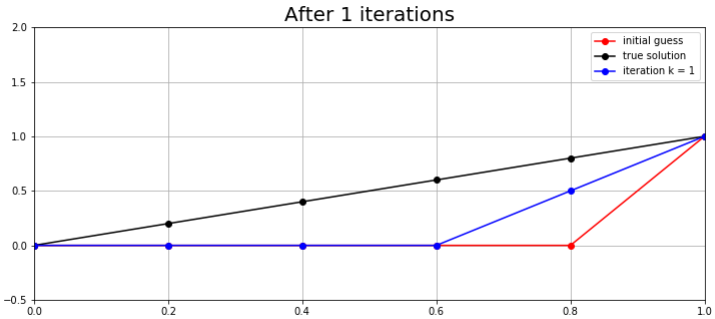
AMath 585
Homework #6
Due Thursday, March 12, 2020

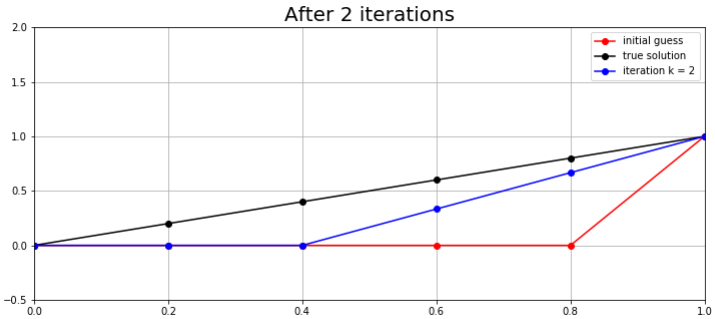
Homework is due to Canvas by 11:00pm PDT on the due date.

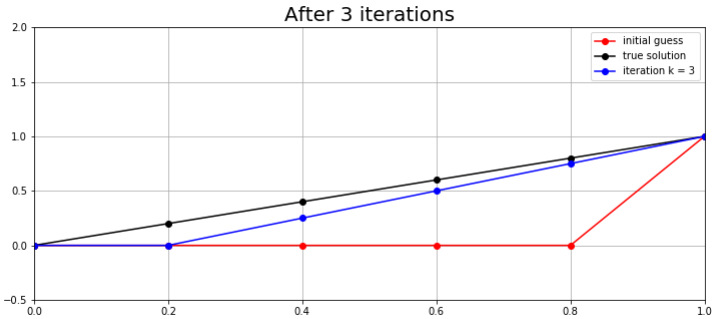
To submit, see `https://canvas.uw.edu/courses/1352870/assignments/5284853`

Problem 1. Consider the BVP $u''(x) = 0$ on $0 \le x \le 1$ with Dirichlet boundary conditions $u(0) = 0$ and $u(1) = 1$. The exact solution is $u(x) = x$.

Discretize with the standard centered approximation using m equally spaced interior points. If we apply the Conjugate-Gradient method with initial data $u_i^{[0]} = 0$ for $i = 1, 2, \ldots, m$ then we see the sort of behavior that is illustrated in the plots below for the case $m = 4$. For $k < m$ the approximate solution is always piecewise linear and has $u_i^{[k]} = 0$ for $i \le m - k$. After m iterations, $u^{[m]}$ is equal to the exact solution.







(a) For the case $m = 3$, work through the C-G algorithm by hand to explicitly calculate the vectors $r^{[k]}$, $b^{[k]}$, and $u^{[k]} \in \mathbb{R}^3$ in each iteration. This should help you see why the behavior seen in the plots makes sense.

(b) To show this behavior is seen for general m, show by induction that each residual $r^{[k]}$ is a unit vector (all zeros except in one element). Hint: Use the fact that we know that all the residuals generated in C-G are pairwise orthogonal to one another, and that the only elements that can change from one iteration to the next are those in which the search direction $b^{[k]}$ has nonzero components, which can also be determined in general.

