STAT 310: MATHEMATICAL COMPUTATIONS II
WINTER 2012
PROBLEM SET 4

Let $A \in \mathbb{R}^{n\times n}$ be symmetric positive definite and $\mathbf{b} \in \mathbb{R}^n$. As usual, we write

$$\mathbf{r}_k = \mathbf{b} - A\mathbf{x}_k. \tag{0.1}$$

We assume that $\mathbf{x}_0$ is initialized in some manner. In the lectures we assumed $\mathbf{x}_0 = \mathbf{0}$ and so $\mathbf{r}_0 = \mathbf{b}$ but we will do it a little more generally here.

1. Consider the quadratic functional

$$\varphi(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} - 2\mathbf{b}^\top \mathbf{x}.$$

(a) Show that

$$\nabla\varphi(\mathbf{x}_k) = -2\mathbf{r}_k$$

and hence if $\mathbf{x}_* \in \mathbb{R}^n$ is a stationary point of φ, then

$$A\mathbf{x}_* = \mathbf{b}.$$

Show also that $\mathbf{x}_*$ must be a minimizer of φ.

(b) Consider an iterative method

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k \tag{1.2}$$

where $\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \dots$ are search directions to be chosen later. Show that if we want α_k so that the function $f : \mathbb{R} \to \mathbb{R}$,

$$f(\alpha) = \varphi(\mathbf{x}_k + \alpha\mathbf{p}_k)$$

is minimized, then we must have

$$\alpha_k = \frac{\mathbf{r}_k^\top \mathbf{p}_k}{\mathbf{p}_k^\top A\mathbf{p}_k}. \tag{1.3}$$

(c) Deduce that

$$\varphi(\mathbf{x}_{k+1}) - \varphi(\mathbf{x}_k) = -\frac{(\mathbf{r}_k^\top \mathbf{p}_k)^2}{\mathbf{p}_k^\top A\mathbf{p}_k}$$

and therefore $\varphi(\mathbf{x}_{k+1}) < \varphi(\mathbf{x}_k)$ as long as $\mathbf{r}_k^\top \mathbf{p}_k \neq 0$.

2. Notations here follow those in Problem **1**.

(a) Show that if we choose

$$\mathbf{p}_k = \mathbf{r}_k, \tag{2.4}$$

we obtain the steepest decent method discussed in the lectures.

(b) Let the eigenvalues of A be $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ and $P \in \mathbb{R}[t]$. Show that

$$\|P(A)\mathbf{x}\|_A \leq \max_{1\leq i\leq n} |P(\lambda_i)| \|\mathbf{x}\|_A$$

for every $\mathbf{x} \in \mathbb{R}^n$. [*Hint:* $A \succ 0$ and so has an eigenbasis].

Date: February 10, 2012 (Version 1.1); due: February 16, 2012.

(c) Using (b) and $P_\alpha(t) = 1 - \alpha t$, show that
$$\|\mathbf{x}_k - \mathbf{x}_*\|_A \le \max_{1\le i\le n} |P_\alpha(\lambda_i)| \|\mathbf{x}_{k-1} - \mathbf{x}_*\|_A$$
for all $\alpha \in \mathbb{R}$.

(d) Using properties of Chebyshev polynomials, show that
$$\min_{\alpha\in\mathbb{R}} \max_{\lambda_n\le t\le\lambda_1} |1 - \alpha t| = \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}$$
and hence deduce that
$$\|\mathbf{x}_k - \mathbf{x}_*\|_A \le \left(\frac{\kappa - 1}{\kappa + 1}\right)^k \|\mathbf{x}_0 - \mathbf{x}_*\|_A$$
where $\kappa = \lambda_1/\lambda_n$ is the condition number of A.

3. Notations here follow those in Problem **1**. As in Problem **2**, we set
$$\mathbf{p}_0 = \mathbf{r}_0, \quad \alpha_0 = \frac{\mathbf{r}_0^\top \mathbf{p}_0}{\mathbf{p}_0^\top A\mathbf{p}_0}, \quad \mathbf{x}_1 = \mathbf{x}_0 + \alpha_0\mathbf{p}_0, \quad \mathbf{r}_1 = \mathbf{b} - A\mathbf{x}_1,$$
but we will no longer require (2.4) for $k \ge 1$. Instead for each $k \in \mathbb{N}$, we would like to get a 2-dimensional search space
$$\Pi_k = \{\mathbf{x}_k + \xi\mathbf{r}_k + \eta\mathbf{p}_{k-1} \in \mathbb{R}^n \mid \xi, \eta \in \mathbb{R}\}$$
and determine our next search direction $\mathbf{p}_k$ from this affine plane.

(a) Show that if we want $(\xi_k, \eta_k) \in \mathbb{R}^2$ so that the function $f : \mathbb{R}^2 \to \mathbb{R}$,
$$f(\xi, \eta) = \varphi(\mathbf{x}_k + \xi\mathbf{r}_k + \eta\mathbf{p}_{k-1})$$
is minimized, then we must have
$$\begin{cases} \xi_k\mathbf{r}_k^\top A\mathbf{r}_k + \eta_k\mathbf{r}_k^\top A\mathbf{p}_{k-1} = \mathbf{r}_k^\top\mathbf{r}_k, \\ \xi_k\mathbf{r}_k^\top A\mathbf{p}_{k-1} + \eta_k\mathbf{p}_{k-1}^\top A\mathbf{p}_{k-1} = \mathbf{0}. \end{cases} \tag{3.5}$$

(b) Show that as long as $\mathbf{r}_k \ne \mathbf{0}$, ξ_k in (3.5) can always be chosen to be non-zero and therefore
$$\mathbf{p}_k := \mathbf{r}_k + \frac{\eta_k}{\xi_k}\mathbf{p}_{k-1} \in \Pi_k \tag{3.6}$$
is always well-defined and that this choice of $\mathbf{p}_k$ gives the best descent direction in Π_k.

