

Métodos iterativos para la solución de problemas inversos discretos

Rivas Jazmin et. al.

Contents

1	Introducción	2
1.1	Metodos Iterativos	2
1.1.1	Métodos estandar	4
1.1.2	Métodos de Krylov	5
1.1.3	Metodos de Proyección	6
1.1.4	Algoritmo de Arnoldi	7

Chapter 1

Introducción

En este texto se pretende estudiar el problema de enfoque de imágenes. A partir de una imagen borrosa con ruido y una matriz de desenfoque, nuestro objetivo es encontrar nuestra imagen original utilizando su versión borrosa. Este problema se puede formular a través de un sistema lineal $Ax = b$, donde A es la matriz de desenfoque, b es el ruido y x es la imagen a restaurar. Debido a su mal condicionamiento, nos interesan los métodos de regularización para la reconstrucción de las mismas. En particular, nos interesaremos por los métodos de regularización de Krylov. Estos métodos están basados en procesos de proyección sobre un tipo de subespacios muy particulares, de los cuales el método toma el nombre. Este capítulo servirá de introducción a la estructura general de esta clase de métodos.

1.1 Metodos Iterativos

Consideremos el problema lineal dado por el sistema $Ax = b$. Sabemos que existen métodos directos para la resolución de los mismos (descomposiciones QR, de Cholesky, PALU, entre otros), pero estos métodos traen consigo limitaciones. El primer obstáculo que nos podemos encontrar viene con la cantidad de operaciones que requieren y el tiempo computacional que inducen cuando la matriz A sube de dimensión, la estructura que puede tener o incluso no tenerla de forma explícita y que venga en forma de operador lineal o función.

A raíz de estas limitaciones con los métodos directos, surgen los métodos iterativos. La idea es que dado un sistema $Ax = b$ y un vector inicial x_0 , el

método produce una sucesión $\{x_n\}_{n \in \mathbb{N}}$ de soluciones del sistema que, dadas las condiciones necesarias, convergen a x .

Para plantear un método, dado el sistema lineal y la matriz A , la reescribimos como $A = M + N$, lo que nos deja con el sistema:

$$(M - N)x = b$$

El cual se traduce a:

$$Mx = b + Nx$$

Y si resolvemos para x tenemos que:

$$x = M^{-1}(b + Nx)$$

Y de acá podemos deducir la fórmula recursiva de método:

$$x_{n+1} = M^{-1}(b + Nx_n)$$

Y si observamos que $N = M - A$ concluimos que:

$$x_{n+1} = M^{-1}(b + (M - A)x_n) = M^{-1}b + x_n - M^{-1}Ax_n$$

Luego,

$$x_{n+1} = x_n + M^{-1}(b - Ax_n)$$

La cual es la fórmula general para cualquier método iterativo.

Pero habíamos dicho que para asegurar que la sucesión $\{x_n\}_{n \in \mathbb{N}}$ generada por el método resultase convergente necesitábamos que se den ciertas condiciones, vamos a verlas:

Error de los métodos

Partiendo de la forma del método:

$$x = M^{-1}b + (I - M^{-1}A)x$$

Tenemos que la forma iterativa viene dada por:

$$x_{n+1} = M^{-1}b + (I - M^{-1}A)x_n$$

Y entonces tenemos que la diferencia entre iteraciones nos da el error del paso $n + 1$:

$$e_{n+1} = (I - M^{-1}A)e_n$$

Entonces si llamamos $B = I - M^{-1}A$ tenemos que:

$$e_{n+1} = Be_n = BB e_{n-1} = \cdots = B^{n+1}e_0$$

Y por lo tanto tenemos que:

$$\|e_{n+1}\| \leq \|B\|^{n+1} \|e_0\|$$

Notemos que si $\|B\| < 1$, entonces $\|e_n\| \xrightarrow{n \rightarrow \infty} 0$.

