

# **MÉTODOS DE KRYLOV**

ALAN REYES-FIGUEROA  
MÉTODOS NUMÉRICOS II

(AULA 10) 09.AGOSTO.2021

# Métodos de Richardson

Recordemos los métodos iterativos para resolver sistemas lineales. En el Aula 06, mencionamos que una técnica general para diseñar métodos iterativos lineales se basa en una descomposición aditiva de  $A$ , en la forma  $A = P - N$ , con  $P$  no singular.  $P$  se llama **matriz de preconditionamiento** o **precondicionador**.

Dado  $\mathbf{x}^{(0)} \in \mathbb{R}^n$ , se puede calcular la secuencia de iterados  $\{\mathbf{x}^{(k)}\}_{k \geq 0}$ , como  $P\mathbf{x}^{(k+1)} = N\mathbf{x}^{(k)} + \mathbf{b}$ ,  $k \geq 0$ . La ecuación anterior puede escribirse en la forma

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + P^{-1}\mathbf{r}^{(k)}, \quad k \geq 0. \quad (1)$$

donde  $\mathbf{r}^{(k)} = \mathbf{b} - A\mathbf{x}^{(k)}$  es el vector residual en el paso  $k$ .

## Definición

Un **método de** RICHARDSON, es cualquier método de la forma

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k P^{-1}\mathbf{r}^{(k)}, \quad k \geq 0, \quad (2)$$

donde  $\alpha_k \in \mathbb{R}$ , es llamado el **parámetro de aceleración**.

# Métodos de Richardson

Implementar un método de RICHARDSON, involucra los siguientes pasos:

- Calcular  $\mathbf{z}^{(k)} = P^{-1}\mathbf{r}^{(k)}$ .
- Calcular el parámetro de aceleración  $\alpha_k$ .
- Calcular  $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{z}^{(k)}$ .
- Calcular el residuo  $\mathbf{r}^{(k+1)} = \mathbf{b} - A\mathbf{x}^{(k+1)} = \mathbf{r}^{(k)} - \alpha_k A\mathbf{z}^{(k)}$ .

Si  $P = I$  se dice que el método está **no preconditionado**.

## Ejemplos:

- El método de JACOBI se obtiene con  $\alpha_k = 1$  y  $P = D$ .
- El método de SEIDEL se obtiene con  $\alpha_k = 1$  y  $P = D - L$ .

Consideremos la iteración de RICHARDSON no preconditionada (2)

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha \mathbf{r}^{(k)} = \mathbf{x}^{(k)} - \alpha(\mathbf{b} - A\mathbf{x}^{(k)}), \quad k \geq 0.$$

# Métodos de Richardson

que se puede reescribir como

$$\mathbf{x}^{(k+1)} = (I - \alpha A)\mathbf{x}^{(k)} + \alpha \mathbf{b}, \quad (3)$$

En este caso, la matriz de iteración es  $T_\alpha = I - \alpha A$ , y  $\mathbf{c}_\alpha = \alpha \mathbf{b}$ . Si los autovalores de  $A$  son  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ , entonces los autovalores  $\mu_i = 1 - \alpha \lambda_i$  de  $T_\alpha$  satisfacen

$$1 - \alpha \lambda_1 = \mu_1 \leq \mu_2 \leq \dots \leq \mu_n = 1 - \alpha \lambda_n.$$

## Observaciones:

- Si  $\lambda_n < 0$  y  $\lambda_1 > 0$ , el método diverge.
- Si los autovalores de  $A$  son todos positivos, se ha de cumplir que

$$\mu_n = 1 - \alpha \lambda_n < 1, \quad \text{y} \quad \mu_1 = 1 - \alpha \lambda_1 > -1.$$

En particular,  $-1 < \lambda_i < 1, \forall i$ , y  $\rho(I - \alpha A) < 1$ . Además,  $0 < \alpha < \frac{2}{\lambda_1}$ .

