad p_k \perp_A p_s.$$

Propiedades "gratuitas" del método de gradiente conjugado

(que salen de la teoría anterior)

- Los vectores $p_0, p_1, p_2, \ldots$ se obtienen por las fórmulas de Gram-Schmidt a partir de $r_0, r_1, r_2, \ldots$, por eso
$$\forall k < s \qquad p_k \perp_A p_s.$$

- Los vectores $p_0, p_1, p_2, \ldots$ son A-ortogonales, por eso el método del gradiente conjugado es un caso particular de métodos de direcciones conjugadas, y

Propiedades “gratuitas” del método de gradiente conjugado

(que salen de la teoría anterior)

- Los vectores $p_0, p_1, p_2, \ldots$ se obtienen por las fórmulas de Gram-Schmidt a partir de $r_0, r_1, r_2, \ldots$, por eso

$$\forall k < s \qquad p_k \perp_A p_s.$$

- Los vectores $p_0, p_1, p_2, \ldots$ son A-ortogonales, por eso el método del gradiente conjugado es un caso particular de métodos de direcciones conjugadas, y

$$\forall k < s \qquad p_k \perp r_s.$$

Condición de terminación

Si para algún s obtenemos $r_s = 0_n$, entonces al algoritmo se termina, es decir, no se definen p_k y r_s con $k > s$.

Condición de terminación

Si para algún s obtenemos $r_s = 0_n$, entonces al algoritmo se termina, es decir, no se definen p_k y r_s con $k > s$.

Por la teoría de métodos de direcciones conjugadas, el algoritmo termina su trabajo en no más de n pasos.

Condición de terminación

Si para algún s obtenemos $r_s = 0_n$, entonces al algoritmo se termina, es decir, no se definen p_k y r_s con $k > s$.

Por la teoría de métodos de direcciones conjugadas, el algoritmo termina su trabajo en no más de n pasos.

En la práctica, terminamos el trabajo cuando r_s es suficientemente pequeño.

Fórmula para el coeficiente α_s en el método de gradiente conjugado

Proposición

$$\alpha_s = \frac{\|r_s\|^2}{\|p_s\|_A^2}.$$

Demostración

Recordamos que

$$p_s =$$

Demostración

Recordamos que

$$p_s = r_s - \sum_{k=0}^{s-1} \beta_{s,k} p_k.$$

Obtenemos

$$\langle r_s, p_s \rangle =$$

Demostración

Recordamos que

$$p_s = r_s - \sum_{k=0}^{s-1} \beta_{s,k} p_k.$$

Obtenemos

$$\langle r_s, p_s \rangle = \|r_s\|^2 - \sum_{k=0}^{s-1} \beta_{s,k} \langle p_k, r_s \rangle.$$

Los últimos productos internos son cero por la propiedad de residuos y direcciones en métodos de direcciones conjugadas. Luego

$$\alpha_s = \frac{\langle r_s, p_s \rangle}{\langle A p_s, p_s \rangle} = \frac{\|r_s\|^2}{\|p_s\|_A^2}.$$

Relación entre r_s, r_{s+1} y p_s

$$r_{s+1} = r_s - \alpha_s A p_s .$$

Relación entre r_s, r_{s+1} y p_s

$$r_{s+1} = r_s - \alpha_s A p_s.$$

De aquí,

$$A p_s =$$

Relación entre r_s, r_{s+1} y p_s

$$r_{s+1} = r_s - \alpha_s A p_s.$$

De aquí,

$$\boxed{A p_s = \frac{r_s - r_{s+1}}{\alpha_s}.}$$

Igualdad de ciertos subespacios en el método de gradiente conjugado

Teorema

Para cada s,

$$\ell(p_0, \ldots, p_s) = \ell(r_0, \ldots, r_s) = K_{s+1}(A, r_0).$$

Igualdad de ciertos subespacios en el método de gradiente conjugado

Teorema

Para cada s,

$$\ell(p_0, \ldots, p_s) = \ell(r_0, \ldots, r_s) = K_{s+1}(A, r_0).$$

Observación. La demostración estará basada en las fórmulas

$$r_{s+1} = r_s - \alpha_s A p_s, \qquad p_{s+1} = r_{s+1} - \sum_{k=0}^{s} \beta_{s+1,k} p_k.$$

En esta demostración no serán importantes las fórmulas para α_s y $\beta_{s+1,k}$.

