HW #1 Solutions: M552 Spring 2006

1. (**1.4**-*Trefethen & Bau*) Let $f_1, \ldots, f_8$ be a set of functions defined on the interval $[1, 8]$ with the property that for any numbers $d_1, \ldots, d_8$, there exists a set of coefficients $c_1, \ldots, c_8$ such that

$$\sum_{j=1}^{8} c_j f_j(i) = d_i, \quad i = 1, \ldots, 8.$$

(a) Show by appealing to the theorems of this lecture that $d_1, \ldots, d_8$ determine $c_1, \ldots, c_8$ uniquely.

(b) Let A be the 8×8 matrix representing the linear mapping from data $d_1, \ldots, d_8$ to coefficients $c_1, \ldots, c_8$. What is the i, j entry of A^{-1}?

ANS:

(a) The expression is equivalent to the 8×8 system

$$Fc = d, \quad F_{ij} = f_j(i), \quad c = [c_1, \ldots, c_8]^T, \quad d = [d_1, \ldots, d_8]^T.$$

To say there exists a c given any d is equivalent to saying that the columns of F span $\mathbb{R}^8$ (or $\mathbb{C}^8$). But F is 8×8, so $rank(F) = 8$, and by Theorem 1.3 F is invertible. Hence $c = F^{-1}d$, i.e., c is *uniquely* determined.

(b) We have $Ad = c$ or $d = A^{-1}c$. But form (a), $d = Fc$, so we must have $A^{-1} = F$. Therefore, $A^{-1}_{ij} = F_{ij} = f_j(i)$.

2. (**2.2**-*Trefethen & Bau*) The Pythagorean theorem asserts hat for a set of n orthogonal vectors $\{x_i\}$,

$$\left\| \sum_{i=1}^{n} x_i \right\|^2 = \sum_{i=1}^{n} \|x_i\|^2.$$

(a) Prove this in the case $n = 2$ by an explicit computation of $\|x_1 + x_2\|^2$.

(b) Show that this computation also establishes the general case, by induction.

ANS:

(a) A direct calculation using $\|x\|^2 = x^*x$ for a general vector x shows:

$$\begin{aligned} \|x_1 + x_2\|^2 &= (x_1 + x_2)^*(x_1 + x_2) \\ &= {x_1}^*x_1 + {x_1}^*x_2 + {x_2}^*x_1 + {x_2}^*x_2 \\ &= {x_1}^*x_1 + 0 + 0 + {x_2}^*x_2 \\ &= \|x_1\|^2 + \|x_2\|^2. \end{aligned}$$

(b) Above we established the result for $n=2$. Assume (the induction hypothesis) that the result is true for $n-1$, i.e., $\left\| \sum_{i=1}^{n-1} x_i \right\|^2 = \sum_{i=1}^{n-1} \|x_i\|^2$ for $n-1$ mutually orthogonal vectors $x_1, \ldots, x_{n-1}$. Then given n mutually orthogonal vectors $x_1, \ldots, x_n$, let $x = \sum_{i=1}^{n-1} x_i$ and $y = x_n$. Since x_n is orthogonal to each x_i for $i = 1, \ldots, n-1$, by linearity of the inner product we have $x^*y = 0$. Then applying (a) to x and y gives

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 = \left\| \sum_{i=1}^{n-1} x_i \right\|^2 + \|x_n\|^2,$$

and applying the induction hypothesis to the term $\left\| \sum_{i=1}^{n-1} x_i \right\|^2$ gives the desired result.

3. (**2.3**-*Trefethen & Bau*) Let $A \in \mathbb{C}^{m\times m}$ be hermitian. An eigenvector of A is a nonzero vector $x \in \mathbb{C}^m$ such that $Ax = \lambda x$ for some $\lambda \in \mathbb{C}$, the corresponding eigenvalue.

(a) Prove that all eigenvalues are real.

(b) Prove that if x and y are eigenvectors corresponding to distinct eigenvalues, then x and y are orthogonal.

ANS: Suppose $Ax = \lambda_x x$ and $Ay = \lambda_y y$, where $A^* = A$, $x \neq 0$ and $y \neq 0$, and $\lambda_x \neq \lambda_y$.

(a) $Ax = \lambda_x x \quad \Rightarrow \quad x^*Ax = x^*(\lambda_x x) = \lambda_x x^* x \quad \Rightarrow \quad \lambda_x = \dfrac{x^*Ax}{x^*x}$, noting that $x^*x \neq 0$.

$\Rightarrow \quad \overline{\lambda_x} = \dfrac{\overline{x^*Ax}}{\overline{x^*x}} = \dfrac{(x^*Ax)^*}{(x^*x)^*} = \dfrac{x^*A^*x}{x^*x} = \dfrac{x^*Ax}{x^*x} = \lambda_x \quad \Rightarrow \quad \overline{\lambda_x} = \lambda_x \quad \Rightarrow \quad \lambda_x \in \mathbb{R}.$

(b) We have

$$\begin{aligned}
(\lambda_x - \lambda_y)x^*y &= \lambda_x x^*y - \lambda_y x^*y \\
&= (\overline{\lambda_x}x)^*y - x^*(\lambda_y y) \\
&= (\lambda_x x)^*y - x^*(\lambda_y y) \quad \text{since } \lambda_x \in \mathbb{R} \\
&= (Ax)^*y - x^*(Ay) \\
&= x^*A^*y - x^*Ay \\
&= x^*Ay - x^*Ay \quad \text{since } A^* = A \\
&= 0
\end{aligned}$$

$\Rightarrow \quad x^*y = 0$ since $\lambda_x \neq \lambda_y \quad \Rightarrow \quad x$ and y are orthogonal.

4. (**2.4**-*Trefethen & Bau*) What can be said about the eigenvalues of a unitary matrix?

   ANS: Suppose Q is unitary ($Q^{-1} = Q^*$) and (λ, x) is an e-pair, i.e. $Qx = \lambda x$. Then for any vector y

$$\|Qy\| = \sqrt{(Qy)^* Qy} = \sqrt{y^* Q^* Qy} = \sqrt{y^* (Q^* Q) y} = \sqrt{y^* y} = \|y\|$$

   So multiplying any vector by Q preserves the length of the vector. Now let $y = x$. We have

$$\|x\| = \|Qx\| = \|\lambda x\| = \sqrt{(\lambda x)^* \lambda x} = \sqrt{\overline{\lambda} \lambda x^* x} = |\lambda| \|x\|,$$

   and $\|x\| \neq 0$ since x is an e-vector, so we must have $|\lambda| = 1$. Thus all the e-values of a unitary matrix must lie on the unit circle in the complex plane.

5. (**2.5**-*Trefethen & Bau*) Let $S \in \mathbb{C}^{m\times m}$ be *skew-hermitian*, i.e., $S^* = -S$.

(a) Show by using Exercis 2.3 that the eigenvalues of S are pure imaginary.

(b) Show that $I - S$ is nonsingular.

(c) Show that the matrix $Q = (I - S)^{-1}(I + S)$, known as the *Cayley transform* of S, is unitary. (This is the matrix analogue of a linear fractional transformation $(1 + s)/(1 - s)$, which maps the left half os the complex s-plane conformally onto the unit disk.)

ANS: Suppose $Sx = \lambda x$ where $S^* = -S$, $x \neq 0$.

(a) $Sx = \lambda x \quad \Rightarrow \quad x^*Sx = x^*(\lambda x) = \lambda x^*x \quad \Rightarrow \quad \lambda = \dfrac{x^*Sx}{x^*x}$, noting that $x^*x \neq 0$.

$\Rightarrow \quad \overline{\lambda} = \dfrac{\overline{x^*Sx}}{\overline{x^*x}} = \dfrac{(x^*Sx)^*}{(x^*x)^*} = \dfrac{x^*S^*x}{x^*x} = -\dfrac{x^*Sx}{x^*x} = -\lambda \quad \Rightarrow \quad \overline{\lambda} = -\lambda,$

that is, λ is pure imaginary.

(b) Suppose (λ, x) is an e-pair of S, and $p(t) = a_nt^n + a_{n-1}t^{n-1} + \ldots + a_1t + a_0$ is a polynomial. Then,

$$p(S)x = (a_nS^n + a_{n-1}S^{n-1} + \ldots + a_1S + a_0I)x = (a_n\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0I)x = p(\lambda)x,$$

i.e., $(p(\lambda), x)$ is an e-pair of $p(S)$. Let $p(t) = 1 - t$. Then the e-values of $I - S = p(S)$ are $p(\lambda) = 1 - \lambda$ where λ is an e-value of S. From (a) we know that λ is pure imaginary, thus $1 - \lambda \neq 0$, hence no e-value of $I - S$ equals 0, i.e., $I - S$ is nonsingular. The same argument shows $I + S$ is also nonsingular.

(c) Recall $(AB)^* = B^*A^*$, $(A+B)^* = A^* + B^*$, $I^* = I$, and if A is nonsingular that $(A^*)^{-1} = (A^{-1})^*$. Then

$$\begin{aligned}
QQ^* &= (I - S)^{-1}(I + S)[(I - S)^{-1}(I + S)]^* \\
&= (I - S)^{-1}(I + S)(I + S)^*[(I - S)^{-1}]^* \\
&= (I - S)^{-1}(I + S)(I + S^*)(I - S^*)^{-1} \\
&= (I - S)^{-1}(I + S)(I - S)(I + S)^{-1} \\
&= (I - S)^{-1}(I - S)(I + S)(I + S)^{-1}, \quad \text{since } (I - S) \text{ and } (I + S) \text{ commute} \\
&= I * I = I
\end{aligned}$$

Thus $Q^* = Q^{-1}$, showing Q is unitary.

6. Let $A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ and $b = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Solve $Ax = b$ (by hand) using the spectral decomposition of A. Show all details.

ANS: $A^T = A \Rightarrow A$ is symmetric, hence diagonalizable by a orthogonal matrix, $A = UDU^{-1} = UDU^T$, where $D = diag(\lambda_1, \lambda_2)$. First we need to find the e-values and a corresponding unit e-vector for each, i.e., the e-pairs (λ_1, u_1) and (λ_2, u_2).

$$det(A-\lambda I) = det \begin{pmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{pmatrix} = (2-\lambda)^2 - (-1)^2 = \lambda^2 - 4\lambda + 3 = (\lambda-1)(\lambda-3) \Rightarrow \lambda_1 = 1, \lambda_2 = 3.$$

$$\lambda = 1: \quad (A-1I)u = 0 \;\Rightarrow\; \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} u = 0 \;\Rightarrow\; u = (1,1)^T \;\Rightarrow\; u_1 = \frac{u}{\|u\|} = (1/\sqrt{2}, 1/\sqrt{2})^T.$$

$$\lambda = 3: \quad (A-3I)u = 0 \;\Rightarrow\; \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} u = 0 \;\Rightarrow\; u = (-1,1)^T \;\Rightarrow\; u_2 = \frac{u}{\|u\|} = (-1/\sqrt{2}, 1/\sqrt{2})^T.$$

So we have $Au_1 = \lambda_1 u_1$ and $Au_2 = \lambda_2 u_2$ where both u_1 and u_2 are unit vectors. Thus, $U = [u_1|u_2]$, or

$$U = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \quad \Rightarrow \quad U^{-1} = U^T = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}.$$

Then,

$$U^{-1}b = U^T b = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}, \quad \text{or} \quad b = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} u_1 + \frac{-1}{\sqrt{2}} u_2,$$

giving the *coordinates* of b in the e-basis $\{u_1, u_2\}$. So finally,

$$x = \frac{1}{\lambda_1}\frac{1}{\sqrt{2}} u_1 + \frac{1}{\lambda_2}\frac{-1}{\sqrt{2}} u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} + \frac{-1}{3\sqrt{2}} \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} + \begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix} = \begin{pmatrix} 2/3 \\ 1/3 \end{pmatrix}.$$

In other words, $x = A^{-1}b = (UDU^T)^{-1}b = UD^{-1}(U^T b)$.

7. Write a MATLAB function M-file **trilu** to find the LU decomposition as discussed in class, $A = LU$, for the tridiagonal $m \times m$ matrix A,

$$A = \begin{pmatrix} a_1 & c_1 & & & \\ b_2 & a_2 & c_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & c_{m-1} \\ & & & b_m & a_m \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ \beta_2 & 1 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & \beta_m & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 & c_1 & & & \\ & \alpha_2 & c_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & c_{m-1} \\ & & & & \alpha_m \end{pmatrix} = LU.$$

The function should output the two m-vectors α and β, and its first line should read:

```
function [alpha,beta] = trilu(a,b,c)
```

Next, write an M-file function **trilu_solve** to solve $Ax = f$, which takes the vectors α, β, c and f and returns x. Its first line should read:

```
function x = trilu_solve(alpha,beta,c,f)
```

Test your code with the 5×5 system with $a_i = 2$, $b_i = -1$, $c_i = -1$, and RHS $f = [1, 0, 0, 0, 1]^T$. The exact solution is clearly $x = [1, 1, 1, 1, 1]^T$. Use MATLAB's **diary** command to save your MATLAB session output showing that your code works properly. Include a copy of both codes.

ANS: First, let's test the code:

```
>> a=2*ones(5,1);
>> b=-ones(5,1); b(1)=0;
>> c=-ones(5,1); c(5)=0;
>> f=[1 0 0 0 1]';
>> [alpha,beta]=trilu(a,b,c);

      alpha     beta
    ----------------
    2.0000         0
    1.5000   -0.5000
    1.3333   -0.6667
    1.2500   -0.7500
    1.2000   -0.8000

>> x=trilu_solve(alpha,beta,c,f);
>> x

x =

    1.0000
    1.0000
    1.0000
    1.0000
    1.0000
```

Here are the codes:

```
function [alpha,beta]=trilu(a,b,c)
%
%TRILU - Reduced LU decomposition of tridiagonal matrix.
%
m=length(a);
alpha=zeros(m,1);
beta=zeros(m,1);

alpha(1)=a(1);
for k=2:m
    beta(k)=b(k)/alpha(k-1);
    alpha(k)=a(k)-beta(k)*c(k-1);
end

function x=trilu_solve(alpha,beta,c,f)
%
%TRILU_SOLVE - solve tridiagonal system using decompostion
%              produced by TRILU.
%
m=length(c);
x=zeros(m,1);
z=zeros(m,1);

% solve Lz=f by forward substitution
z(1)=f(1);
for k=2:m
    z(k)=f(k)-beta(k)*z(k-1);
end

% solve Ux=z by backward substitution
x(m)=z(m)/alpha(m);
for k=m-1:-1:1
    x(k)=(z(k)-c(k)*x(k+1))/alpha(k);
end
```

8. Consider the 2-point BVP

$$\begin{cases} -y'' + (4x^2+2)y = 2x(1+2x^2) \\ y(0) = 1,\ y(1) = 1+e \end{cases}$$

Show $y(x) = x + e^{x^2}$ is the exact solution. Write a MATLAB function M-file to solve the problem using the 2nd order centered FD scheme we discussed in class, $-D_+D_-u_i + c_iu_i = f_i$. Use meshsize $h = 1/2^p$, where p is a positive integer. Your code should use your M-files **trilu** and **trilu_solve**. For $p = 1:4$, plot the exact solution ($y(x)$ vs. x) and the numerical solution (u_i vs. x_i), including the boundary points. The 4 plots should appear separately in one figure, with axes labeled and a title for each indicating p. Investigate **subplot** in MATLAB for how to have multiple plots in a single figure window. For $p = 1:20$ present a table with the following data - column 1: h; column 2: $\|u_h - y_h\|_\infty$; column 3: $\|u_h - y_h\|_\infty/h^2$; column 4: cpu time; column 5: (cpu time)/m, where $h = 1/(m+1)$. Discuss the trends in each column. Include a copy of your code.

ANS: First we check $y(x)$ is the solution. The boundary conditions are easily seen to be satisfied, and

$$y' = 1+2xe^{x^2} \Rightarrow y'' = (4x^2+2)e^{x^2} \Rightarrow -y''+(4x^2+2)y = -(4x^2+2)e^{x^2}+(4x^2+2)(x+e^{x^2}) = 2x(1+2x^2).$$

Here is the code for the first part of the problem. A listing of the M-file *bvp_solve* appears at the end of the solution of this problem. It requires the M-files *trilu* and *trilu_solve* from problem 7.

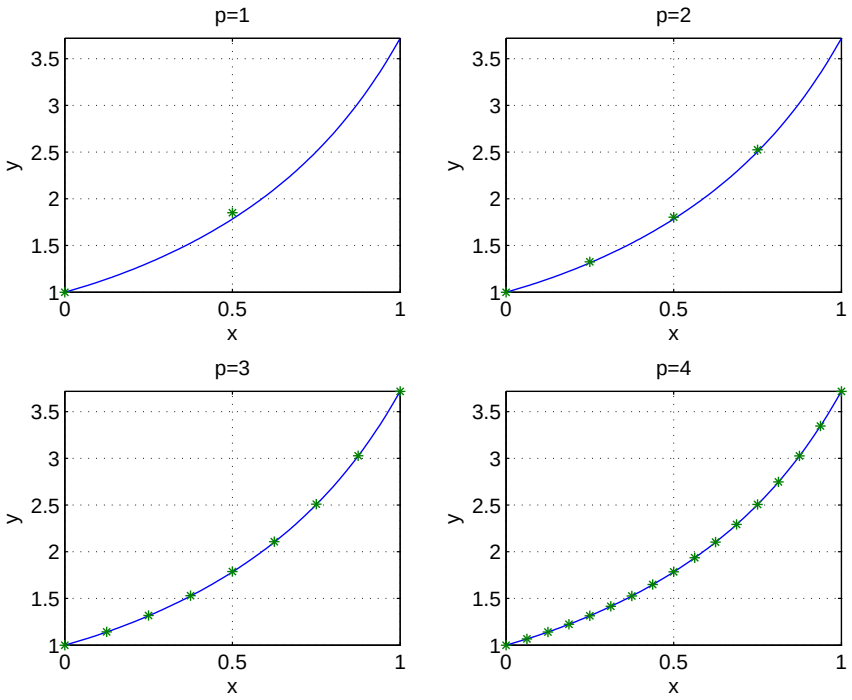
```
xx=0:0.01:1;xx=xx';
yy=xx+exp(xx.^2);                % exact solution
c='4*x^2+2'; f='2*x*(1+2*x^2)';  % functions for BVP
clf;
for p=1:4
    [x,u]=bvp_solve(2^p-1,0,1,1,1+exp(1),c,f);
    subplot(2,2,p),plot(xx,yy,x,u,'*'),grid
    axis('tight'),xlabel('x'),ylabel('y'),title(['p=',num2str(p)])
end
```

Here is a general BVP solver for our problem:

```
function [xv,uv,stime]=bvp_solve(m,a,b,ya,yb,c,f)
%
h=(b-a)/(m+1);
xv=a:h:b; xv=xv';
fv=zeros(m+2,1);
cv=zeros(m+2,1);
for i=1:m+2
    x=xv(i);
    cv(i)=eval(c);
    fv(i)=eval(f);
end
%
av=(2*ones(m,1)+h*h*cv(2:m+1))/(h*h);
bv=-ones(m,1)/(h*h); bv(1)=0;
cv=-ones(m,1)/(h*h); cv(m)=0;
```

```
fv=fv(2:m+1);
fv(1)=fv(1)+ya/(h*h); fv(m)=fv(m)+yb/(h*h);
%
tic;
[alpha,beta]=trilu(av,bv,cv);
uv=trilu_solve(alpha,beta,cv,fv);
stime=toc;
uv=[ya;uv;yb];
```

Here is the graph.



Next we solve the BVP with $m = 2^p - 1$ for $p = 1, \ldots, 20$. Here is the code:

```
c='4*x^2+2'; f='2*x*(1+2*x^2)';
clf;
h=zeros(20,1);
m=zeros(20,1);
times=zeros(20,1);
err_inf=zeros(20,1);
for p=1:20
    [x,u,stime]=bvp_solve(2^p-1,0,1,1,1+exp(1),c,f);
    h(p)=1/(2^p);
    m(p)=2^p-1;
    times(p)=stime;
```

```
    y=x+exp(x.^2);
    err_inf(p)=max(abs(u-y));
end
disp(' ')
disp('        h        inf_err      err/h^2      cputime     cputime/m  ')
disp('  ------------------------------------------------------------------')
disp(' ')
disp([h err_inf err_inf./h.^2 times times./m])
```

The results are:

```
       h        inf_err      err/h^2      cputime     cputime/m
  ------------------------------------------------------------------

   5.0000e-01   6.8077e-02   2.7231e-01   7.5100e-04   7.5100e-04
   2.5000e-01   2.0012e-02   3.2019e-01   1.8200e-04   6.0667e-05
   1.2500e-01   5.4792e-03   3.5067e-01   1.5800e-04   2.2571e-05
   6.2500e-02   1.3851e-03   3.5458e-01   1.4500e-04   9.6667e-06
   3.1250e-02   3.4860e-04   3.5697e-01   2.2400e-04   7.2258e-06
   1.5625e-02   8.7213e-05   3.5722e-01   2.5800e-04   4.0952e-06
   7.8125e-03   2.1807e-05   3.5729e-01   3.8300e-04   3.0157e-06
   3.9062e-03   5.4521e-06   3.5731e-01   6.4600e-04   2.5333e-06
   1.9531e-03   1.3630e-06   3.5731e-01   1.1700e-03   2.2896e-06
   9.7656e-04   3.4076e-07   3.5731e-01   2.2510e-03   2.2004e-06
   4.8828e-04   8.5188e-08   3.5730e-01   5.2390e-03   2.5594e-06
   2.4414e-04   2.1303e-08   3.5740e-01   8.6270e-03   2.1067e-06
   1.2207e-04   5.3250e-09   3.5736e-01   1.7026e-02   2.0786e-06
   6.1035e-05   5.2771e-09   1.4166e+00   3.4211e-02   2.0882e-06
   3.0518e-05   7.8709e-10   8.4513e-01   6.7547e-02   2.0614e-06
   1.5259e-05   2.9901e-10   1.2842e+00   1.3695e-01   2.0897e-06
   7.6294e-06   1.4168e-09   2.4340e+01   2.6656e-01   2.0337e-06
   3.8147e-06   3.9363e-08   2.7050e+03   5.3246e-01   2.0312e-06
   1.9073e-06   3.6684e-07   1.0084e+05   1.0608e+00   2.0234e-06
   9.5367e-07   8.2977e-07   9.1234e+05   2.1497e+00   2.0501e-06
```

We can see from the err/h^2 column that the expected $O(h^2)$ error is observed until $h \approx 1.22707e-4$ since err/h^2 rapidly approaches a constant. But then we lose accuracy. Why? Roundoff error begins to dominate! Thus, while theoretically as $h \to 0$ we have convergence, floating point errors in the form of roundoff error eventually dominates! Note, however, that the cputime stills scales linearly with m as evidenced by the $cputime/m$ column.

Note: I have only timed the linear solver portion of the code.