Homework #2 . MF 5012.

(Due on thursday March 27, 2008).

(4pts) ① The weighted Jacobi method for solving the matrix equation $Ax=b$ is given by

$$x^{k+1} = (1-\omega)\, x^{k} + \omega\, x_{J,k}$$
$$x_{J,k} = -D^{-1}(A-D)\, x^{k} + D^{-1} b$$

where D is the diagonal matrix such that $d_{ii} = a_{ii}$ for $i = 1, 2, \dots, n$.

Assuming that A is symmetric and positive definite and that $a_{ij} \le 0$ for $i \ne j$, for what values of ω is the method convergent?

(4pts) ② The Kantorovich inequality is (equivalent to) the following inequality

$$\left(\sum_{i=1}^{n} \xi_i \mu_i \right)\left(\sum_{i=1}^{n} \xi_i \mu_i^{-1} \right) \le \frac{(M+m)^2}{4Mm}$$

where $0 < m \le \mu_i \le M$ and $\xi_i \geqslant 0$ for $i = 1, \dots, n$, and $\sum_{i=1}^{n} \xi_i = 1$.

Prove the inequality.

Hint: First, show that for each μ_i you can find p_i and q_i such that

$$\mu_i = p_i M + q_i m$$
$$\mu_i^{-1} = p_i M^{-1} + q_i m^{-1}$$

$$p_i, q_i \geqslant 0$$
$$p_i + q_i \leqslant 1 .$$

Then, set $p = \sum_{i=1}^{n} \xi_i p_i$ and $q = \sum_{i=1}^{n} \xi_i q_i$, show that

$$\left(\sum_{i=1}^{n} \xi_i \mu_i\right)\left(\sum_{i=1}^{n} \xi_i \mu_i^{-1}\right) = (pM + qm)\,(pM^{-1} + qm^{-1}),$$

and conclude.

(12 pts) 3 Let Ω be the unit square and let U_{ij} denote the classic finite difference approximation to the exact solution of

$$(P) \qquad \begin{cases} -\Delta u = 1 & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases}$$

that is

$$(P)_h \qquad \begin{cases} -\frac{1}{h^2}\left(U_{i+1,j} + U_{i-1,j} + U_{i,j+1} + U_{i,j-1} - 4\,U_{ij}\right) = 1 \\ \qquad\qquad \text{if} \quad i,j = 1, \dots, N-1 \\ U_{ij} = 0 \quad \text{if} \quad i=0 \text{ or } i=N \text{ or } j=0 \text{ or } j=N. \end{cases}$$

Here $h = \frac{1}{N}$.

The objective of this exercise is to compare the performance of the Jacobi, Gauss-Seidel, SOR (with the parameter ω of your choice!), steepest descent and conjugate gradient methods.

Given a value for $h = \frac{1}{N}$, and for each of these methods, compute the number of iterations that reduce the initial error (Take $x^0 = 0$!) 10^6 times. Take $N = 4, 8, 16, \ldots$ Fill the following Table:

$N = 1/h$	J	G-S	SOR	SD	CG
4	...				
8					
16					
$\vdots$					

Explain the results by using the theory, if possible.

Homework #2. MF 5012 - A solution

① The iteration matrix of the weighted Jacobi method is

$$G = Id - \omega D^{-1} A.$$

It can be easily verified that the method is consistent regardless of the value of the parameter ω.

To find for what values of ω, G has a spectrum strictly less than one, we are going to use the theorem on page 4 of the notes on "Classic Iterative Methods". The result states that if we write

$$A = N - P,$$

and set

$$G = N^{-1} P,$$

we have that $\rho(G) < 1$ provided A is symmetric and positive definite and provided

$$Q = N + N^T - A$$

is positive definite.

In our case, we have that

$$\begin{aligned} Id - \omega D^{-1} A &= G \\ &= N^{-1} P \\ &= N^{-1}(N - A) \\ &= Id - N^{-1} A \end{aligned}$$

and so we must have $N = D/\omega$. This implies that

$$Q = \frac{2}{\omega} D - A,$$

and it only remains to find for what values of the parameter ω this matrix is positive definite.