- El valor de óptimo para  $\alpha$  es

$$\alpha_{opt} = \frac{2}{\lambda_1 + \lambda_n}.$$

# Precondicionadores

Dado un método iterativo

$$\mathbf{x}^{(k+1)} = T\mathbf{x}^{(k)} + \mathbf{c},$$

éste puede verse como una técnica para resolver el sistema  $(I - T)\mathbf{x} = \mathbf{c}$ .

Comparando con la iteración

$$\mathbf{x}^{(k+1)} = P^{-1}N\mathbf{x}^{(k)} + P^{-1}\mathbf{b},$$

se tiene que  $I - T = P^{-1}N$ , así  $T = I - P^{-1}N = P^{-1}(P - N) = P^{-1}A$ . Así, el método iterativo se puede ver como una técnica para resolver el sistema preconditionado

$$P^{-1}A\mathbf{x} = P^{-1}\mathbf{b}.$$

## Ejemplos:

- JACOBI,  $P = D$ .
- SEIDEL,  $P = D - L$ .
- SOR,  $P = \frac{1}{\omega}(D - \omega L)$ .

# Precondicionadores

Estos preconditionadores son de la forma  $P^{-1} = f(A)$ . Un caso más sofisticado son los preconditionadores de NEWMAN. En éstos, se asume que la matriz  $A$  se escribe como

$$A = D - C = (I - CD^{-1})D,$$

con lo que

$$A^{-1} = D^{-1}(I - CD^{-1})^{-1} = D^{-1}(I + CD^{-1} + (CD^{-1})^2 + \dots).$$

Se obtienen los preconditionadores de NEUMAN truncando la serie. Este método funciona si  $\rho(CD^{-1}) < 1$ .

# Métodos de Krylov

Recordemos que en el método de Richardson (no preconditionado, esto es  $\alpha = 1$ ,  $P = I$ ), podemos escribir la secuencia de iterados como

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{r}^{(k)}, \quad k \geq 0. \quad (4)$$

Como  $\mathbf{r}^{(k)} = \mathbf{b} - A\mathbf{x}^{(k)}$ , entonces tenemos

$$\mathbf{r}^{(k+1)} = \mathbf{b} - A\mathbf{x}^{(k+1)} = \mathbf{b} - A(\mathbf{x}^{(k)} + \mathbf{r}^{(k)}) = (\mathbf{b} - A\mathbf{x}^{(k)}) - A\mathbf{r}^{(k)} = \mathbf{r}^{(k)} - A\mathbf{r}^{(k)} = (I - A)\mathbf{r}^{(k)}.$$

En particular, podemos escribir

$$\mathbf{r}^{(k)} = (I - A)\mathbf{r}^{(k-1)} = (I - A)^2\mathbf{r}^{(k-2)} = (I - A)^3\mathbf{r}^{(k-3)} = \dots = (I - A)^k\mathbf{r}^{(0)}, \quad k \geq 0.$$

De allí que

$$\begin{aligned} \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \mathbf{r}^{(k)} = \mathbf{x}^{(k-1)} + \mathbf{r}^{(k-1)} + \mathbf{r}^{(k)} = \mathbf{x}^{(k-2)} + \mathbf{r}^{(k-2)} + \mathbf{r}^{(k-1)} + \mathbf{r}^{(k)} \\ &= \dots \\ &= \mathbf{x}^{(0)} + \sum_{i=0}^k \mathbf{r}^{(i)} = \mathbf{x}^{(0)} + \sum_{i=0}^k (I - A)^i \mathbf{r}^{(0)}. \end{aligned}$$

# Métodos de Krylov

Vimos que el método de Richardson construye la secuencia de iterados

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(0)} + \sum_{i=0}^k (I - A)^i \mathbf{r}^{(0)}, \quad k \geq 0. \quad (5)$$

De allí,  $\mathbf{x}^{(k+1)}$ , pertenece al espacio generado  $\langle \mathbf{x}^{(0)}, \mathbf{r}^{(0)}, A\mathbf{r}^{(0)}, A^2\mathbf{r}^{(0)}, \dots, A^k\mathbf{r}^{(0)} \rangle$ .

