TMA4180
Solutions to recommended exercises in Chapter 5 of N&W

Exercise 5.1

An Matlab implementation (`cg.m`) of CG applied to the problem $Ax = b$, where A is the Hilbert matrix is found on the wiki. The number of iterations needed to reach $\|r_k\| < 10^{-6}$ is reported below for different matrix dimensions n, along with the condition number $\kappa(A) = \frac{\lambda_n}{\lambda_1}$. We set $x_0 = (0, 0, \ldots, 0)^T$ and $b = (1, 1, \ldots, 1)^T$.

n	5	8	12	20
Iterations	6	19	39	76
$\kappa(A)$	476607	$1.5 \cdot 10^{10}$	$1.7 \cdot 10^{16}$	$2.0 \cdot 10^{18}$

We conclude that the required number of iterations increases fast with the matrix dimension. This can be explained by Equation (5.36) in N&W, since the condition number is growing *very* fast. For $n = 20$, Matlab's own linear solver (\) reported convergence problems due to badly scaled matrix.

Exercise 5.8

To create matrices with the desired eigenvalue distribution, we do as follows. First, create a random $n \times n$ matrix M, then apply QR factorization, so that $M = QR$, where Q is a unitary matrix ($Q^T Q = I$). Now let D be a $n \times n$ diagonal matrix with the wanted eigenvalues on the diagonal. Now, $A = QDQ^T$ is SPD and has the same eigenvalues as D. This algorithm is implemented in `cg.m`. Three different cases are considered:

1. *Clustered*: A has only 5 distinct eigenvalues: 1, 2, 10, 20 and 100.
2. *Perturbed*: A has the same eigenvalues, only slightly perturbed so that no are identical, but still in clusters.
3. *Uniform*: A has uniformly distributed eigenvalues between 1 and 100.

Observe that all realizations of A has the same condition number $\kappa(A) = 100$. For $n = 1000$, $x_0 = (0, 0, \ldots, 0)^T$ and $b = (1, 1, \ldots, 1)^T$, the error norm, $\|x^* - x\|_A$ is plotted against the iteration number in Figure 1. The exact solution x^* is obtained by using Matlab's own linear solver. We see that on iteration 5, the error goes to 0 for the clustered case, while for the perturbed case, we see a significant drop in the error on the 6th iteration. In the uniform case, the error norm decays close to linearly, and has very slow convergence. These observations are in agreement with the discussion on pages 116-117 of N&W.

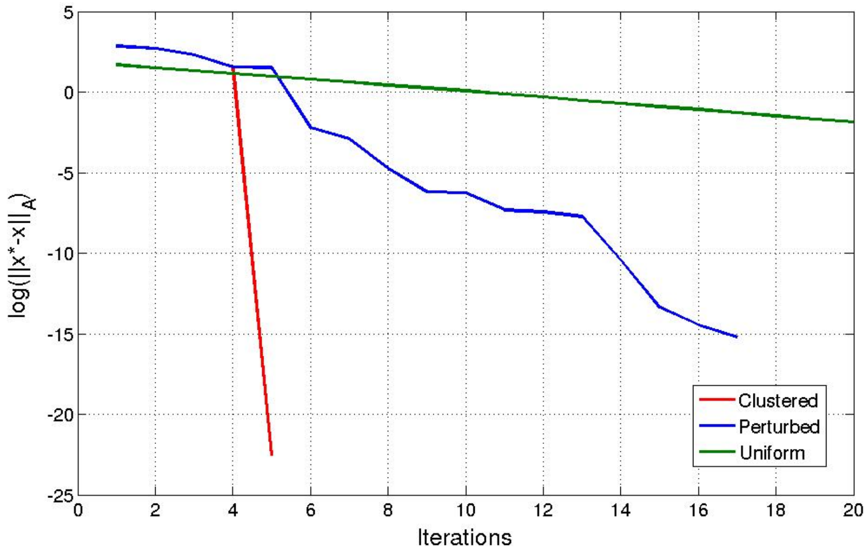

Figure 1: Error norm vs. iterations for CG applied to three different cases.

