Chapter III: Conditioning and Stability

Aili Shao

1 Lecture 12 Conditioning and Stability (28/06/2018)

Exercise 1.1. *By the formula derived in the book, we know that*

$$\kappa(A) = \frac{\sigma_1}{\sigma_{202}}.$$

We assume that A is non-singular, or otherwise, $\kappa(A) = \infty$.
By Theorem 5.3, we know that $\|A\|_2 = \sigma_1$ and $\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_{202}^2}$. $\|A\|_2 = 100$ implies that $\sigma_1 = 100$. $\|A\|_F = 101$ implies that $\sigma_2^2 + \cdots + \sigma_{202}^2 = 101^2 - 100^2 = 201$. By the non-decreasing property of singular values, we know that $0 \leq \sigma_{202} \leq \sigma_{201} \leq \cdots \leq \sigma_1$. Thus, $\sigma_{202} \leq \sqrt{201/(202-1)} = 1$. This shows that

$$\kappa(A) = \frac{\sigma_1}{\sigma_{202}} \geq 100.$$

Exercise 1.2. *(a) Suppose that the degree $n-1$ polynomial is of the following form*

$$p(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_{n-1} x^{n-1}.$$

Let A_x be the Vandermonde matrix of x, that is

$$A_x = \begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-1} \\ 1 & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{bmatrix}.$$

Then

$$A_x \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}.$$

Let A_y be the Vandermonde matrix of x, that is

$$A_y = \begin{bmatrix} 1 & y_1 & \cdots & y_1^{n-1} \\ 1 & x_2 & \cdots & y_2^{n-1} \\ 1 & \vdots & \ddots & \vdots \\ 1 & y_n & \cdots & y_n^{n-1} \end{bmatrix}.$$

Then

$$A_y \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix} = \begin{bmatrix} p(y_1) \\ p(y_2) \\ \vdots \\ p(y_m) \end{bmatrix}.$$

Since $A \in \mathbb{C}^{m\times n}$ maps an n-vector of data at $\{x_i\}$ to an m-vector of sampled values $\{p(y_j)\}$ where p is a polynomial of degree $n-1$, we have

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} p(y_1) \\ p(y_2) \\ \vdots \\ p(y_m) \end{bmatrix}.$$