(c) Since (3.6) is only dependent on the ratio η_k/ξ_k and not on the values of η_k and ξ_k, we will let $\beta_{k-1} = \eta_k/\xi_k$. Now combine this with (0.1), (1.2), (1.3) and (3.6) to get an iterative method:
$$\begin{aligned} \alpha_k &= \frac{\mathbf{r}_k^\top\mathbf{p}_k}{\mathbf{p}_k^\top A\mathbf{p}_k}, & \mathbf{x}_{k+1} &= \mathbf{x}_k + \alpha_k\mathbf{p}_k, \\ & & \mathbf{r}_{k+1} &= \mathbf{b} - A\mathbf{x}_{k+1}, \\ \beta_k &= -\frac{\mathbf{r}_{k+1}^\top A\mathbf{p}_k}{\mathbf{p}_k^\top A\mathbf{p}_k}, & \mathbf{p}_{k+1} &= \mathbf{r}_{k+1} + \beta_k\mathbf{p}_k. \end{aligned} \tag{3.7}$$
Show that despite what it appears, (3.7) can be achieved with just one matrix-vector multiply per iteration and that it is identical to the conjugate gradient method derived in the lectures.

4. Notations here follow those in Problem **3**.

(a) Show that the residuals and search directions have the following properties.
   (i) $\mathbf{p}_i^\top\mathbf{r}_j = 0$ for all $i < j$;
   (ii) $\mathbf{r}_i^\top\mathbf{r}_j = 0$ for all $i \ne j$;
   (iii) $\mathbf{p}_i^\top A\mathbf{p}_j = 0$ for all $i \ne j$;

(iv) $\operatorname{span}\{\mathbf{r}_0, \dots, \mathbf{r}_k\} = \operatorname{span}\{\mathbf{p}_0, \dots, \mathbf{p}_k\} = \operatorname{span}\{\mathbf{r}_0, A\mathbf{r}_0, \dots, A^k\mathbf{r}_0\} =: K_{k+1}(A, \mathbf{r}_0)$ for all $k \in \mathbb{N}$.

(b) Show that $\mathbf{x}_k$ satisfies

$$\varphi(\mathbf{x}_k) = \min\{\varphi(\mathbf{x}) \mid \mathbf{x} \in \mathbf{x}_0 + K_k(A, \mathbf{r}_0)\}$$

and

$$\|\mathbf{x}_k - \mathbf{x}_*\|_A = \min\{\|\mathbf{x} - \mathbf{x}_*\|_A \mid \mathbf{x} \in \mathbf{x}_0 + K_k(A, \mathbf{r}_0)\}$$

where $\mathbf{x}_0 + S := \{\mathbf{x}_0 + \mathbf{y} \mid \mathbf{y} \in S\}$.

(c) Show that every $\mathbf{x} \in \mathbf{x}_0 + K_k(A, \mathbf{r}_0)$ satisfies

$$\mathbf{x} - \mathbf{x}_* = A^{-1}P(A)\mathbf{r}_0$$

for some $P \in \mathbb{R}[t]$ with $\deg(P) \le k$ and $P(0) = 1$. Deduce that

$$\|\mathbf{x}_k - \mathbf{x}_*\|_A \le \min_P \max_{\lambda_n \le t \le \lambda_1} |P(t)| \|\mathbf{x}_0 - \mathbf{x}_*\|_A$$

where the minimum is taken over all $P \in \mathbb{R}[t]$, $\deg(P) \le k$, $P(0) = 1$.

(d) Show that the solution to the minimax problem above is given by

$$P_k(t) = \frac{T_k\left(\dfrac{b + a - 2t}{b - a}\right)}{T_k\left(\dfrac{b + a}{b - a}\right)}$$

where $\lambda_n = a$ and $\lambda_1 = b$. Deduce that

$$\|\mathbf{x}_k - \mathbf{x}_*\|_A \le 2\left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^k \|\mathbf{x}_k - \mathbf{x}_*\|_A$$

where $\kappa = \lambda_1/\lambda_n$ is the condition number of A.

5. This last problem is for Part II of the course.

(a) Implement Newton's method ((3.38) in Nocedal and Wright, or from lecture notes).

(b) Apply the method to the following function (Fenton's function); which you can download from Mihai's website, from lecture list area, including with links to the automatic differentiation package discussed in class — if you wish to go this route. If not you may have to compute the gradient and Hessian by hand. It is also an easy function to code in AMPL if you want to test your work — but this is not required).

$$f(x_1, x_2) = \frac{1}{10}\left[12 + x_1^2 + \frac{1 + x_2^2}{x_1^2} + \frac{x_1^2 x_2^2 + 100}{x_1^4 x_2^4}\right].$$

(c) Initialize the method at $\mathbf{x} = (3, 2)$. Describe what you observe.

(d) Initialize the method at $\mathbf{x} = (3, 4)$. Describe what you observe.

(e) For the cases where the method converges, proposed a way to estimate the order of convergence numerically, and compare to what we discussed theoretically in class.