Demostración de $\ell(p_0, \ldots, p_s) = \ell(r_0, \ldots, r_s)$

Usar la inducción sobre s y las siguientes ideas.

$$p_{s+1} = r_{s+1} - \sum_{k=0}^{s} \beta_{s+1,k} p_k \overset{\text{hip. ind.}}{\in} \ell(r_0, \ldots, r_{s+1}),$$

$$r_{s+1} = p_{s+1} + \sum_{k=0}^{s} \beta_{s+1,k} p_k \in \ell(p_0, \ldots, p_{s+1}).$$

Demostración de $\ell(p_0, \ldots, p_s) = \ell(r_0, \ldots, r_s)$

Usar la inducción sobre s y las siguientes ideas.

$$p_{s+1} = r_{s+1} - \sum_{k=0}^{s} \beta_{s+1,k} p_k \overset{\text{hip. ind.}}{\in} \ell(r_0, \ldots, r_{s+1}),$$

$$r_{s+1} = p_{s+1} + \sum_{k=0}^{s} \beta_{s+1,k} p_k \in \ell(p_0, \ldots, p_{s+1}).$$

Es la propiedad de conservación de subespacios en el método de Gram–Schmidt.

Demostración de $\ell(r_0, \ldots, r_s) \subseteq K_{s+1}(A, r_0)$

Inducción sobre s. Por la hipótesis de inducción,

$$r_0, \ldots, r_s \in K_{s+1}(A, r_0).$$

$$r_{s+1} = \underbrace{r_s}_{\substack{\cap \\ K_{s+1}(A, r_0)}} - \alpha_s A \underbrace{p_s}_{\substack{\cap \\ K_{s+1}(A, r_0)}}$$

Como $p_s \in K_{s+1}(A, r_0)$, obtenemos $Ap_s \in K_{s+2}(Ar_0)$ y

$$r_{s+1} \in K_{s+2}(A, r_0).$$

Demostración de $K_{s+1}(A, r_0) \subseteq \ell(p_0, \ldots, p_s)$

Por inducción sobre s. Suponemos que

$$r_0, Ar_0, \ldots, A^s r_0 \in \ell(p_0, \ldots, p_s).$$

Notamos que

$$Ap_s = \frac{r_s - r_{s+1}}{\alpha_s} \in \ell(r_0, \ldots, r_{s+1}), \tag{1}$$

Luego

$$A^{s+1} r_0 = A(A^s r_0) \overset{\text{hip. ind.}}{\in} A\ell(p_0, \ldots, p_s) = \ell(Ap_0, \ldots, Ap_s)$$
$$\overset{(1)}{\subseteq} \ell(r_0, \ldots, r_{s+1}) = \ell(p_0, \ldots, p_{s+1}).$$

Ejercicio. Usando las ideas previas, demostrar directamente que

$$\ell(p_0) = \ell(r_0),$$

$$\ell(p_0, p_1) = \ell(r_0, r_1) = \ell(r_0, Ar_0),$$

$$\ell(p_0, p_1, p_2) = \ell(r_0, r_1, r_2) = \ell(r_0, Ar_0, A^2 r_0).$$

Ciertas ortogonalidades en el método de gradiente conjugado

Hemos demostrado que para cada s,

$$\ell(p_0, \ldots, p_s) = \ell(r_0, \ldots, r_s) = K_{s+1}(A, r_0).$$

Corolario

Para cada s,

$$p_s \perp_A K_s(A, r_0), \qquad r_s \perp K_s(A, r_0).$$

Los vectores $p_0, p_1, p_2, \ldots$ son A-ortogonales por pares, y los vectores $r_0, r_1, r_2, \ldots$ son ortogonales por pares.