Since A is symmetric and positive definite and since its off-diagonal terms are negative, it can be shown that the matrix

$$2D - A$$

is positive definite. Using this fact, we see that

$$Q = 2\left(\frac{1-\omega}{\omega}\right) D + (2D - A)$$

is positive definite if $\left(\frac{1-\omega}{\omega}\right) > 0$, that is, if $\omega \in [0, 1)$. The value $\omega = 0$, however, must be ruled out since then $G = Id$. Thus the method is stable for $\omega \in (0, 1)$.

If the theorem in the notes escaped your attentive reading, let us deal with a simpler case and assume the hypothesis used in class, namely, that the matrix A is strictly diagonal dominant.

So, let e be an eigenvector of G and let λ be its corresponding eigenvalue. Then the equation

$$Ge = \lambda e$$

becomes

$$e - \omega D^{-1} A e = \lambda e$$

$$\Rightarrow \quad e(1-\omega) - \omega D^{-1}(A-D)\, e = \lambda e$$

$$\Rightarrow \quad (\lambda - 1 + \omega)\, e = -\omega D^{-1}(A-D)\, e$$

or equivalently

$$(\lambda - 1 + \omega)\, e_i = - \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{a_{ij}}{a_{ii}}\, \omega\, e_j \qquad i = 1, \dots, n.$$

Taking absolute values we get

$$|\lambda - 1 + \omega|\, |e_i| \leq - \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{a_{ij}}{a_{ii}}\, |\omega|\, |e_j| \qquad i = 1, \dots, n$$

If i is such that $|e_j| \leq |e_i|$ $j = 1, \dots, n$, we get

$$|\lambda - 1 + \omega| \leq - \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{a_{ij}}{a_{ii}}\, |\omega| =: \theta$$

or

$$-\theta \leq \lambda - 1 + \omega \leq \theta .$$

this implies that

$$1 - \omega - \theta \leq \lambda \leq \theta - \omega + 1$$

and if we want $|\lambda| < 1$, we must have

$$-1 < 1 - \omega - \theta \quad \text{and} \quad \theta - \omega + 1 < 1.$$

the second inequality implies that ω must be nonnegative and that $\theta < \omega$, which is nothing but our assumption of the matrix being strictly diagonal dominant. the second inequality now holds if and only if

$$\omega < \frac{2}{1 - \sum_{\substack{i \neq j \\ j=1}}^{n} \frac{a_{ij}}{a_{ii}}}$$

Hence, the method converges if

$$\omega \in \left(0, \frac{2}{1 - \sum_{\substack{i \neq j \\ j=1}}^{n} \frac{a_{ij}}{a_{ii}}}\right).$$

(2) By definition of p_i and q_i, namely,

$$\mu_i = p_i M + q_i m$$
$$\mu_i^{-1} = p_i M^{-1} + q_i m^{-1}$$

we easily obtain that

$$p_i = \frac{\mu_i^2 - m^2}{M^2 - m^2} \cdot \frac{M}{\mu_i}$$

$$q_i = \frac{M^2 - \mu_i^2}{M^2 - m^2} \cdot \frac{m}{\mu_i}.$$

We then easily see that p_i and q_i are non-negative since $\mu_i \in [m, M]$. Now

$$p_i + q_i = \frac{1}{M^2 - m^2}\left[M\mu_i - \frac{m^2 M}{\mu_i} + \frac{M^2 m}{\mu_i} - m\mu_i\right]$$

$$= \frac{1}{M+m}\left[\mu_i + \frac{mM}{\mu_i}\right]$$

$$\leq \frac{1}{M+m} \max_{\mu \in [m,M]} \left[\mu + \frac{mM}{\mu}\right]$$

$$= 1.$$

Since

$$\sum_{i=1}^{n} \xi_i \mu_i = pM + qm$$
$$\sum_{i=1}^{n} \xi_i \mu_i^{-1} = pM^{-1} + qm^{-1}$$

we readily get that

$$\left(\sum_{i=1}^{n} \xi_i \mu_i\right)\left(\sum_{i=1}^{n} \xi_i \mu_i^{-1}\right) = (pM + qm)(pM^{-1} + qm^{-1}).$$