Exercise 5.9

We are asked to prove that CG equals PCG when doing the change of variables $\hat{x} = Cx$. The corresponding objective function is given as

$$\hat{\phi}(\hat{x}) = \frac{1}{2}\hat{x}^T(C^{-T}AC^{-1}\hat{x} - (C^{-T}b)^T\hat{x},$$

and the linear system of equations is

$$C^{-T}AC^{-1}\hat{x} = C^{-T}b.$$

If we apply CG (Algorithm 5.2 in N&W) on this system we get that

$$\hat{r}_0 = C^{-T}AC^{-1}Cx_0 - C^{-T}b = C^{-T}(Ax_0 - b), \tag{1}$$

$$\hat{p}_0 = -\hat{r}_0, \tag{2}$$

$$\hat{\alpha}_k = \frac{\hat{r}_k^T\hat{r}_k}{\hat{p}_k^T C^{-T}AC^{-1}\hat{p}_k}, \tag{3}$$

$$\hat{x}_{k+1} = \hat{x}_k + \hat{\alpha}_k\hat{p}_k, \tag{4}$$

$$\hat{r}_{k+1} = \hat{r}_k + \hat{\alpha}_k C^{-T}AC^{-1}\hat{p}_k, \tag{5}$$

$$\hat{\beta}_{k+1} = \frac{\hat{r}_{k+1}^T\hat{r}_{k+1}}{\hat{r}_k^T\hat{r}_k}, \tag{6}$$

$$\hat{p}_{k+1} = -\hat{r}_{k+1} + \hat{\beta}_{k+1}\hat{p}_k. \tag{7}$$

We now define $p_k := C^{-1}\hat{p}_k$ and $r_k := C^T\hat{r}_k$. This immediately gives $r_0 = Ax_0 - b$ and

$p_0 = -r_0$. By premultiplying (4) by C^{-1}, we get

$$x_{k+1} = x_k + \hat{\alpha}_k C^{-1} \hat{p}_k = x_k + \hat{\alpha}_k p_k.$$

Further, from (5),

$$r_{k+1} = r_k + \hat{\alpha}_k C^T C^{-T} A C^{-1} \hat{p}_k = r_k + \hat{\alpha}_k A p_k,$$

and from (7),

$$p_{k+1} = -C^{-1} C^{-T} r_{k+1} + \hat{\beta}_{k+1} p_k = -y_{k+1} + \hat{\beta}_{k+1} p_k,$$

where $M = C^T C$ and $M y_k = r_k$. Finally, from (3) and (6), we get that

$$\alpha_k := \hat{\alpha}_k = \frac{(C^{-T} r_k)^T (C^{-T} r_k)}{(C p_k)^T C^{-T} A C^{-1} (C p_k)} = \frac{r_k^T y_k}{p_k^T A p_k},$$
$$\beta_k := \hat{\beta}_k = \frac{(C^{-T} r_{k+1})^T (C^{-T} r_{k+1})}{(C^{-T} r_k)^T (C^{-T} r_k)} = \frac{r_{k+1}^T y_{k+1}}{r_k^T y_k}.$$

All together, we have arrived at the PCG method (Algorithm 5.3 in N&W).

Exercise 5.11

For a quadratic function $f(x) = \frac{1}{2} x^T A x - b^T x$, we have $\nabla f_k = r_k = A x_k - b$. For the nonlinear conjugate gradient methods, we have

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k p_k, \\ p_{k+1} &= -r_{k+1} + \beta_k p_k, \end{aligned} \tag{8}$$

where β_k could be one of the following:

$$\beta_{k+1}^{FR} = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k} = \frac{\|r_{k+1}\|^2}{\|r_k\|^2},$$
$$\beta_{k+1}^{PR} = \frac{r_{k+1}^T (r_{k+1} - r_k)}{\|r_k\|^2} = \frac{\|r_{k+1}\|^2 - r_{k+1}^T r_k}{\|r_k\|^2},$$
$$\beta_{k+1}^{HS} = \frac{r_{k+1}^T (r_{k+1} - r_k)}{(r_{k+1} - r_k)^T p_k}.$$