Simplificación de la fórmula para p_{s+1}

Corolario

Para cada s,

$$\forall k \leq s-1 \qquad r_{s+1} \perp_A p_k.$$

Por consecuencia, $\beta_{s+1,0} = \ldots = \beta_{s+1,s-1} = 0$ y

$$p_{s+1} = r_{s+1} - \beta_{s+1,s} p_s.$$

Simplificación de la fórmula para p_{s+1}

Corolario

Para cada s,

$$\forall k \leq s-1 \qquad r_{s+1} \perp_A p_k.$$

Por consecuencia, $\beta_{s+1,0} = \ldots = \beta_{s+1,s-1} = 0$ y

$$p_{s+1} = r_{s+1} - \beta_{s+1,s} p_s.$$

Demostración. Para cualquier k en $\{0, \ldots, s-1\}$,

$$\langle r_{s+1}, p_k \rangle_A = \langle r_{s+1}, A p_k \rangle = \frac{1}{\alpha_k} \langle r_j - r_{j+1}, r_{s+1} \rangle = 0.$$

Fórmula para $\beta_{s,s-1}$

Corolario

$$\beta_{s+1,s} = -\frac{\|r_{s+1}\|^2}{\|r_s\|^2},$$

Fórmula para $\beta_{s,s-1}$

Corolario

$$\beta_{s+1,s} = -\frac{\left\|r_{s+1}\right\|^2}{\left\|r_s\right\|^2},$$

Demostración. Recordemos que $\beta_{s+1,s} = \frac{\langle r_{s+1}, p_s \rangle_A}{\|p_s\|_A^2}$.

Fórmula para $\beta_{s,s-1}$

Corolario

$$\beta_{s+1,s} = -\frac{\left\|r_{s+1}\right\|^2}{\left\|r_s\right\|^2},$$

Demostración. Recordemos que $\beta_{s+1,s} = \frac{\langle r_{s+1}, p_s\rangle_A}{\|p_s\|_A^2}$.
Simplificamos el numerador y el denominador:

$$\langle r_{s+1}, p_s\rangle_A =$$

Fórmula para $\beta_{s,s-1}$

Corolario

$$\beta_{s+1,s} = -\frac{\|r_{s+1}\|^2}{\|r_s\|^2},$$

Demostración. Recordemos que $\beta_{s+1,s} = \frac{\langle r_{s+1}, p_s\rangle_A}{\|p_s\|_A^2}$.
Simplificamos el numerador y el denominador:

$$\langle r_{s+1}, p_s\rangle_A = \langle r_{s+1}, Ap_s\rangle =$$

Fórmula para $\beta_{s,s-1}$

Corolario

$$\beta_{s+1,s} = -\frac{\|r_{s+1}\|^2}{\|r_s\|^2},$$

Demostración. Recordemos que $\beta_{s+1,s} = \frac{\langle r_{s+1}, p_s \rangle_A}{\|p_s\|_A^2}$.
Simplificamos el numerador y el denominador:

$$\langle r_{s+1}, p_s \rangle_A = \langle r_{s+1}, Ap_s \rangle = \frac{1}{\alpha_s} \langle r_s - r_{s+1}, r_{s+1} \rangle =$$

Fórmula para $\beta_{s,s-1}$

Corolario

$$\beta_{s+1,s} = -\frac{\|r_{s+1}\|^2}{\|r_s\|^2},$$

Demostración. Recordemos que $\beta_{s+1,s} = \frac{\langle r_{s+1}, p_s\rangle_A}{\|p_s\|_A^2}$.
Simplificamos el numerador y el denominador:

$$\langle r_{s+1}, p_s\rangle_A = \langle r_{s+1}, Ap_s\rangle = \frac{1}{\alpha_s}\langle r_s - r_{s+1}, r_{s+1}\rangle = -\frac{1}{\alpha_s}\|r_{s+1}\|^2.$$

Fórmula para $\beta_{s,s-1}$

Corolario

$$\beta_{s+1,s} = -\frac{\|r_{s+1}\|^2}{\|r_s\|^2},$$

Demostración. Recordemos que $\beta_{s+1,s} = \frac{\langle r_{s+1}, p_s \rangle_A}{\|p_s\|_A^2}$.
Simplificamos el numerador y el denominador:

$$\langle r_{s+1}, p_s \rangle_A = \langle r_{s+1}, A p_s \rangle = \frac{1}{\alpha_s} \langle r_s - r_{s+1}, r_{s+1} \rangle = -\frac{1}{\alpha_s} \|r_{s+1}\|^2.$$

$$\|p_s\|_A^2 =$$

Fórmula para $\beta_{s,s-1}$

Corolario

$$\beta_{s+1,s} = -\frac{\|r_{s+1}\|^2}{\|r_s\|^2},$$

Demostración. Recordemos que $\beta_{s+1,s} = \frac{\langle r_{s+1}, p_s\rangle_A}{\|p_s\|_A^2}$.
Simplificamos el numerador y el denominador:

$$\langle r_{s+1}, p_s\rangle_A = \langle r_{s+1}, Ap_s\rangle = \frac{1}{\alpha_s}\langle r_s - r_{s+1}, r_{s+1}\rangle = -\frac{1}{\alpha_s}\|r_{s+1}\|^2.$$

$$\|p_s\|_A^2 = \langle Ap_s, p_s\rangle =$$

Fórmula para $\beta_{s,s-1}$

Corolario

$$\beta_{s+1,s} = -\frac{\|r_{s+1}\|^2}{\|r_s\|^2},$$

Demostración. Recordemos que $\beta_{s+1,s} = \frac{\langle r_{s+1}, p_s \rangle_A}{\|p_s\|_A^2}$.
Simplificamos el numerador y el denominador:

$$\langle r_{s+1}, p_s \rangle_A = \langle r_{s+1}, Ap_s \rangle = \frac{1}{\alpha_s} \langle r_s - r_{s+1}, r_{s+1} \rangle = -\frac{1}{\alpha_s} \|r_{s+1}\|^2.$$

$$\|p_s\|_A^2 = \langle Ap_s, p_s \rangle = \frac{1}{\alpha_s} \langle r_s - r_{s+1}, p_s \rangle =$$

Fórmula para $\beta_{s,s-1}$

Corolario

$$\beta_{s+1,s} = -\frac{\|r_{s+1}\|^2}{\|r_s\|^2},$$

Demostración. Recordemos que $\beta_{s+1,s} = \frac{\langle r_{s+1}, p_s\rangle_A}{\|p_s\|_A^2}$.
Simplificamos el numerador y el denominador:

$$\langle r_{s+1}, p_s\rangle_A = \langle r_{s+1}, Ap_s\rangle = \frac{1}{\alpha_s}\langle r_s - r_{s+1}, r_{s+1}\rangle = -\frac{1}{\alpha_s}\|r_{s+1}\|^2.$$

$$\|p_s\|_A^2 = \langle Ap_s, p_s\rangle = \frac{1}{\alpha_s}\langle r_s - r_{s+1}, p_s\rangle = \frac{1}{\alpha_s}\langle r_s, p_s\rangle =$$

Fórmula para $\beta_{s,s-1}$

Corolario

$$\beta_{s+1,s} = -\frac{\|r_{s+1}\|^2}{\|r_s\|^2},$$

Demostración. Recordemos que $\beta_{s+1,s} = \frac{\langle r_{s+1}, p_s \rangle_A}{\|p_s\|_A^2}$.
Simplificamos el numerador y el denominador:

$$\langle r_{s+1}, p_s \rangle_A = \langle r_{s+1}, Ap_s \rangle = \frac{1}{\alpha_s} \langle r_s - r_{s+1}, r_{s+1} \rangle = -\frac{1}{\alpha_s} \|r_{s+1}\|^2.$$

$$\|p_s\|_A^2 = \langle Ap_s, p_s \rangle = \frac{1}{\alpha_s} \langle r_s - r_{s+1}, p_s \rangle = \frac{1}{\alpha_s} \langle r_s, p_s \rangle = \frac{1}{\alpha_s} \|r_s\|^2.$$

Pequeña tarea adicional

Demostrar o refutar (con un contraejemplo) la siguiente afirmación. Para cada s,

$$\|r_{s+1}\| \leq \|r_s\|.$$

Plan

Algoritmo

```
function [x, s] = cg(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;  al = (p' * r) / (p' * q);
        x = x + al * p;  r = r - al * q;
        rrold = rr;  rr = r' * r;
        be = - rr / rrold;  p = r - be * p;
        s = s + 1;
    end
end
```

Pruebas

Se recomienda aplicar el método del gradiente y el método del gradiente conjugado a sistemas $Ax = b$, con $A \in \mathcal{M}_n(\mathbb{R})$, $A > 0$.

Poner $n = 512$, $n = 1024$, $n = 2048$, $\mathrm{smax} = 100$, $\mathrm{tol} = 10^{-10}$.

Comparar el número de pasos.

Para construir $A > 0$, se puede usar la siguiente receta:

```
B = randn(n, n);
A = B' * B + eye(n);
```