Método de gradiente conjugado con precondicionamiento

Estos apuntes son casi una copia de los apuntes de Greg Fasshauer:

http://www.math.iit.edu/~fass/477577_Chapter_16.pdf

Objetivos. Estudiar la idea del método de gradiente conjugado precondicionado.

Requisitos. El método de gradiente conjugado. El número de condición de una matriz.

1. Introducción breve. Sea A una matriz simétrica positiva definida. Consideramos el sistema de ecuaciones $Ax = b$, donde b es algún vector dado. Se sabe que el método de gradiente conjugado para este sistema de ecuaciones converge mejor cuando la matriz A es *bien condicionada*, esto es, cuando su número condición no es muy grande. En algunos casos se puede construir una matriz positiva simétrica M tal que el producto $M^{-1}A$ es mucho mejor condicionado que la matriz original A.

2. Factorización de la matriz de precondicionamiento. Como M es simétrica y positiva definida, se puede encontrar una matriz L invertible tal que

$$M^{-1} = LL^\top.$$

Por ejemplo, para calcular L de manera explícita, uno puede primero calcular la factorización de Cholesky de la matriz M, y luego invertir el factor inferior izquierdo de la factorización de Cholesky. Es un procedimiento más costoso que resolver el sistema original $Ax = b$, y en algoritmo que vamos a deducir no se utilizará la matriz L en la forma explícita. Sin embargo, la matriz L sirve para entender el algoritmo.

3. Transformación del sistema. Transformamos el sistema $Ax = b$:

$$\begin{aligned} Ax = b \quad &\iff \quad M^{-1}Ax = M^{-1}b \quad \iff \quad LL^\top Ax = LL^\top b \\ &\iff \quad L^\top Ax = L^\top b \quad \iff \quad L^\top ALL^{-1}x = L^\top b. \end{aligned}$$

Utilizamos la siguiente notación:

$$\widehat{A} \coloneqq L^\top AL, \qquad \widehat{x} \coloneqq L^{-1}x, \qquad \widehat{b} \coloneqq L^\top b.$$

Hemos mostrado que el sistema original $Ax = b$ es equivalente al sistema $\widehat{A}\widehat{x} = \widehat{b}$. Es fácil verificar que la matriz $\widehat{A}$ es simétrica y positiva definida.

4. La matriz $\widehat{A}$ es simétrica y positiva definida. En efecto,

$$\widehat{A}^\top = L^\top A^\top (L^\top)^\top = L^\top A L = \widehat{A},$$

y si $v \in \mathbb{R}^n \setminus \{\mathbf{0}_n\}$, entonces $Lv \neq \mathbf{0}_n$ y

$$v^\top \widehat{A} v = (Lv)^\top A (Lv) > 0.$$

Apliquemos el método de gradiente conjugado a la matriz $\widehat{A}$, pero transformemos las fórmulas del método de tal manera que en vez de $\widehat{A}$ o L o $\widehat{b}$ se utilicen solamente los objetos originales A, M y b.

5. Duda. En la introducción pedimos que sea pequeño el número $\varkappa(M^{-1}A)$. No veo cómo relacionarlo con $\varkappa(\widehat{A})$.

Deducción de las fórmulas del algoritmo

Escribimos las fórmulas principales del método de gradiente conjugado aplicado a la matriz $\widehat{A}$ y el vector $\widehat{b}$. Luego transformamos estas fórmulas de tal manera que solamente se utilicen A, M y b.

