CS412 Spring Semester 2012

Homework Assignment #4

Due Thursday May 3rd 2012, in class

Sum of all problems : 120%, Maximum possible score : 100%.

1. [60%] In this problem you will construct a proof that the Jacobi method converges when applied to matrices that are diagonally dominant. In the following subproblems, you are free to use any of the individual questions (even if you didn't prove it) to answer the ones that come after it.

   (a) [15%] If $\mathbf{x} \in \mathbf{R}^n$ and $\mathbf{T}$ is an $n \times n$ matrix, show that for any positive integer k the following inequality holds:

   $$\|\mathbf{T}^k \mathbf{x}\| \leq \|\mathbf{T}\|^k \|\mathbf{x}\|$$

   (b) [5%] Show that the Jacobi method can be written as

   $$\mathbf{x}^{(k+1)} = \mathbf{T}\mathbf{x}^{(k)} + \mathbf{c}$$

   where $\mathbf{T} = \mathbf{D}^{-1}(\mathbf{L} + \mathbf{U})$ (using the decomposition $\mathbf{A} = \mathbf{D} - \mathbf{L} - \mathbf{U}$).

   (c) [10%] Define the *error* after the k-th iteration to be $\mathbf{e}^{(k)} = \mathbf{x}^{(k)} - \mathbf{x}^*$, where $\mathbf{x}^*$ is the *exact* solution to equation $\mathbf{A}\mathbf{x} = \mathbf{b}$. Show that:

   $$\mathbf{e}^{(k+1)} = \mathbf{T}\mathbf{e}^{(k)}$$

   Hint: You can use (after explaining why it is true) that $\mathbf{x}^* = \mathbf{T}\mathbf{x}^* + \mathbf{c}$.

   (d) [15%] If $\mathbf{A}$ is diagonally dominant by rows, and $\mathbf{T}$ is the matrix defined in (b) above, show that $\|\mathbf{T}\|_\infty < 1$.

   Hint: The way to prove this question may become more apparent if you write out explicitly what the first few rows of $\mathbf{T} = \mathbf{D}^{-1}(\mathbf{L} + \mathbf{U})$ look like.

   (e) [15%] Show that when $\mathbf{A}$ is diagonally dominant by rows , then:

   $$\|\mathbf{e}^{(k)}\|_\infty \leq \|\mathbf{T}\|_\infty^k \|\mathbf{e}^{(0)}\|_\infty$$

   Explain why this result implies that Jacobi is guaranteed to converge with diagonally dominant matrices.

2. [60%] In this problem you will show (in 2 different ways) that the **QR** factorization can, in fact, be used to construct the least-squares solution of an overdetermined system. First let us review some notation:

   Let $\mathbf{A}\mathbf{x} = \mathbf{b}$ be an overdetermined linear system of equations with m equations and n unknowns, i.e $\mathbf{A}$ is an $m \times n$ matrix with $m > n$. Assume we have obtained the factorization $\mathbf{A} = \mathbf{QR}$ where $\mathbf{Q}$ is an $m \times m$ orthogonal matrix ($\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}_{m\times m}$) and $\mathbf{R}$ is an $m \times n$ upper triangular matrix ($[\mathbf{R}]_{ij} = 0$ when $i > j$). We also write $\mathbf{Q}$ and $\mathbf{R}$ in the following block form:

$$\mathbf{Q} = \left(\hat{\mathbf{Q}} \mid \mathbf{Q}^*\right), \quad \mathbf{R} = \left(\begin{array}{c} \hat{\mathbf{R}} \\ \mathbf{0}_{(m-n)\times n} \end{array} \right)$$

where $\hat{\mathbf{Q}} \in \mathbf{R}^{m\times n}$ contains the first m columns of $\mathbf{Q}$, and $\mathbf{Q}^* \in \mathbf{R}^{m\times(m-n)}$ includes the last $m-n$ columns. Likewise, the upper triangular $n \times n$ matrix $\hat{\mathbf{R}}$ contains the first n rows of $\mathbf{R}$. We will also assume that the columns of $\mathbf{A}$ were linearly independent, and consequently according to the theory presented in class the diagonal elements of $\mathbf{R}$ (and $\hat{\mathbf{R}}$, too) are nonzero.

(a) [15%] Show that the normal equations for this system can be written equivalently as

$$\hat{\mathbf{R}}^T\hat{\mathbf{R}}\mathbf{x} = \hat{\mathbf{R}}^T\hat{\mathbf{Q}}^T\mathbf{b}$$

(b) [10%] Show that the least squares solution can be obtained from the system

$$\hat{\mathbf{R}}\mathbf{x} = \hat{\mathbf{Q}}^T\mathbf{b}$$

(c) [15%] Show that the 2-norm of the residual satisfies the equality:

$$\|\mathbf{b} - \mathbf{Ax}\|_2^2 = \|\mathbf{Q}^T\mathbf{b} - \mathbf{Rx}\|_2^2$$

Hint: You may use the equality $\mathbf{y}^T\mathbf{y} = \|\mathbf{y}\|_2^2$ which was proved in class. Expand both sides of the equality and show they are equal.

(d) [5%] Show that if $\mathbf{r}_1 \in \mathbf{R}^n$, $\mathbf{r}_2 \in \mathbf{R}^{m-n}$ and

$$\mathbf{r} = \left(\begin{array}{c} \mathbf{r}_1 \\ \mathbf{r}_2 \end{array} \right) \in \mathbf{R}^m$$

is the vector that results from stacking $\mathbf{r}_1$ and $\mathbf{r}_2$, then

$$\|\mathbf{r}\|_2^2 = \|\mathbf{r}_1\|_2^2 + \|\mathbf{r}_2\|_2^2$$

(e) [15%] Show that the squared residual norm can also be written as:

$$\|\mathbf{b} - \mathbf{Ax}\|_2^2 = \|\hat{\mathbf{Q}}^T\mathbf{b} - \hat{\mathbf{R}}\mathbf{x}\|_2^2 + \|\mathbf{Q}^*\mathbf{b}\|_2^2$$

Hint: Replace $\mathbf{Q}$ and $\mathbf{R}$ with their block forms and use (d).
The significance of the last expression is that the term $\|\hat{\mathbf{Q}}^T\mathbf{b} - \hat{\mathbf{R}}\mathbf{x}\|_2^2$ can be driven to zero by just using the $\mathbf{x}$ which solves $\hat{\mathbf{R}}\mathbf{x} = \hat{\mathbf{Q}}^T\mathbf{b}$, while the term $\|\mathbf{Q}^*\mathbf{b}\|_2^2$ which is independent of $\mathbf{x}$ is the "inevitable" minimum that the residual norm cannot go below.