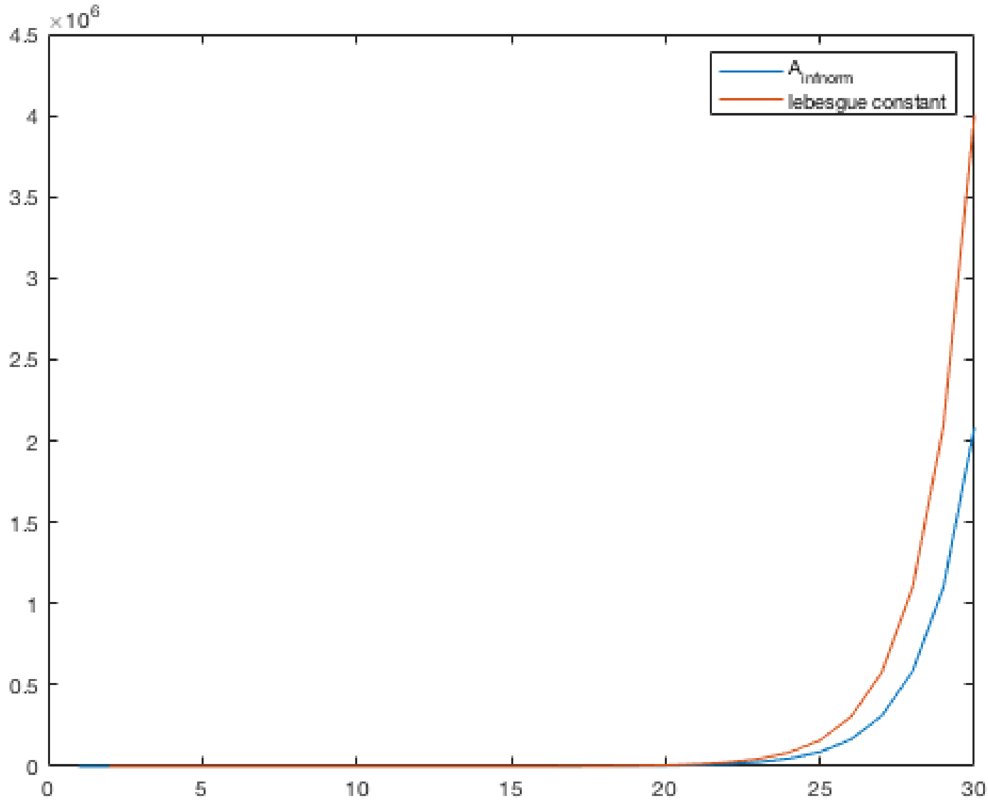
That is $p(y_i) = x_1 a_{i1} + x_2 a_{i2} + \cdots + x_n a_{in}$. Thus, $A = A_y A_x^{-1}$.
(b)

```
  function [A,A_inf] = A_and_infnorm(n)
% This function generate the m-by-n matrix A which maps an n-vector data at
% {x_j} ti an m-vector of sampled values {p(y_j)} where p is the degree n-1
% polynomial interpolant of the data.

% Note that {x_j} and {y_j} are equispaced points from [-1, 1]
m=2*n-1;
x=linspace(-1,1,n)';
y=linspace(-1,1,m)';
Ay=zeros(m,n);
Ax=zeros(n,n);

for i=1:n;
    Ay(:,i)=y.^(i-1);
    Ax(:,i)=x.^(i-1);
end
A=Ay*Ax^(-1);
A_inf=norm(A,inf);
end
```

```
  A_infnorm=[];
for n=1:30;
   [A, A_inf]=A_and_infnorm(n);
   A_infnorm=[ A_infnorm A_inf];
end
nn=1:30;
figure(1);clf
plot(nn,A_infnorm);
hold on;
lebesgue_f=2.^nn./(exp(1)*(nn-1).*log(nn))
semilogy(nn,lebesgue_f);
legend('A_{infnorm}', 'lebesgue constant')
```



(c)

$$A_x \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

This implies $c_0 = 1, c_i = 0$ *for* $1 \leq i \leq n-1$.
Also

$$A_y \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

By (12.6), we know that

$$\kappa = \frac{\|J(x)\|_\infty}{\|f(x)\|_\infty / \|x\|_\infty} = \|J(x)\|_\infty = \|A\|_\infty.$$

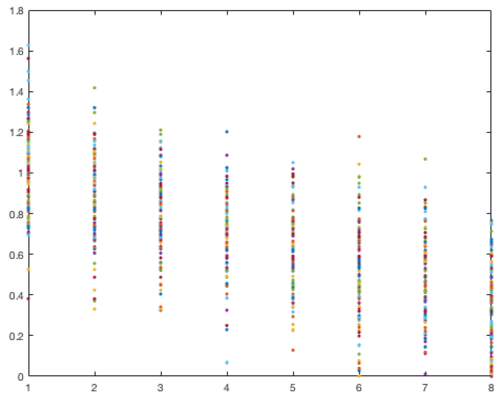
For $n = 1, 2, \cdots 30, m = 2n-1$, *the* ∞*-norm condition number is*

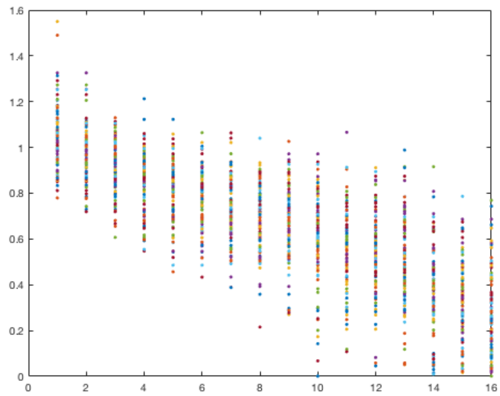
```
1.0e+06 *
   0.000001000000000
   0.000001000000000
   0.000001250000000
   0.000001625000000
   0.000002171875000
   0.000002992187500
   0.000004263671875
   0.000006293945313
   0.000009619323730
   0.000015183441162
   0.000024660987854
   0.000041047313690
   0.000069737399578
   0.000120509180783
   0.000211184145869
   0.000374409046097
   0.000670263206741
   0.001209770202991
   0.002198873910791
   0.004020914016712
   0.007391694568172
   0.013651721403415
   0.025318140580175
   0.047129263899626
   0.088025258439049
   0.164909275124768
   0.309806524714292
   0.583512768723697
   1.101470543059379
   2.084477422390027
```