6. Fórmula para el residuo. El residuo correspondiente al s-ésimo paso se puede escribir como

$$\widehat{r}_s = \widehat{b} - \widehat{A}\widehat{x}_s = L^\top b - (L^\top A L) L^{-1} x_s = L^\top r_s.$$

Denotemos por z_s al siguiente *residuo modificado*:

$$z_s := M^{-1} r_s.$$

7. La norma del residuo.

$$\|\widehat{r}_s\|^2 = (L^\top r_s)^\top (L^\top r_s) = r_s^\top L L^\top r_s = r_s^\top M^{-1} r_s = r_s^\top z_s. \tag{1}$$

8. Direcciones conjugadas. Denotamos por $\widehat{p}_s$ a las direcciones en el método de gradiente conjugado aplicado a la matriz $\widehat{A}$. Se sabe que los vectores $\widehat{p}_s$ son $\widehat{A}$-ortogonales. Pongamos

$$p_s := L\widehat{p}_s.$$

Notamos que los vectores p_s son A-ortogonales.

$$p_s^\top A p_t = \widehat{p}_s^\top L^\top A L \widehat{p}_t. \tag{2}$$

Si $s \neq t$, entonces este producto es 0.

9. Fórmula para la longitud del paso. Por (1) y (2),

$$\widehat{\alpha}_s = \frac{\widehat{r}_s^\top \widehat{r}_s}{\widehat{p}_s^\top \widehat{A} \widehat{p}_s} = \frac{r_s^\top z_s}{p_s^\top A p_s}.$$

10. Fórmula para el coeficiente de ortogonalización. Notamos que

$$\widehat{p}_s^\top \widehat{r}_{s+1} = p_s^\top (L^{-1})^\top L^\top r_s = p_s^\top r_s.$$

Por eso

$$\widehat{\beta}_{s+1} = \frac{\widehat{p}_s^\top \widehat{r}_{s+1}}{\widehat{p}_s^\top \widehat{A} \widehat{p}_s} = \frac{p_s^\top r_{s+1}}{p_s^\top A p_s}. \tag{3}$$

11. Renovación del vector.

$$\widehat{x}_{s+1} = \widehat{x}_s + \alpha_s \widehat{p}_s \quad \Longleftrightarrow \quad L^{-1} x_{s+1} = L^{-1} x_s + \alpha_s L^{-1} p_s \quad \Longleftrightarrow \quad x_{s+1} = x_s + \alpha_s p_s.$$

12. Renovación del residuo.

$$\begin{aligned} \widehat{r}_{s+1} = \widehat{r}_s - \alpha_s \widehat{A} \widehat{p}_s \quad &\Longleftrightarrow \quad L^\top r_{s+1} = L^\top r_s - \alpha_s L^\top A L\, L^{-1} p_s \\ &\Longleftrightarrow \quad r_{s+1} = r_s - \alpha_s A p_s. \end{aligned}$$

13. Renovación de la dirección.

$$\begin{aligned} \widehat{p}_{s+1} = \widehat{r}_{s+1} + \beta_{s+1} \widehat{p}_s \quad &\Longleftrightarrow \quad L^{-1} p_{s+1} = L^\top r_{s+1} + \beta_{s+1} L^{-1} p_s \\ &\Longleftrightarrow \quad p_{s+1} = z_{s+1} + \beta_{s+1} p_s. \end{aligned}$$

Algoritmo

14. Método de gradiente conjugado con precondicionamiento, en el lenguaje Matlab/Octave. Supongamos que está dada una función `solveM` que resuelve sistemas de ecuaciones lineales con la matriz M. Por ejemplo, si M es una matriz dispersa, entonces se puede usar la función estándar para resolver sistemas:

```
solveM = @ (w) M \ w;
```

Entonces el método de gradiente conjugado con el precondicionador M se puede programar así:

```
   function [x, s] = precondconjgrad(A, b, tol, smax, solveM),
   n = length(b);
   x = zeros(n, 1);
   r = b;
   z = solveM(r);
   p = z;
   delta = r' * p;
   nr = norm(r);
   s = 0;
   while (s < smax) && (nr >= tol),
      q = A * p;
      alpha = delta / (p' * q);
      x = x + alpha * p;
      rold = r;
      r = r - alpha * q;
      z = solveM(r);
      deltaold = delta;
      delta = r' * z;
      nr = norm(r);
      beta = delta / deltaold;
      p = z + beta * p;
      s = s + 1;
   end
end
```