Since p_i, q_i are non-negative and $p_i + q_i \leq 1$, we have that p, q are non-negative and $p+q \leq 1$. Then

$$\left(\sum_{i=1}^{n} \xi_i \mu_i\right)\left(\sum_{i=1}^{n} \xi_i \mu_i^{-1}\right) \leq \max_{\substack{p,q \geq 0 \\ p+q \leq 1}} \Phi(p,q),$$

where

$$\Phi(p,q) = (pM + qm)(pM^{-1} + qm^{-1}).$$

But

$$\Phi(p,q) = p^2 + q^2 + pq\left(\frac{M}{m} + \frac{m}{M}\right)$$
$$= (p+q)^2 + pq\left(\frac{M}{m} + \frac{m}{M} - 2\right)$$

and since $\frac{M}{m} + \frac{m}{M} \geq 2$,

$$\Phi(p,q) \leq \Phi\left(\tfrac{1}{2}, \tfrac{1}{2}\right)$$
$$= 1 + \frac{1}{4}\left(\frac{M}{m} + \frac{m}{M} - 2\right)$$
$$= \frac{1}{4}\left(\frac{M}{m} + \frac{m}{M} + 2\right)$$
$$= \frac{(M+m)^2}{4Mm}.$$

Hence

$$\left(\sum_{i=1}^{n} \xi_i \mu_i\right)\left(\sum_{i=1}^{n} \xi_i \mu_i^{-1}\right) \leq \frac{(M+m)^2}{4Mm},$$

as wanted.

(3) Here we recall what is needed for the theoretical interpretation of the numerical results (this is based on the notes in "Classic Iterative methods).
We begin by noting that

$$u(x,y) = \sin(\pi n x)\,\sin(\pi m y)$$

is such that

$$-\Delta u = \pi^2 (n^2+m^2)\, u \qquad \text{in } \Omega,$$
$$u = 0 \qquad \text{in } \partial\Omega.$$

In other words, u is an eigenfunction of $-\Delta$ (with zero boundary conditions) and $\lambda = \pi^2(n^2+m^2)$ is its corresponding eigenvalue; n and m are arbitrary integers, of course.

Inspired in this result, let us see if we have a similar result for the eigenfunctions $\{U_{ij}\}_{i,j=0\to N}$ of the discrete version of the above problem given by $(P)_h$. So, take

$$U_{ij} = \sin\left(\pi n \frac{i}{N}\right) \sin\left(\pi m \frac{j}{N}\right)$$

for $i=0\to N$ and $j=0\to N$. First, we see that U_{ij} satisfies the boundary conditions since $U_{ij}=0$ for $i=0$, $i=N$, $j=0$, or $j=N$. Note also that both n and m are integers bigger than 0 and smaller than N!

Now set

$$\Theta_{ij} = -\frac{1}{h^2}\left(U_{i+1,j} + U_{i-1,j} + U_{i,j+1} + U_{i,j-1} - 4\, U_{ij} \right).$$

Then, noting that

$$U_{i+1,j} + U_{i-1,j} = \left(\sin\left(\pi n \frac{i+1}{N}\right) + \sin\left(\pi n \frac{i-1}{N}\right)\right) \sin\left(\pi m \frac{j}{N}\right)$$

$$= 2\cos\left(\pi \frac{n}{N}\right) \sin\left(\pi n \frac{i}{N}\right) \sin\left(\pi m \frac{j}{N}\right)$$

$$= 2\cos\left(\pi \frac{n}{N}\right) U_{ij},$$

and that, similarly,

$$U_{i,j+1} + U_{i,j-1} = 2\cos\left(\pi \frac{m}{N}\right) U_{ij},$$

we obtain that

$$\Theta_{ij} = \left[-\frac{1}{h^2}\left(2\cos\left(\pi \frac{m}{N}\right) + 2\cos\left(\pi \frac{n}{N}\right) - 4 \right)\right] U_{ij}$$