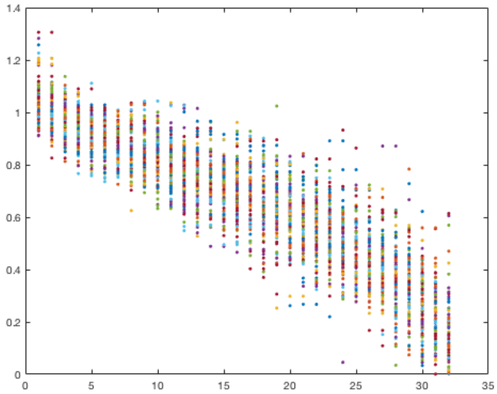
(d) The condition number at $n = 11$ is 24.660987854005842. From Figure 11.1, we see that the bound is approximately 4. Hence our answer is a bit far from the implicit bound.

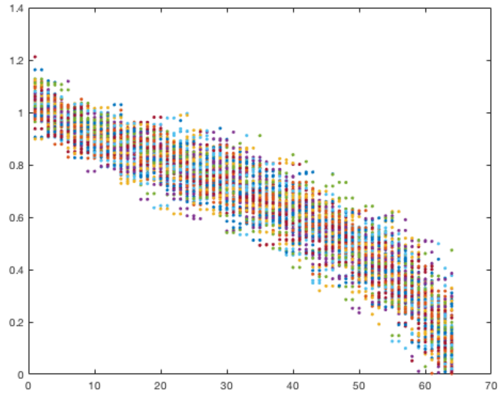
Exercise 1.3. *(a) The code is adapted from cs.dartmouth homework solution:*

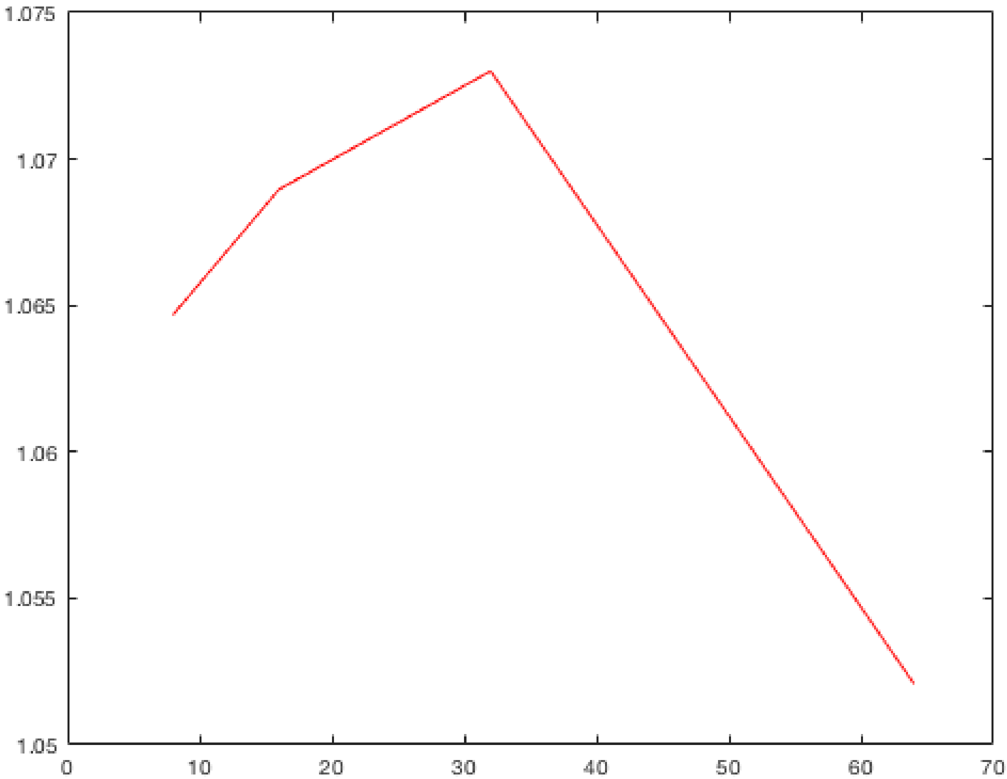
```
spec_mean=zeros(4,1);
for j=3:6
    figure;clf;
    spec=zeros(100,1);
    m=2.^j;
    eigs = zeros(100,m);
    for i=1:100
        A = randn(m,m)/sqrt(m);
         eigs(i,:)=eig(A)';
        plot([1:m],abs(eigs(i,:)),'.');
        spec(i) = max(abs(eigs(i,:)));
        hold on;
    end
spec_mean(j-2) = mean(spec);
end
figure;
plot(2.^[3:6],spec_mean,'r-');
```







