Numerical Optimization 2024 - Homework 5

Aral Cimcim, k11720457

Artificial Intelligence, JKU Linz

April 22, 2024

Problem 1.

Implement Algorithm 5.2 and use to it solve linear systems in which A is the Hilbert matrix, whose elements are $A_{i,j} = \frac{1}{(i+j-1)}$. Set the right-hand-side to $b = (1, 1, \ldots, 1)^T$ and the initial point to $x_0 = 0$. Try dimensions $n = 5, 8, 12, 20$ and report the number of iterations required to reduce the residual below 10^{-6}.

Algorithm 5.2:

Given x_0;
Set $r_0 \leftarrow Ax_0 - b, p_0 \leftarrow -r_0, k \leftarrow 0$;
while $r_k \neq 0$ do

$$\alpha_k \leftarrow \frac{r_k^T r_k}{p_k^T A p_k}; \tag{1}$$

$$x_{k+1} \leftarrow x_k + \alpha_k p_k; \tag{2}$$

$$r_{k+1} \leftarrow r_k + \alpha_k A p_k; \tag{3}$$

$$\beta_{k+1} \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}; \tag{4}$$

$$p_{k+1} \leftarrow -r_{k+1} + \beta_{k+1} p_k; \tag{5}$$

$$k \leftarrow k + 1; \tag{6}$$

end (while)

Python implementation:

```
import matplotlib.pyplot as plt
import numpy as np
from scipy.linalg import hilbert

def conjugate_gradient(A, b, x_0, n, tol=1e-6, max_iters=10**6):
    r = A @ x_0 - b
    p = -r
    x_k = 0
    r_k = 0
    k = 0

    while k < max_iters and np.linalg.norm(r) > tol:
        alpha = (r.T @ r) / (p.T @ A @ p)
        x_k = x_0 + alpha * p
        r_k = r + alpha * A @ p
        beta = (r_k.T @ r_k) / (r.T @ r)
        p_k = -r_k + beta * p

        p = p_k
        r = r_k
        x_0 = x_k
        k += 1

    return x_k, r_k, k

dimensions = [5, 8, 12, 20]
iterations = []

for n in dimensions:
    A = hilbert(n)
    b = np.ones(n)
    x_0 = np.zeros(n)
    x, r, num_iter = conjugate_gradient(A, b, x_0, n)
    iterations.append(num_iter)

print(f'Num of iters: {iterations} | Corresponding dims: {dimensions}')
Number of iterations: [6, 19, 38, 67] | Corresponding dimensions: [5, 8, 12, 20]
```

The code returns the number of iterations 6, 19, 38, 67 to reduce the residual below 10^{-6} for dimensions $n = 5, 8, 12, 20$ respectively.

Problem 2.

Show that if the nonzero vectors $p_0, p_1, \dots, p_l$ satisfy (5.5), where A is symmetric and positive definite, then these vectors are linearly independent. (This result implies that A has at most n conjugate directions.)

(5.5) is given as:

$$p_i^T A p_j = 0 \quad \text{for} \quad i \neq j \tag{7}$$

If we suppose that there is a linear relation such that $S = \{p_0, p_1, \dots, p_l\}$:

$$\sum_{i=0}^{l} \alpha_i p_i = 0 \tag{8}$$

We can multiply the above equation with $p_j^T A$:

$$\sum_{i=0}^{l} \alpha_i p_j^T A p_i = \alpha_j p_j^T A p_j = 0 \tag{9}$$

Because matrix A is positive definite, $p_j^T A p_j > 0$ and $\alpha_j = 0$.

Problem 3.

Verify the formula (5.7): $\alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k}$

For the ideal quadratic function that we want to minimize (page 102):

$$\min \phi(\alpha) = \frac{1}{2} x^T A x - b^T x \tag{10}$$

If we take a step towards the direction p_k from x_k:

$$x_{k+1} = x_k + \alpha p_k \tag{11}$$

From this point we need to find the α value that minimizes $\phi(\alpha)$:

$$\phi(\alpha) = \phi(x_k + \alpha p_k) \tag{12}$$

We can rewrite the above equation as:

$$\phi(\alpha) = \frac{1}{2}(x_k + \alpha p_k)^T A (x_k + \alpha p_k) - (x_k + \alpha p_k) b^T \tag{13}$$

$$\phi(\alpha) = \frac{1}{2}(x_k^T + \alpha p_k^T)(A x_k + \alpha A p_k) - (x_k + \alpha p_k) b^T \tag{14}$$

$$\phi(\alpha) = \frac{1}{2}(x_k^T A x_k + \alpha x_k^T A p_k + \alpha p_k^T A x_k + \alpha^2 p_k^T A p_k) - b^T x_k - \alpha b^T p_k \tag{15}$$

$$\phi(\alpha) = \frac{1}{2} x_k^T A x_k + \frac{1}{2} \alpha x_k^T A p_k + \frac{1}{2} \alpha p_k^T A x_k + \frac{1}{2} \alpha^2 p_k^T A p_k - b^T x_k - \alpha b^T p_k \tag{16}$$

Since $x_k^T A p_k = p_k^T A x_k$ we get:

$$\phi(\alpha) = \frac{1}{2} x_k^T A x_k + \alpha x_k^T A p_k + \frac{1}{2} \alpha^2 p_k^T A p_k - b^T x_k - \alpha b^T p_k \tag{17}$$

We see that we can use $\phi(x_k) = \frac{1}{2} x_k^T A x_k - b^T x_k$:

$$\phi(\alpha) = \phi(x_k) + \frac{1}{2} \alpha^2 p_k^T A p_k + \alpha (p_k^T A x_k - p_k^T b)^T \tag{18}$$

The residual at x_k is defined as $r_k = A x_k - b$ (page 102)

$$\phi(\alpha) = \phi(x_k) + \frac{1}{2} \alpha^2 p_k^T A p_k + \alpha (r_k^T p_k) \tag{19}$$

Taking the derivative of $\phi(\alpha)$ with respect to α and setting it to 0:

$$\alpha p_k^T A p_k + r_k^T p_k = 0 \tag{20}$$

$$\alpha = -\frac{r_k^T p_k}{p_k^T A p_k} \tag{21}$$