$$= \left[+\frac{1}{h^2}\left(+4\sin^2\left(\frac{\pi n}{2N}\right) + 4\sin^2\left(\frac{\pi m}{2N}\right)\right)\right] U_{ij}$$

$$= \left[\left(\frac{\sin\left(\frac{\pi n}{2N}\right)}{\frac{\pi n}{2N}} \right)^2 \pi^2 n^2 + \left(\frac{\sin\left(\frac{\pi m}{2N}\right)}{\frac{\pi m}{2N}} \right)^2 \pi^2 m^2 \right] U_{ij}$$

given that $h = \frac{1}{N}$. We thus see that U_{ij} is indeed an eigenfunction of the discrete operator defined by (P), whose corresponding eigenvalue is

$$\lambda_{nm} = \left(\frac{\sin\left(\frac{\pi n}{2N}\right)}{\frac{\pi n}{2N}} \right)^2 \pi^2 n^2 + \left(\frac{\sin\left(\frac{\pi m}{2N}\right)}{\frac{\pi m}{2N}} \right)^2 \pi^2 m^2 .$$

Since the function $\frac{1}{\theta}\sin\theta$ goes to one as θ goes to zero, we see that λ_{nm} goes to $\pi(n^2+m^2)$ as N goes to infinity, as expected!

With this information, we can obtain the spectral radius of the iteration matrix of the Jacobi method, $I - D^{-1}A$. Since $D = \frac{4}{h^2}\,Id$, we get that the eigenvalues of $I - D^{-1}A$ are

$$\lambda^J_{nm} := 1 - \frac{h^2}{4}\lambda_{nm} \qquad 1 \leq n, m \leq N-1$$

Hence

$$\lambda^J_{nm} = 1 - \sin^2\left(\frac{\pi n}{2N}\right) - \sin^2\left(\frac{\pi m}{2N}\right) \qquad 1 \leq n, m \leq N-1 .$$

and the spectral radius is then

$$\rho_J := \max_{1 \leq n, m \leq N-1} \left|\lambda^J_{nm}\right| = \left|\lambda^J_{N-1\; N-1}\right|$$

$$= \left| 1 - 2\sin^2\left(\frac{\pi(N-1)}{2N}\right)\right|$$

$$= \left| 1 - 2\sin^2\left(\frac{\pi}{2N}\right)\right|$$

Since our matrix is symmetric, positive definite matrix who is also block-tridiagonal, we have

$$\rho_{SOR} = \omega_{opt} - 1 < \rho_{GS} = \rho_J^2$$

where

$$\omega_{opt} = \frac{2}{1 + \sqrt{1 - \rho_J^2}}$$

Hence

$$\rho_{GS} = \left| 1 - 2 \sin^2\left(\frac{\pi}{2N}\right) \right|^2$$

$$\rho_{SOR} = \frac{2}{1 + \sqrt{4 \sin^2\left(\frac{\pi}{2N}\right) - 4 \sin^4\left(\frac{\pi}{2N}\right)}} - 1$$

So, for $N \geqslant 2$,

$$\rho_J = 1 - 2 \sin^2\left(\frac{\pi}{2N}\right),$$

$$\rho_{GS} = 1 - \sin^2\left(\frac{\pi}{N}\right),$$

$$\rho_{SOR} = 1 - \frac{2 \sin\left(\frac{\pi}{N}\right)}{1 + \sin\left(\frac{\pi}{N}\right)},$$

and for big N,

$$\rho_J \sim 1 - \frac{\pi^2}{2N^2}$$

$$\rho_{GS} \sim 1 - \frac{\pi^2}{N^2}$$

$$\rho_{SOR} \sim 1 - 2\frac{\pi}{N}$$

To reduce the initial error 10^6 times, we thus need that

$$\rho^k = 10^{-6}$$
$$\Rightarrow \quad k \log \rho = -6 \log 10$$
$$\Rightarrow \quad k = -6 \log 10 / \log \rho$$

Hence, since $\log_e(1-x) \sim -x$ for small $x > 0$,

$$k_J = -6 \log_e 10 / \log_e \rho_J \sim \frac{12}{\pi^2} \log_e 10 \cdot N^2$$

$$k_{GS} = -6 \log_e 10 / \log_e \rho_{GS} \sim \frac{6}{\pi^2} \log_e 10 \cdot N^2$$

$$k_{SOR} = -6 \log_e 10 / \log_e \rho_{SOR} \sim \frac{3}{\pi} \log_e 10 \cdot N$$

To see how many iterations we must do for the SD and CG methods, let us recall that

$$\|x^k - x\|_A \leq \left(\frac{K-1}{K+1}\right)^k \|x^0 - x\|_A \qquad \text{(SD)}$$
$$\|x^k - x\|_A \leq 2\left(\frac{\sqrt{K}-1}{\sqrt{K}+1}\right)^k \|x^0 - x\|_A \qquad \text{(CG)}$$

where K is the condition number of A.

then

$$k_{SD} = -6 \log 10 / \log\left(\frac{K-1}{K+1}\right),$$
$$k_{CG} = \log(10^{-6}/2) / \log\left(\frac{\sqrt{K}-1}{\sqrt{K}+1}\right).$$

To evaluate K is easy since A being symmetric and positive definite,

$$K = \frac{\max\limits_{1\le n,m \le N-1} \lambda_{nm}}{\min\limits_{1 \le n,m \le N-1} \lambda_{nm}}$$

$$= \frac{\lambda_{N-1,N-1}}{\lambda_{11}}$$

$$= \frac{\sin^2\left(\frac{\pi}{2}\,\frac{N-1}{N}\right)}{\sin^2\left(\frac{\pi}{2N}\right)}$$

$$= \frac{1}{\sin^2\left(\frac{\pi}{2N}\right)} - 1$$

$$\sim \frac{4N^2}{\pi^2}$$

for big N. So, if N is big, so is K and

$$k_{SD} \sim \frac{12}{\pi^2} \log_e 10 \cdot N^2$$

$$k_{CG} \sim \left[6 \log_e 10 + \log_e 2\right] \frac{N}{\pi}.$$

Hence, for N big, we have

$$k_J \sim 2.8 N^2$$
$$k_{GS} \sim 1.4 N^2$$
$$k_{SOR} \sim 2.2 N$$
$$k_{SD} \sim 2.8 N^2$$
$$k_{CG} \sim 4.6 N$$

So, we have the following theoretical result:

$N = 1/h$	J	GS	SOR	SD	CG
4	45	22	9	22	18
8	179	90	18	90	37
16	717	358	35	358	73
32	2,867	1433	70	1433	147

(You can now compare your results with these theoretical predictions!)

Now, if you did not used the SOR method with optimal ω, you have to proceed as follows. First, (see page 9 of the notes on "Classical Iterative methods), note that in our case the eigenvalues of the iteration matrix of the SOR method are

$$\lambda_{nm}^{SOR} = \frac{\sqrt{\lambda_{nm}^J}}{\omega} + \frac{1}{\sqrt{\lambda_{nm}^J}} - \frac{1}{\omega\sqrt{\lambda_{nm}^J}} \qquad 1 \le n, m \le N-1 .$$

Hence

$$\lambda_{nm}^{SOR} = \frac{\omega - \sin^2\left(\frac{\pi n}{2N}\right) - \sin^2\left(\frac{\pi m}{2N}\right)}{\omega\sqrt{1 - \sin^2\left(\frac{\pi n}{2N}\right) - \sin^2\left(\frac{\pi m}{2N}\right)}}$$

and, assuming that $\omega > 1$,

$$\lambda_{nm}^{SOR} \leq \lambda_{11}^{SOR} \left(=: \rho^{SOR} \right)$$

$$\sim \frac{\omega - \left(\frac{\pi}{2N}\right)^2 2}{\omega}$$

$$\sim 1 - \frac{\pi^2}{2\omega} \cdot \frac{1}{N^2}$$

Hence

$$k^{SOR} = -6 \log_e 10 / \log_e \rho^{SOR} \sim \frac{12\omega}{\pi^2} \log_e 10 \cdot N^2 .$$

We then see that the behavior of SOR in this case is by far inferior to that of SOR with optimal choice of ω.