Exercise 4

February 06, 2024

Problem 1.

Let

$$A = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 2 \end{pmatrix} \qquad \text{and} \qquad b = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Use the CG-method with initialization $x_0 = 0$ for solving the linear system $Ax = b$.

Solution.

By applying Algorithm 5.2 with the initial value $x_0 = (0, 0, 0)$, we obtain the following:

$$r_0 = (-1, 0, -1),\ p_0 = (1, 0, 1),\ \alpha_0 = 1,$$

$$x_1 = (1, 0, 1),\ r_1 = (0, -2, 0),\ \beta_1 = 2,\ p_1 = (2, 2, 2),\ \alpha_1 = 1,$$

$$x_2 = (3, 2, 3),\ r_2 = (0, 0, 0).$$

Since $r_2 = 0$, we stop the process here and conclude that $x = (3, 2, 3)$ solves the liner system $Ax = b$ in two iterative steps.

Problem 2.

Implement the linear CG method (N&W, p. 112, Algorithm 5.2) and use it to 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}.

Solution.

See the file `CG_Hilbert.py` on the wiki-page.

The number of iterations needed are as follows:

n	5	8	12	20
iterations	6	19	38	70

That is, the CG method always requires more iterations than predicted by the theory. The reason for this discrepancy is the fact that the Hilbert matrix is extremely ill-conditioned (and because of that often used for numerical

tests). For $n = 5$ for instance, the linalg package in numpy computes the condition of the Hilbert matrix as approximately 476607; for $n = 20$, it computes it at approximately 10^{18}. (Because the matrix is that ill-conditioned, these values should only be taken as rough approximations to the true condition numbers; the ill-conditioning of the matrix makes it also difficult to compute its condition number.)

Problem 3.

Implement a non-linear CG-method (see N&W, Algorithm 5.4). Test it on the *Rosenbrock function* $f\colon \mathbb{R}^2 \to \mathbb{R}$, $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$. Use the strong Wolfe conditions with $0 < c_1 < c_2 < 1/2$ for step length selection.

Hint: You can use a bracketing method as discussed in the lecture for finding a suitable step length. In this case, you have to recall, though, that the two conditions

$$\begin{aligned} f(x_k + \alpha p_k) &\leq f(x_k) + c_1 \alpha \langle \nabla f(x_k), p_k \rangle, \\ \langle \nabla f(x_k + \alpha p_k), p_k \rangle &\leq -c_2 \langle \nabla f(x_k), p_k \rangle, \end{aligned}$$

indicate that a step length α is not too large (that is, if one of those is violated, then the current step length *is* too large), while only the weak curvature condition

$$\langle \nabla f(x_k + \alpha p_k), p_k \rangle \geq c_2 \langle \nabla f(x_k), p_k \rangle$$

prevents too small step lengths α.

Alternatively, you can use the algorithm that is described in N&W, Section 3.5.

Solution.

See possible solutions on the wiki page.

Problem 4.

Let $Q \in \mathbb{R}^{d \times d}$ by symmetric and positive definite and let $b \in \mathbb{R}^d$. Define $\Phi\colon \mathbb{R}^d \to \mathbb{R}$,

$$\Phi(x) = \frac{1}{2}\langle x, Qx \rangle - \langle x, b \rangle.$$

Consider the minimisation of Φ using a non-linear CG method *with exact line search*, where the search directions are chosen either with the Polak–Ribière formula (N&W, eq. (5.44)), the Hestenes–Stiefel formula (N&W, eq. (5.46)), or the Fletcher–Reeves formula (N&W, eq. (5.41a)). Show that all of these methods reduce to the CG-method for linear systems.

Hint: Use induction to show that $\langle r_{k+1}, p_k\rangle = 0$, $\langle r_k, p_k\rangle = -\|r_k\|^2$ and $\langle r_{k+1}, r_k\rangle = 0$ for all k.

Solution.