Convergencia de los métodos

Definición. Decimos que un método iterativo converge si para todo $x_0 \in \mathbb{R}^n$ existe x tal que:

$$\lim_{n \rightarrow \infty} x_n = x$$

Teorema. Sea $A \in \mathbb{R}^{n \times n}$, entonces:

$$\rho(A) = \inf_{\|\cdot\|} \|A\|$$

Teorema. Sea $A \in \mathbb{R}^{n \times n}$, entonces:

$$A^n \xrightarrow{n \rightarrow \infty} 0 \iff \rho(A) < 1$$

Teorema. Dado el sistema lineal $Ax = b$, consideremos el método iterativo:

$$x_{n+1} = M^{-1}b + (I - M^{-1}A)x_n$$

y sea $B = I - M^{-1}A$, entonces el método converge si y solo si $\rho(B) < 1$.

1.1.1 Métodos estandar

Habíamos quedado en que un método iterativo viene dado por la fórmula recursiva

$$x_{n+1} = x_n + M^{-1}(b - Ax_n)$$

Pero para poder hacer esto, necesitábamos descomponer nuestra matriz A y toda la iteración iba a depender de eso.

Dado un sistema lineal, podemos descomponer a su matriz A como $A = L + D + U$, con L su parte triangular inferior estricta, D su diagonal y U como su parte triangular superior estricta. Y tenemos que

$$Ax = b \iff Dx = -(L + U)x + b \iff (L + D)x = -Ux + b$$

Con esto podemos definir los dos métodos iterativos más comunes.

Método de Jacobi

Sabiendo que $Dx = -(L+U)x+b$, podemos escribir la forma recursiva como:

$$Dx_{n+1} = -(L+U)x_n + b$$

Y esto nos dice que:

$$x_{n+1} = -D^{-1}(L+U)x_n + D^{-1}b$$

Y esta es la iteración del método de Jacobi.

Método de Gauss-Seidel

Partiendo de que $(L+D)x = -Ux+b$ es equivalente al sistema $Ax = b$, vemos que su forma recursiva está dada por:

$$(L+D)x_{n+1} = -Ux_n + b$$

Y esto lo podemos reescribir como:

$$x_{n+1} = -(L+U)^{-1}Ux_n + (L+U)^{-1}b$$

Que es la iteración del método de Gauss-Seidel.

En general, podemos decir que las iteraciones de los métodos vienen dadas por $x_{n+1} = Bx_n + C$, y en el caso de Jacobi, $B_J = -D^{-1}(L+U)$ y $C_J = D^{-1}b$ y para Gauss-Seidel $B_{GS} = -(L+U)^{-1}U$ y $C_{GS} = (L+U)^{-1}b$.

1.1.2 Métodos de Krylov

Definición. Dada A una matriz y $v, Av, \dots, A^{m-1}v \in \mathbb{C}^m$ vectores linealmente independientes, llamamos subespacio de Krylov de dimensión m a:

$$\mathcal{K}_m(A, v) = \langle v, Av, \dots, A^{m-1}v \rangle$$

Los métodos de Krylov son métodos iterativos que utilizan subespacios de Krylov que, dado x_0 una aproximación inicial de la solución del sistema, se define el subespacio $\mathcal{K}_m(A, r_0)$, con $r_0 = b - Ax_0$.

Para entender de donde surge la idea de utilizar subespacios de Krylov, notemos que dado la solución $x_{true} = A^{-1}b$ y x_0 una aproximación inicial para el método, resolver el sistema original es equivalente a resolver $Az = r_0$

si $z = x_{true} - x_0$. Sea $p(x)$ el polinomio de menor grado i tal que $p(A)r_0 = 0$. Por el teorema de Cayley-Hamilton, tenemos que $i \leq n$. Entonces tenemos que:

$$p(A) = a_0 + Aa_1 + \cdots + a_{i-1}A^{i-1} + a_iA^i = 0$$

Por lo que vale que

$$A^{-1}a_0 = -(a_1I + a_2A + \cdots + a_{i-1}A^{i-2} + a_iA^{i-1})$$

Y si multiplicamos por r_0 tenemos que:

$$z = A^{-1}r_0 = -\frac{1}{a_0}(a_1I + a_2A + \cdots + a_{i-1}A^{i-2} + a_iA^{i-1})r_0$$