(c) Explain how the result of (b) implies the behavior seen in the plots.

Solution:

(a)

$$Au = f$$

With the discretization and m = 3:

$$A = \begin{bmatrix} -32 & 16 & 0 \\ 16 & -32 & 16 \\ 0 & 16 & -32 \end{bmatrix}$$

$$f = \begin{bmatrix} 0 \\ 0 \\ -16 \end{bmatrix}$$

$$u^{[0]} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Iteration 1:

$$r_0 = \begin{bmatrix} 0 \\ 0 \\ -16 \end{bmatrix} - \begin{bmatrix} -32 & 16 & 0 \\ 16 & -32 & 16 \\ 0 & 16 & -32 \end{bmatrix} * \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -16 \end{bmatrix}$$

$$p_0 = \begin{bmatrix} 0 \\ 0 \\ -16 \end{bmatrix}$$

$$\omega_0 = \begin{bmatrix} -32 & 16 & 0 \\ 16 & -32 & 16 \\ 0 & 16 & -32 \end{bmatrix} * \begin{bmatrix} 0 \\ 0 \\ -16 \end{bmatrix} = \begin{bmatrix} 0 \\ -256 \\ 512 \end{bmatrix}$$

$$\alpha_0 = \frac{r_0^T r_0}{p_0^T \omega_0} = \frac{-1}{32}$$

$$u_1 = u_0 + \alpha_0 p_0 = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{2} \end{bmatrix}$$

$$r_1 = r_0 - \alpha_0 \omega_0 = \begin{bmatrix} 0 \\ -8 \\ 0 \end{bmatrix}$$

$$\beta_0 = \frac{r1^T r1}{r_0^T r_0} = \frac{1}{4}$$

$$p_1 = r_1 + \beta_0 p_0 = \begin{bmatrix} 0 \\ -8 \\ -4 \end{bmatrix}$$

Iteration 2:

$$\omega_1 = Ap_1 = \begin{bmatrix} -128 \\ 192 \\ 0 \end{bmatrix}$$

$$\alpha_1 = \frac{r_1^T r_1}{p_1^T \omega_1} = \frac{-1}{24}$$

$$u_2 = u_1 + \alpha_1 p_1 = \begin{bmatrix} 0 \\ \frac{1}{3} \\ \frac{2}{3} \end{bmatrix}$$

$$r_2 = r_1 - \alpha_1 \omega_1 = \begin{bmatrix} \frac{16}{3} \\ 0 \\ 0 \end{bmatrix}$$

$$\beta_1 = \frac{r_2^T r_2}{r_1^T r_1} = \frac{4}{9}$$

$$p_2 = r_2 + \beta_1 p_1 = \begin{bmatrix} \frac{16}{3} \\ -\frac{32}{9} \\ -\frac{16}{9} \end{bmatrix}$$

Iteration 3:

$$\omega_2 = Ap_2 = \begin{bmatrix} \frac{1024}{9} \\ 0 \\ 0 \end{bmatrix}$$

$$\alpha_2 = \frac{r_2^T r_2}{p_2^T \omega_2} = \frac{-3}{64}$$

$$u_3 = u_2 + \alpha_2 p_2 = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \\ \frac{3}{4} \end{bmatrix}$$

$$r_3 = r_2 - \alpha_2 \omega_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\beta_2 = \frac{r_3^T r_3}{r_2^T r_2} = 0$$

$$p_3 = r_3 + \beta_2 p_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

(b) In general the discretization with the standard centered approximation with m interior points for the BVP $u''(x) = 0$ on $0 \leq x \leq 1$ is:

$$A(mxm) = \frac{1}{h^2} \begin{bmatrix} -2 & 1 & \dots & \\ 1 & -2 & 1\dots & \\ 0 & 1 & -2 & 1\dots \\ & & \ddots & \ddots \end{bmatrix}$$

and f has the form:

$$f = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \frac{1}{h^2} \end{bmatrix}$$

Additionally, the initial guess, u_0 is the the zero vector. In general, $p_0 = r_0 = f - Au_0$ so:

$$r_0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \end{bmatrix} - \frac{1}{h^2} \begin{bmatrix} -2 & 1 & \dots & \\ 1 & -2 & 1\dots & \\ 0 & 1 & -2 & 1\dots \\ & & \ddots & \ddots \end{bmatrix} * \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

This means that both r_0 and p_0 will have one non-zero element in the mth row. For induction, we must first show that r_1 is a multiple of a unit vector.

$$r_1 = f - Au_1 = f - Au_0 - \alpha_0 Ap_0 = r_0 - \alpha_0 Ap_0$$

where

$$\alpha_0 = \frac{r_0^T r_0}{p_0^T A p_0}$$

$$r_0^T r_0 = \frac{1}{h^4}$$

$$p_0^T A = \begin{bmatrix} 0 & 0 & \dots & \frac{-1}{h^2} \end{bmatrix} * -\frac{1}{h^2} \begin{bmatrix} -2 & 1 & \dots & \\ 1 & -2 & 1\dots & \\ 0 & 1 & -2 & 1\dots \\ & & \ddots & \ddots \end{bmatrix} = \begin{bmatrix} 0 & \dots & \frac{1}{h^4} & \frac{-2}{h^4} \end{bmatrix}$$

$$p_0^T A p_0 = \frac{2}{h^6}$$

$$\alpha_0 = -\frac{h^2}{2}$$

$$r_1 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \frac{-1}{h^2} \end{bmatrix} + \frac{h^2}{2} \begin{bmatrix} 0 \\ \vdots \\ -\frac{1}{h^4} \\ \frac{2}{h^4} \end{bmatrix}$$

$$r_1 = \begin{bmatrix} 0 \\ \vdots \\ -\frac{1}{2h^2} \\ 0 \end{bmatrix}$$

which is a scalar multiple of a unit vector, where the nonzero element was in the m-1th element of the vector. It is also important to note that since r_1 has this form:

$$p_1 = r_1 + \beta_0 r_0$$

which has the form:

$$p_1 = \begin{bmatrix} 0 \\ \vdots \\ * \\ * \end{bmatrix}$$

where the * represents nonzero entries. So, p_1 has the same nonzero entries in the same row as r_1 and below.

Now, to assume that it is true for the kth iteration: If it is true for the kth iteration, then the nonzero element will be in the m-kth row of r. We want to show that it will move to the m - (k+1)th element.

Let

$$r_k = \begin{bmatrix} 0 \\ \vdots \\ * \\ \vdots \\ 0 \end{bmatrix}, p_k = \begin{bmatrix} 0 \\ \vdots \\ * \\ \vdots \\ * \end{bmatrix}$$

$$r_{k+1} = r_k - \alpha_k * A * p_k$$

Since A is tridiagonal, when you multiply A by p_k, it adds one nonzero element above the previous nonzero element, and since α is a scalar multiple it does not change the number of nonzero entries, which gives you:

$$r_k = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ * \\ \vdots \\ 0 \end{bmatrix}, \alpha_k A p_k = \begin{bmatrix} 0 \\ \vdots \\ * \\ * \\ \vdots \\ * \end{bmatrix}$$

Thus, when you form the next r_{k+1} it has the form:

$$r_{k+1} = \begin{bmatrix} 0 \\ \vdots \\ * \\ * \\ \vdots \\ * \end{bmatrix}$$

Now, we know that r_k is orthogonal to r_{k+1} and r_k has a nonzero element in the m-kth row and r_{k+1}has nonzero elements in the m-(k+1)th through mth row.
In order for these two vectors to be orthogonal, all of the entries in $r + k + 1$ must be zero except for the k-(m+1)th element.

Thus we conclude that r_{k+1} is has only one nonzero element if r_k only has one nonzero element and we have shown that the p_k direction vectors have a specific form. Thus, it is true for every iteration k by induction.

(c) From the plots above, you can see that the iterations converge toward the true solution one point at a time. The internal point that moves is the nonzero element of the r residual vector. This makes sense the the residual is Au - f and we are assuming that we start at the zero vector.

Problem 2. Consider a linear system $Au = f$ in which the matrix A is **not** symmetric positive definite, so C-G cannot be applied directly.

(a) Show that if A is nonsingular then the matrix $B = A^T A$ is symmetric positive definite.

(b) So one approach to solving $Au = f$ is multiply both sides by A^T to get $Bu = A^T f$ and then solve this system with C-G. The problem with this approach is that the condition number increases. Show that the 2-norm condition number of B is the square of the 2-norm condition number of A.

(c) On page 93 it is noted that applying C-G to the two-dimensional Poisson problem $Au = f$ on an m by m grid (with second order centered differencing) requires $O(m^3)$ work to converge to a fixed tolerance. Suppose we multiplied both sides by A^T as described above (even though not necessary here since A is already SPD) and solved the resulting system (which is still SPD) by C-G. What order of work would now be required to reach a fixed tolerance?

(d) Given that the global error for this discretization is $O(h^2) = O(1/m^2)$ for smooth solutions, it makes more sense to look at the work required to get the error in the C-G solution down to this level. How does this change the work estimates given above, both for solving $Au = f$ and $A^T Au = A^T f$?

Note: there are better approaches for nonsymmetric matrices than the approach described above that do not magnify the condition number, see Section 4.4 and other references.

Solution:

(a) Assume that A is nonsingular, which means that $|\lambda| > 0$

$$x^T Bx = x^T A^T Ax = (Ax)^T (Ax) > 0 \ \forall x \neq 0$$

Thus, $B = A^T A$ is SPD.
(b) We know that for some matrix M:

$$\|M\| = max\ \sigma = \sigma_1 = \sqrt{max\ \lambda\ of\ M^T M}$$

If M is symmetric positive definite, then

$$\|M\| = max\ \lambda = \lambda_1 = \sqrt{max\ \lambda\ of\ M^T M}$$

Since B is SPD

$$\kappa(B) = \frac{\sigma_1(B)}{\sigma_m(B)} = \frac{\lambda_1(B)}{\lambda_m(B)}$$

Since $B = A^T A$,

$$\kappa(B) = \frac{\lambda_1(B)}{\lambda_m(B)} = \sqrt{\frac{\lambda_1(A)}{\lambda_m(A)}} = \sqrt{\kappa(A)}$$

and thus:

$$\kappa(B) = \kappa(A)^2$$

(c) From page 93, the number of iterations required to reach a desired tolerance is:

$$k = O(\sqrt{\kappa})$$

Where the condition number $\kappa = O(\frac{1}{h^2})$
If you multiply by A^T, then from part (b), we know that this squares the condition number.

$$\kappa = O(\frac{1}{h^4})$$

So now, CG will require $O(m^2)$ iterations to converge (previously O(m)) Each iteration requires $O(m^2)$ work, so in total, it would take $O(m^4)$

(d) From the book, the error bound for the Conjugate Gradient method is: $2e^{\frac{-2k}{\sqrt{\kappa}}}$ if κ is large. If we want the error to be on the order of $\frac{1}{h^2}$ then,

$$2e^{\frac{-2k}{\sqrt{\kappa}}} \geq h^2$$

$$e^{\frac{-2k}{\sqrt{\kappa}}} \geq \frac{h^2}{2}$$

Since$h \geq \frac{1}{m^2}$

$$e^{\frac{-2k}{\sqrt{\kappa}}} \geq \frac{1}{2m^2}$$

$$\ln e^{\frac{-2k}{\sqrt{\kappa}}} \geq \ln \frac{1}{2m^2}$$

$$\frac{-2k}{\sqrt{\kappa}} \geq \ln(1) - \ln(2m^2)$$

$$k \geq \sqrt{\kappa} \ln(2m)$$

Looking at $Au = f$ We know that the condition number $\kappa = O(\frac{1}{h^2}) = O(m^2)$ Thus the number of iterations is about:

$$k \geq \sqrt{m^2} \ln(2m) \sim O(mlog(m))$$

Since there is m^2 work per iteration, the total work is:

$$O(m^3 log(m))$$

Looking at $A^T Au = f$ We know that the condition number $\kappa = O(\frac{1}{h^4}) = O(m^4)$ Thus the number of iterations is about:

$$k \geq \sqrt{m^4} \ln(2m) \sim O(m^2 log(m))$$

Since there is m^2 work per iteration, the total work is:

$$O(m^4 log(m))$$

Thus, to bring the error down to this level, it multiplies the amount of work by $log(m)$

Problem 3. As in the previous homework, consider the one-dimensional BVP

$$\frac{d}{dx}\left(\kappa(x)u'(x)\right) = 0$$

on $0 \leq x \leq 1$ with Dirichlet boundary conditions $u(0) = 0$ and $u(1) = 1$, again discretizing this problem using the system (2.71) in the text. (Or negate it if you prefer, to make it positive definite.)

Consider the piecewise constant diffusivity

$$\kappa(x) = \begin{cases} \epsilon & \text{if } x < 0.5, \\ 1 & \text{if } x > 0.5. \end{cases}$$

where $\epsilon > 0$.

(a) Generalizing what you did in HW5, determine the exact solution, in terms of the parameter ϵ.

(c) Implement the conjugate gradient method for this problem. For the convergence test require $\|r^{[k]}\|_2 < 10^{-14}$. Allow more than m iterations, if necessary.

Make semilogy plots of the max-norm of the error and the 2-norm of the residual as a function of iteration k for the case $m = 19$ with $\epsilon = 0.1$. Also try $\epsilon = 10^{-3}$. You should observe that more than m iterations are required to get good results. Comment on the behavior of the iterates in each case.

(d) Implement the preconditioned C-G algorithm (PCG) using the diagonal preconditioner and observe that this greatly improves the convergence behavior.

Note: Make sure you do this in a way for which M is symmetric positive definite and not negative definite, as discussed in the notebook `PCG.ipynb` and video that goes with it. This also contains corrections to some typos in the PCG algorithm written on page 95.

The notebook `DarcyFlow.ipynb` provides an implementation of the PCG algorithm for the two dimensional version of this problem that may be useful to follow.

Solution:

(a)

$$\frac{d}{dx}\left(\kappa(x)u'(x)\right) = 0$$
$$\int \frac{d}{dx}\left(\kappa(x)u'(x)\right) = \int 0$$
$$\left(\kappa(x)u'(x)\right) = C$$

First assume $\kappa(x) = 0.1$, $x < 0.5$

$$\int \left(\epsilon u'(x)\right) = \int C$$
$$\epsilon u(x) = Cx + D$$
$$u(x) = \frac{C}{\epsilon}x + \frac{D}{\epsilon}$$
$$u(0) = \frac{D}{\epsilon} = 0$$
$$u(x) = \frac{C}{\epsilon}x$$

Now assume $\kappa(x) = 1$, $x > 0.5$

$$\int \left(u'(x)\right) = \int C$$
$$u(x) = Cx + B$$
$$u(1) = C + B = 1$$
$$u(x) = Cx + 1 - C$$

Since we know that the solution is continuous, the solution must be the same at x = 0.5

$$\frac{C}{\epsilon}(0.5) = C(0.5) + 1 - C$$
$$\frac{C + C\epsilon}{2\epsilon} = -1$$
$$\frac{1+\epsilon}{2\epsilon}C = 1$$
$$C = \frac{2\epsilon}{1+\epsilon}$$

$$u(x) = \begin{cases} \frac{2}{1+\epsilon}x & x \leq 0.5 \\ \frac{2\epsilon}{1+\epsilon}x + \frac{1-\epsilon}{1+\epsilon} & x \geq 0.5 \end{cases}$$

(b)/(c)
These questions are answered in the Jupyter notebook.

Problem 4. Consider the same problem as in Problem 3 but now on the interval $0 \le x \le 4$ with Dirichlet boundary conditions and with $m = 3$ internal grid points (so $h = 1$ for convenience). Now put the jump in κ at the midpoint $x = 2$:

$$\kappa(x) = \begin{cases} \epsilon & \text{if } x < 2, \\ 1 & \text{if } x > 2. \end{cases}$$

(a) Write out the 3×3 matrix A explicitly in this case.

(b) Write out the matrix M that would be used as the "diagonal preconditioner" in this case. Also compute $B = M^{-1}A$ and observe that it is not symmetric.

(c) In this case we can choose C to be $\text{diag}(\sqrt{M_{ii}})$. Write out the matrix $\tilde{A} = C^{-T}AC^{-1}$ in this 3×3 case and observe that it is symmetric.

(d) For the case $\epsilon = 10^{-4}$ compute the eigenvalues and 2-norm condition number of A and B (recall that those of $\tilde{A}$ agree with those of B, but B is easier to work with). You can use the `eig` function in Numpy or Matlab, or do it by hand.

(e) Note that as $\epsilon \to 0$ the matrix A approaches a singluar matrix and the condition number blows up. What does the condition number of B approach as $\epsilon \to 0$? (You should be able to compute this analytically by looking at the limiting matrix.)

Solution:

(a): Make sure that you negate everything to make it SPD instead of SND.

$$A = \begin{bmatrix} 2\epsilon & -\epsilon & 0 \\ -\epsilon & (1+\epsilon) & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

(b): The diagonal pre-conditioner is diag(A)

$$M = \begin{bmatrix} 2\epsilon & 0 & 0 \\ 0 & (1+\epsilon) & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$B = M^{-1}A = \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ \frac{\epsilon}{\epsilon+1} & 1 & \frac{\epsilon}{\epsilon+1} \\ 0 & -\frac{1}{2} & 1 \end{bmatrix}$$

This is not symmetric

(c):

$$C = diag(\sqrt{M}) = \begin{bmatrix} \sqrt{2\epsilon} & 0 & 0 \\ 0 & \sqrt{(1+\epsilon)} & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix}$$

$$\tilde{A} = C^{-T}AC^{-1} = \begin{bmatrix} 1 & \frac{-\epsilon}{\sqrt{2\epsilon(1+\epsilon)}} & 0 \\ \frac{-\epsilon}{\sqrt{2\epsilon(1+\epsilon)}} & 1 & \frac{-1}{\sqrt{2(1+\epsilon)}} \\ 0 & \frac{-1}{\sqrt{2(1+\epsilon)}} & 1 \end{bmatrix}$$

(d):
The eigenvalues of A are:
0.0002, 0.3820, 2.6181
The 2 norm condition number of A is:
1.3092e+04
The eigenvalues of B are:
1.0000, 0.2929,1.7071
The 2 norm condition number of B is:
7.0063 (from MATLAB)
The eigenvalues of $\tilde{A}$ are:
1.0000, 0.2929,1.7071
which are the same as B
The 2 norm condition number of $\tilde{A}$ is:
5.8284
Since $\tilde{A}$ is symmetric, the 2 norm condition number is the largest eigenvalue over the smallest eigenvalue. Although the eigenvalues for B and $\tilde{A}$ are the same, the condition numbers are different because B is not symmetric and therefore the condition number in the 2-norm does not equal the ratio of largest and smallest eigenvalues.

(e):
As A becomes a singular matrix, one of the eigenvalues approaches 0, and thus, the condition number approaches infinity.

For $\tilde{A}$ as $\epsilon \to 0$, the matrix has the form:

$$\tilde{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{-1}{2} \\ 0 & -\frac{1}{2} & 1 \end{bmatrix}$$

which has a condition number of about 5.8284.