Similarmente, para los residuos se tiene

$$\mathbf{r}^{(k+1)} = \mathbf{b} - A\mathbf{x}^{(0)} + A \sum_{i=0}^k (I - A)^i \mathbf{r}^{(0)} = \mathbf{r}^{(0)} - A \sum_{i=0}^k (I - A)^i \mathbf{r}^{(0)}, \quad k \geq 0, \quad (6)$$

de modo que  $\mathbf{r}^{(k+1)}$ , pertenece al espacio generado  $\langle \mathbf{r}^{(0)}, A\mathbf{r}^{(0)}, A^2\mathbf{r}^{(0)}, \dots, A^{k+1}\mathbf{r}^{(0)} \rangle$ .

Estos espacios de la forma  $\langle \mathbf{v}, A\mathbf{v}, A^2\mathbf{v}, \dots, A^k\mathbf{v} \rangle$  son de mucha importancia en el álgebra lineal numérica.



## Definición

El espacio  $\langle \mathbf{r}^{(0)}, A\mathbf{r}^{(0)}, A^2\mathbf{r}^{(0)}, \dots, A^{k-1}\mathbf{r}^{(0)} \rangle$  se llama **subespacio de KRYLOV** de dimensión  $k$ , correspondiente a la matriz  $A$  y al residuo inicial  $\mathbf{r}^{(0)}$ .

$$\mathcal{K}_k(A, \mathbf{r}^{(0)}) = \langle \mathbf{r}^{(0)}, A\mathbf{r}^{(0)}, A^2\mathbf{r}^{(0)}, \dots, A^{k-1}\mathbf{r}^{(0)} \rangle.$$

**Pregunta:** ¿Se puede usar la información contenida en  $\mathcal{K}_k(A, \mathbf{r}^{(0)})$  de forma más eficiente que con el método de Richardson?

Si consideramos  $\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + \sum_{i=0}^{k-1} \alpha_i \mathbf{r}^{(i)}$ , entonces  $\mathbf{x}^{(k)}$  pertenece al espacio

$$W_k = \{ \mathbf{v} = \mathbf{x}^{(0)} + \mathbf{y} : \mathbf{y} \in \mathcal{K}_k(A, \mathbf{r}^{(0)}) \} = \mathbf{x}^{(0)} + \mathcal{K}_k(A, \mathbf{r}^{(0)}).$$

Se pueden plantear métodos de la forma

$$\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + q_{k-1}(A)\mathbf{r}^{(0)},$$

donde  $q_{k-1}$  es un polinomio de grado  $k - 1$  que se aplica sobre la matriz  $A$ , y se selecciona de forma que  $\mathbf{x}^{(k)}$  sea la mejor aproximación de  $\mathbf{x}$  en el espacio  $W_k$ .

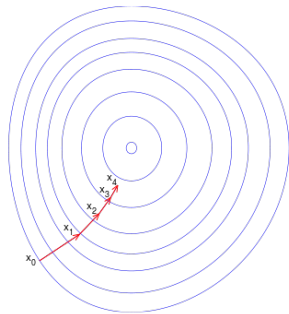
# Métodos de Krylov

**Ejemplo:** Método de descenso más rápido (*steepest descent*)

Por ejemplo, el método de descenso más rápido para optimización, se basa en iteraciones de la forma

$$\begin{aligned}\mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \alpha_k \mathbf{r}^{(k)}, \\ \mathbf{r}^{(k+1)} &= \mathbf{b} - \mathbf{A}\mathbf{x}^{(k+1)} = \mathbf{r}^{(k)} - \alpha_k \mathbf{A}\mathbf{r}^{(k)}.\end{aligned}$$

Para este método  $\mathbf{x}^{(k)} \in W_k$ .



# Métodos de Krylov

Los métodos de proyección de un paso se basan en una combinación óptima de los dos últimos vectores base del subespacio de KRYLOV.

¿Es posible contruir una combinación lineal óptima de todos los vectores base del subespacio de KRYLOV?

La respuesta en dos pasos:

1. Primero veremos cómo construir una base adecuada para  $\mathcal{K}_k(A, \mathbf{r}^{(0)})$ ,
2. Después veremos cómo construir una aproximación óptima como una combinación lineal de los vectores base (al menos para matrices simétricas).

La base más simple para el subespacio de KRYLOV  $\mathcal{K}_k(A, \mathbf{r}^{(0)})$  es la base:  $\mathbf{r}^{(0)}, A\mathbf{r}^{(0)}, A^2\mathbf{r}^{(0)}, \dots, A^{k-1}\mathbf{r}^{(0)}$ . Sin embargo, esta base es mal condicionada ya que  $A^{k-1}\mathbf{r}^{(0)}$  apunta cada vez más en la dirección del autovector dominante de  $A$ .

Una base estable y ortogonal se puede construir usando el método de ARNOLDI.

# Métodos de Krylov

**Algoritmo:** (Iteración de ARNOLDI).

*Inputs:*  $A \in \mathbb{R}^{n \times n}$ .

*Outputs:*  $Q \in \mathbb{R}^{n \times n}$ , ortogonal con una base  $\{\mathbf{q}_i\}$  ortonormal de  $\mathcal{K}_n(A, \mathbf{q}_1)$  como columnas,  $H \in \mathbb{R}^{n \times n}$  en la forma de Hessemberg.

Initialize  $H = \mathbf{0}$ , choose  $\mathbf{q}_1 \in \mathbb{R}^n$  con  $\|\mathbf{q}_1\|_2 = 1$ .

For  $k = 1, 2, \dots, n$ :

$\mathbf{v} = A\mathbf{q}_k$ .

    For  $i = 1, 2, \dots, k$ :

$H_{ik} = \mathbf{v}^T \mathbf{q}_i$ ,

$\mathbf{v} = \mathbf{v} - H_{ik} \mathbf{q}_i$ , (ortogonalización)

$H_{k+1,k} = \|\mathbf{v}\|_2$ ,

    If  $(H_{k+1,k} = 0)$ : stop. (Subespacio invariante).

$\mathbf{q}_{k+1} = \mathbf{v} / H_{k+1,k}$ .

Return  $Q, H$ .

Los vectores  $\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$  son llamados los **vectores de Arnoldi** de  $A$ .

# Métodos de Krylov

El método de ARNOLDI se puede resumir en el siguiente proceso. En cada paso, se construye una matriz  $H_k$  en la forma de Hessemberg

$$H_k = \begin{pmatrix} h_{11} & \dots & \dots & h_{1k} \\ h_{21} & \ddots & & \vdots \\ & \ddots & \ddots & \vdots \\ & & h_{k,k-1} & h_{kk} \end{pmatrix},$$

y una matriz ortogonal  $Q_k = (\mathbf{q}_1 \quad \mathbf{q}_2 \quad \dots \quad \mathbf{q}_k)$ , de forma que

$$AQ_k = Q_k H_k + h_{k+1,k} \mathbf{q}_k \mathbf{e}_k^T,$$

donde  $\mathbf{e}_k$  es el  $k$ -ésimo vector de la base canónica de  $\mathbb{R}^k$ .

En el caso en que  $A$  es simétrica, de acuerdo a la iteración de ARNOLDI,  $Q_k^T A Q_k = H_k$ .

En este caso,  $A$  simétrica implica que

$$H_k^T = (Q_k^T A Q_k)^T = Q_k^T A Q_k = H_k,$$

así que  $H_k$  es simétrica y Hessenberg superior. Luego,  $H_k$  es tridiagonal.