Por lo que

$$z \in \langle r_0, Ar_0, \dots, A^{i-1}r_0 \rangle = \mathcal{K}_m(A, r_0)$$

o bien,

$$x_{true} \in x_0 + \mathcal{K}_m(A, r_0).$$

Así nos podemos convencer que la formulación de un subespacio de Krylov para resolver sistemas lineales sale naturalmente de entender el problema.

1.1.3 Metodos de Proyección

Debido a la gran escala de las matrices y su mal condicionamiento, estos sistemas se suelen resolver a través de la proyección a subespacios de Krylov dimensión más pequeña.

En el apartado anterior, vimos que la solución x está contenida en el subespacio afín $x_0 + \mathcal{K}_m(A, r_0)$. Naturalmente, esto nos lleva a pensar que la solución x_n puede ser computada en el subespacio $x_0 + \mathcal{K}_n(A, r_0)$. Equivalentemente,

$$x_n = x_0 + y_n, \quad y_n \in \mathcal{K}_n(A, r_0)$$

para $n \in \mathbb{N}$. Nuestro objetivo es encontrar la mejor combinación lineal de coeficientes para el subespacio de Krylov, con el fin de que a partir de una cierta iteración m , se obtenga la solución exacta (o suficientemente aproximada).

Sea $x_0 \in \mathbb{C}^n$ un vector inicial y sea $r_0 := b - Ax_0$. El residuo correspondiente a la iteración n está dado por:

$$r_n := b - Ax_n = b - A(x_0 + y_n) = r_0 - Ay_n, \quad r_n \in \mathcal{K}_{n+1}(A, r_0)$$

Para poder determinar la solución x_n de manera óptima, se imponen ciertas restricciones, que suelen describir a partir de nociones de ortogonalidad. Estos son algunos de los enfoques:

Enfoque de Ritz-Galerkin

Para cada $n \in \mathbb{N}$, pedimos

$$r_n \perp \mathcal{K}_n(A, r_0)$$

Enfoque de Petrov-Galerkin

Para este enfoque, pedimos que el residuo (definido previamente) sea ortogonal a un subespacio $\mathcal{L} \subset \mathbb{C}^n$ de dimensión n , es decir,

$$r_n \perp \mathcal{L}_n, \text{ con } \dim(\mathcal{L}_n) = n$$

Proposición 1. *El enfoque de Ritz-Galerkin es el enfoque de Petrov-Galerkin con la elección $\mathcal{L}_n = \mathcal{K}_n(A, r_0)$.*

Para poder representar este problema matricialmente, consideramos $V = [v_1, \dots, v_m]$ y $W = [w_1, \dots, w_m]$, cuyas columnas definen bases de \mathcal{K} y \mathcal{L} respectivamente. La solución aproximada \tilde{x} ahora puede ser escrita como $\tilde{x} = x_0 + Vy$, $y \in \mathbb{C}^n$. Si definimos $r := b - A\tilde{x}$, para que éste sea ortogonal a \mathcal{L} pedimos:

$$W^T r = 0 \iff W^T(b - A\tilde{x}) = 0 \iff W^T(b - A(x_0 + Vy)) = 0 \iff \boxed{W^T AVy = W^T r_0}.$$

Si $W^T AV$ es inversible, entonces la aproximación está dada por:

$$\tilde{x} = x_0 + V(W^T AV)^{-1}W^T r_0$$

1.1.4 Algoritmo de Arnoldi

FOM

K

GMRES

K

Fausto implementa algo medio copado c: