This assignment is due on Friday, 11 April 2014, a week after classes end. Since we won't have a class on that day, you can either

- drop the assignment off at my office, BA 4228,
- leave it for me in my mailbox in BA 4239,
- give it to Lynda Barnes, the receptionist in BA 4283, and ask her to leave it for me in my mailbox,
- email it to me at krj@cs.toronto.edu.

1. [10 marks]

   Do question 5.1 on page 133 of your textbook.

   Also discuss whether your numerical results are consistent with the theory for the convergence of the conjugate gradient method developed on pages 112–118 of your textbook.

   If you do this questions in MatLab, you can use the MatLab function `hilb` to construct the Hilbert matrix and the MatLab functions `eig` and `cond` to compute the eigenvalues and condition number, respectively, of the Hilbert matrix.

2. [10 marks]

   Do question 5.9 on page 134 of your textbook.

3. [10 marks]

   Suppose that a minimization algorithm for the problem

   $$\min_{x \in \mathbb{R}^n} f(x) \tag{1}$$

   generates a sequence of points x_0, x_1, $x_2, \ldots$ As we noted a few times this term, your textbook says that the algorithm is *globally convergent* if

   $$\lim_{k \to \infty} \|\nabla f(x_k)\| = 0 \tag{2}$$

   or possibly even

   $$\liminf_{k \to \infty} \|\nabla f(x_k)\| = 0 \tag{3}$$

   See, for example, the third line from the top of page 40 and Theorem 4.5 on page 80.

   The condition (3) means that, for any $\epsilon > 0$ and any $K \geq 0$, there is a $k \geq K$ such that $\|\nabla f(x_k)\| < \epsilon$.

As I've mentioned a few times in class, both (2) and (3) are weaker than

$$\lim_{k \to \infty} x_k = x^* \tag{4}$$

where x^* is a local minimizer of (1), which is what we usually mean by a globally convergent algorithm. By weaker here I mean that we can have a sequence $x_0, x_1, x_2, \ldots$ that satisfies either (2) or (3), but does not satisfy (4). For example, for the function $f(x) = \mathrm{e}^{-x}$ and the sequence $x_k = k$, for $k = 0, 1, 2, \ldots$, both (2) and (3) are satisfied, but (4) is not satisfied, since x_k does not converge to any real value.

However, we can sometimes get the stronger result (4) from either (2) or (3). As a case in point, prove the following theorem.

Theorem If

(a) $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable (i.e., $\nabla^2 f(x)$ exists and is continuous for all $x \in \mathbb{R}^n$),

(b) the level set $\mathcal{L} = \{x \in \mathbb{R}^n : f(x) \le f(x_0)\}$ is convex and there exist positive constants m and M such that

$$m\|z\|_2^2 \le z^T \nabla^2 f(x)\, z \le M\|z\|_2^2$$

for all $z \in \mathbb{R}^n$ and all $x \in \mathcal{L}$,

(c) the sequence of points $x_0, x_1, x_2, \ldots$ generated by the algorithm satisfies $f(x_{k+1}) \le f(x_k)$ for all $k = 0, 1, 2, \ldots$,

(d) the sequence of points $x_0, x_1, x_2, \ldots$ generated by the algorithm also satisfies

$$\liminf_{k \to \infty} \|\nabla f(x_k)\| = 0$$

then f has a unique strict minimizer $x^* \in \mathcal{L}$ and

$$\lim_{k \to \infty} x_k = x^*$$

By f having a unique strict minimizer $x^* \in \mathcal{L}$, we mean that there exists a point $x^* \in \mathcal{L}$ that satisfies $f(x^*) < f(x)$ for all $x \in \mathcal{L}$ for which $x \ne x^*$.

This is quite a difficult result to prove. If you find there are pieces of the proof that you can't actually prove yourself (such as, for example, the existence of the unique strict minimizer x^*), clearly state the result you need as an unproven assumption.

Of course, the fewer of these unproven assumptions you need in your proof, the better your proof will be.