For an exact line search, we have

$$\alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k}.$$

For this problem, we want to show that $\beta_{k+1}^{FR} = \beta_{k+1}^{PR} = \beta_{k+1}^{HS}$. We start by observing that

$$r_{k+1}^T p_k = (A x_{k+1} - b)^T p_k = x_k^T A p_k - \frac{r_k^T p_k}{p_k^T A p_k} p_k^T A p_k - b^T p_k = 0. \tag{9}$$

Further,

$$r_k^T p_k = r_k^T(-r_k + \beta_k p_{k-1}) = -r_k^T r_k, \tag{10}$$

from which it immediately follows that $\beta_{k+1}^{PR} = \beta_{k+1}^{HS}$.

It remains to show that $r_{k+1}^T r_k = 0$, since this implies that $\beta_{k+1}^{FR} = \beta_{k+1}^{PR}$. We know that $r_{k+1} = r_k + \alpha_k A p_k$, thus by (9) and (10),

$$p_k^T A p_k = p_k^T(\alpha_k^{-1}(r_{k+1} - r_k)) = \alpha_k^{-1} r_k^T r_k.$$

Now,

$$\begin{aligned} p_{k+1}^T A p_k &= (-r_{k+1} + \beta_{k+1}^{PR} p_k)^T A p_k \\ &= -r_{k+1}^T A p_k + \beta_{k+1}^{PR} p_k^T A p_k \\ &= -\alpha_k^{-1} r_{k+1}^T (r_{k+1} - r_k) + \beta_{k+1}^{PR} \alpha_k^{-1} r_k^T r_k = 0, \end{aligned}$$

and hence p_k are conjugate. Finally, it follows from Theorem 5.2 in N&W and (8) that

$$r_{k+1}^T r_k = r_{k+1}^T(\beta_k p_{k-1} - p_k) = 0.$$

Exercise 5.12

This proof follows the same steps as the original proof. The inequalities we want to prove are

$$-\frac{1}{1-c_2} \leq \frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2} \leq \frac{2c_2 - 1}{1 - c_2}. \tag{11}$$

We use induction, and by the same arguments as in N&W, (11) hold for $k = 0$. Next, assume (11) hold for some $k \geq 1$. From (5.41b) in N&W we have that

$$\frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} = -1 + \beta_{k+1} \frac{\nabla f_{k+1}^T p_k}{\|\nabla f_{k+1}\|^2}.$$

Thus, we can establish

$$-1 - |\beta_{k+1}| \frac{|\nabla f_{k+1}^T p_k|}{\|\nabla f_{k+1}\|^2} \leq \frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} \leq -1 + |\beta_{k+1}| \frac{|\nabla f_{k+1}^T p_k|}{\|\nabla f_{k+1}\|^2}. \tag{12}$$

By using the strong Wolfe condition

$$|\nabla f_{k+1}^T p_k| \leq -c_2 \nabla f_k^T p_k$$

and the restriction

$$|\beta_k| \leq \beta_k^{FR} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k},$$

we see that

$$-1 - |\beta_{k+1}| \frac{|\nabla f_{k+1}^T p_k|}{\|\nabla f_{k+1}\|^2} \geq -1 + \beta_{k+1}^{FR} \frac{c_2 \nabla f_k^T p_k}{\|\nabla f_{k+1}\|^2} = -1 + \frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2}. \tag{13}$$

Equivalently,

$$-1 + |\beta_{k+1}| \frac{|\nabla f_{k+1}^T p_k|}{\|\nabla f_{k+1}\|^2} \leq -1 - \beta_{k+1}^{FR} \frac{c_2 \nabla f_k^T p_k}{\|\nabla f_{k+1}\|^2} = -1 - \frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2}. \tag{14}$$

Combining (13) and (14) into (12), we have that

$$-1 + \frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2} \leq \frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} \leq -1 - \frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2}.$$

This is equal to an intermediate result in the proof of N&W. By substituting $\frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2}$ from the left-hand-side of the induction hypothesis (11), we obtain (11) for $k+1$.