For a quadratic function $f(x) = \frac{1}{2}\langle x, Qx\rangle - \langle b, x\rangle$ with symmetric and positive definite matrix Q, we have $\nabla f_k = r_k = Qx_k - b$. For the nonlinear conjugate gradient methods, we have

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

where β_k could be one of the following:

$$\beta_{k+1}^{\mathrm{FR}} = \frac{\langle r_{k+1}, r_{k+1}\rangle}{\langle r_k, r_k\rangle}, \quad \beta_{k+1}^{\mathrm{PR}} = \frac{\langle r_{k+1}, r_{k+1} - r_k\rangle}{\langle r_k, r_k\rangle}, \quad \beta_{k+1}^{\mathrm{HS}} = \frac{\langle r_{k+1}, r_{k+1} - r_k\rangle}{\langle r_{k+1} - r_k, p_k\rangle}.$$

Here $\beta_{k+1}^{\mathrm{FR}}$ is the same parameter as used in the standard CG-method for linear systems. For an exact line search, we have

$$\alpha_k = -\frac{\langle r_k, p_k\rangle}{\langle p_k, Qp_k\rangle}.$$

Now, we have to show that $\beta_{k+1}^{\mathrm{FR}} = \beta_{k+1}^{\mathrm{PR}} = \beta_{k+1}^{\mathrm{HS}}$. We start with proving

$$\begin{aligned} \langle r_{k+1}, p_k\rangle &= \langle Qx_{k+1} - b, p_k\rangle \\ &= \langle Qx_k + \alpha_k Qp_k - b, p_k\rangle \\ &= \langle x_k, Qp_k\rangle - \frac{\langle r_k, p_k\rangle}{\langle p_k, Qp_k\rangle}\langle p_k, Qp_k\rangle - \langle b, p_k\rangle \\ &= \langle Qx_k - b, p_k\rangle - \langle r_k, p_k\rangle \\ &= \langle r_k, p_k\rangle - \langle r_k, p_k\rangle \\ &= 0. \end{aligned} \tag{2}$$

Further,

$$\begin{aligned} \langle r_k, p_k\rangle &= \langle r_k, -r_k + \beta_k p_{k-1}\rangle \text{ (by (1))} \\ &= -\langle r_k, r_k\rangle + \beta_k \langle r_k, p_{k-1}\rangle \\ &= -\langle r_k, r_k\rangle + 0 \text{ (by (2))} \\ &= -\langle r_k, r_k\rangle \end{aligned} \tag{3}$$

By using (2) and (3), we get

$$\beta_{k+1}^{\mathrm{HS}} = \frac{\langle r_{k+1}, r_{k+1} - r_k\rangle}{\langle r_{k+1} - r_k, p_k\rangle} = \frac{\langle r_{k+1}, r_{k+1} - r_k\rangle}{\langle r_k, r_k\rangle} = \beta_{k+1}^{\mathrm{PR}}. \tag{4}$$

Now, if we show that $\langle r_{k+1}, r_k \rangle = 0$, we get our desired conclusion. For $k = 0$, we have that $r_0^{\text{PR}} = r_0^{\text{FR}}$ because we use the same initialization x_0 for both PR and FR methods. Further, for $k \to k+1$, the first k steps of both methods follow the linear CG algorithm. Moreover, $x_{k+1}^{\text{PR}} = x_{k+1}^{\text{FR}}$ because α is defined same for both methods, which implies $r_{k+1}^{\text{PR}} = r_{k+1}^{\text{FR}}$. Then by Theorem 5.3 (N&W, inequality 5.16), we have $\langle r_{k+1}, r_k \rangle = 0$.

Thus,

$$\beta_{k+1}^{\text{FR}} = \beta_{k+1}^{\text{PR}}. \tag{5}$$

The expressions (4) and (5) render the conclusion that

$$\beta_{k+1}^{\text{FR}} = \beta_{k+1}^{\text{PR}} = \beta_{k+1}^{\text{HS}}.$$

□

Problem 5.

Consider a non-linear CG-method defined by

$$p_k = -\nabla f(x_k) + \beta_{k-1} p_{k-1},$$

where $\beta_{k-1} \in \mathbb{R}$ satisfies the estimate

$$|\beta_{k-1}| \leq \beta_{k-1}^{\text{FR}} := \frac{\|\nabla f(x_k)\|^2}{\|\nabla f(x_{k-1})\|^2}$$

and the step length $\alpha_k > 0$ satisfies the strong Wolfe conditions with $0 < c_1 < c_2 < 1/2$.

Show that

$$-\frac{1}{1-c_2} \leq \frac{\langle \nabla f(x_k), p_k \rangle}{\|\nabla f(x_k)\|^2} \leq \frac{2c_2 - 1}{1 - c_2}$$

for all k.

Hint: Cf. the proof of Lemma 5.6.

Solution.

Note first that the function $t(\epsilon) = \frac{2\epsilon - 1}{1 - \epsilon}$ is monotonically increasing on the interval $\left[0, \frac{1}{2}\right]$ and that $t(0) = -1$ and $t(\frac{1}{2}) = 0$. Hence, because of $c_2 \in \left(0, \frac{1}{2}\right)$, we have

$$-1 < \frac{2c_2 - 1}{1 - c_2} < 0. \tag{6}$$

Then the descent direction condition $\langle \nabla f_k, p_k \rangle < 0$ follows immediately once we establish the following (inequality 5.53 of Lemma 5.6 (N&W Book))

$$-\frac{1}{1-c_2} \leq \frac{\langle \nabla f_k, p_k \rangle}{\|\nabla f_k\|^2} \leq \frac{2c_2 - 1}{1 - c_2}, \ \forall\ k = 0, 1, 2, \ldots. \tag{7}$$

In order to prove this, we use induction over k.

Base case: for $k = 0$, the middle term of inequality (7) is -1, so by using (6), we see that both inequalities in (7) are satisfied.

Induction step: we assume that (7) holds for some $k \geq 0$. From $p_{k+1} = -\nabla f_{k+1} + \beta_{k+1} p_k$ (expression (5.41b) of Algorithm 5.4, N&W), we have

$$\frac{\langle \nabla f_{k+1}, p_{k+1} \rangle}{\|\nabla f_{k+1}\|^2} = -1 + \beta_{k+1} \frac{\langle \nabla f_{k+1}, p_k \rangle}{\|\nabla f_{k+1}\|^2}. \tag{8}$$

Thus, we can establish

$$-1 - |\beta_{k+1}| \frac{|\langle \nabla f_{k+1}, p_k \rangle|}{\|\nabla f_{k+1}\|^2} \leq -1 + \beta_{k+1} \frac{\langle \nabla f_{k+1}, p_k \rangle}{\|\nabla f_{k+1}\|^2} \leq -1 + |\beta_{k+1}| \frac{|\langle \nabla f_{k+1}, p_k \rangle|}{\|\nabla f_{k+1}\|^2}. \tag{9}$$

By using the condition $|\beta_k| \leq \beta_k^{\mathrm{FR}}$, we get that

$$|\beta_{k+1}| \leq \beta_{k+1}^{\mathrm{FR}} = \frac{\langle \nabla f_{k+1}, \nabla f_{k+1} \rangle}{\langle \nabla f_k, \nabla f_k \rangle}.$$

Therefore,

$$\frac{|\beta_{k+1}||\langle \nabla f_{k+1}, p_k \rangle|}{\|\nabla f_{k+1}\|^2} \leq \frac{|\langle \nabla f_{k+1}, p_k \rangle|}{\|\nabla f_k\|^2}. \tag{10}$$

The strong Wolfe condition gives

$$|\langle \nabla f_{k+1}, p_k \rangle| \leq -c_2 \langle \nabla f_k, p_k \rangle. \tag{11}$$

Inequalities (10) and (11) yield

$$\frac{|\beta_{k+1}||\langle \nabla f_{k+1}, p_k \rangle|}{\|\nabla f_{k+1}\|^2} \leq -\frac{c_2 \langle \nabla f_k, p_k \rangle}{\|\nabla f_k\|^2}. \tag{12}$$

Inequality (12) can be rewritten into the following two forms:

$$-1 + \frac{|\beta_{k+1}||\langle \nabla f_{k+1}, p_k \rangle|}{\|\nabla f_{k+1}\|^2} \leq -1 - \frac{c_2 \langle \nabla f_k, p_k \rangle}{\|\nabla f_k\|^2} \tag{13}$$

and

$$-1 - \frac{|\beta_{k+1}||\langle \nabla f_{k+1}, p_k \rangle|}{\|\nabla f_{k+1}\|^2} \geq -1 + \frac{c_2 \langle \nabla f_k, p_k \rangle}{\|\nabla f_k\|^2}. \tag{14}$$

By combining inequalities (8), (9), (13) and (14), we obtain

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

Substituting for the term $\frac{\langle \nabla f_k, p_k \rangle}{\|\nabla f_k\|^2}$ from the left-hand-side of the induction hypothesis (7), we obtain

$$-1 - \frac{c_2}{1-c_2} \leq \frac{\langle \nabla f_{k+1}, p_{k+1} \rangle}{\|\nabla f_{k+1}\|^2} \leq -1 + \frac{c_2}{1-c_2}.$$

Therefore,

$$-\frac{1}{1-c_2} \leq \frac{\langle \nabla f_{k+1}, p_{k+1} \rangle}{\|\nabla f_{k+1}\|^2} \leq \frac{2c_2 - 1}{1-c_2},$$

which shows that (7) holds for $k+1$ as well. This proves the induction step, and thus it follows that (7) is true for all $k = 0, 1, 2, \ldots$. □