# Métodos de Krylov

Así

$$H_k = \begin{pmatrix} h_{11} & h_{21} & & \\ h_{21} & h_{22} & \ddots & \\ & \ddots & \ddots & h_{k,k-1} \\ & & h_{k,k-1} & h_{kk} \end{pmatrix}.$$

Con  $\alpha_k = h_{kk}$  y  $\beta_k = h_{k,k-1}$ , el método de Arnoldi se simplifica en el **método de LANCZOS**. Con el método de LANCZOS es posible calcular un nuevo vector base ortogonal utilizando sólo los dos vectores base previos. En este caso,

$$H_k = \begin{pmatrix} \alpha_1 & \beta_2 & & \\ \beta_2 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_k \\ & & \beta_k & \alpha_k \end{pmatrix}.$$

# Métodos de Krylov

**Algoritmo:** (Iteración de LANCZOS)

*Inputs:*  $A \in \mathbb{R}^{n \times n}$  simétrica.

*Outputs:*  $Q \in \mathbb{R}^{n \times n}$ , ortogonal con una base  $\{\mathbf{q}_i\}$  ortonormal de  $\mathcal{K}_n(A, \mathbf{q}_1)$  como columnas,  $H \in \mathbb{R}^{n \times n}$  tridiagonal.

Initialize  $H = \mathbf{0}$ , choose  $\mathbf{q}_1 \in \mathbb{R}^n$  con  $\|\mathbf{q}_1\|_2 = 1$ .

$\mathbf{v}_1 = A\mathbf{q}_1$ ,

$\alpha_1 = \mathbf{v}_1^T \mathbf{q}_1$ ,

$\mathbf{v}_1 = \mathbf{v}_1 - \alpha_1 \mathbf{q}_1$ .

For  $i = 2, 3, \dots, k$ :

$\beta_i = \|\mathbf{v}_{i-1}\|_2$ ,

If  $(\beta_i \neq 0)$ :  $\mathbf{q}_i = \mathbf{q}_{i-1} / \beta_i$ .

Else: Choose another  $\mathbf{q}_i \in \langle \mathbf{q}_1, \dots, \mathbf{q}_{i-1} \rangle^\perp$  with  $\|\mathbf{q}_i\| = 1$ .

$\mathbf{v}_i = A\mathbf{q}_i$ ,

$\alpha_i = \mathbf{v}_i^T \mathbf{q}_i$ ,

$\mathbf{v}_i = \mathbf{v}_i - \alpha_i \mathbf{q}_i - \beta_i \mathbf{q}_{i-1}$ , (ortogonalización).

Return  $Q, H$ .

# Cálculo de Autovalores

El método de ARNOLDI y el método de LANCZOS se propusieron originalmente como métodos iterativos para calcular autovalores de la matriz  $A$ .

Como  $H_k = Q_k^T A Q_k$ , para  $k = 1, 2, \dots, n$ , tenemos que

$$H_n = Q_n^T A Q_n,$$

es similar a  $A$ , y portanto posee los mismos autovalores. Los autovalores de  $H_k$  se llaman los **valores de RITZ** de  $A$ .

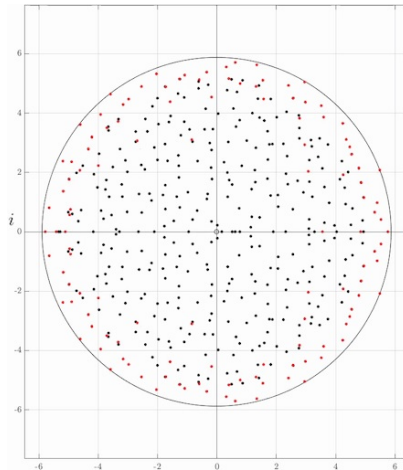
Como  $H_n$  está en la forma de Hessemberg, sus autovalores pueden calcularse de manera eficiente, por ejemplo aplicando el algoritmo  $QR$ .

**Obs!** En la práctica, los autovalores de Ritz convergen a los autovalores de  $A$ . Por lo general, los autovalores de Ritz convergen primero a los autovalores dominantes de  $A$ . Para obtener los autovalores menores de  $A$ , se puede usar la inversa  $A^{-1}$  en su lugar, o alguna técnica de *shift*.

Este es un ejemplo del **método Rayleigh-Ritz**.



# Métodos de Krylov



Iteración 122 del método de ARNOLDI par una matriz aleatoria  $A$ .