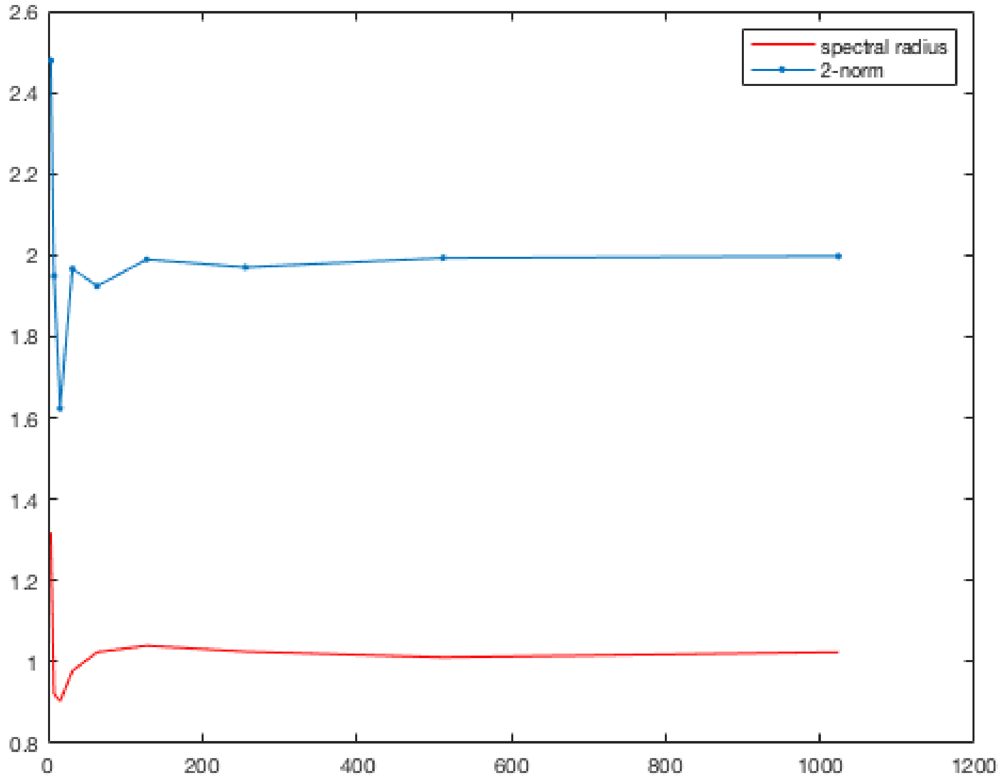




The spectral radius tends to 1 *as* $m \to \infty$. *(b)*

```
spectral=[];
norm2 =[];
figure;clf;
for j=2:10
    m=2.^j;
```

```
    A = randn(m,m)/sqrt(m);
    spec=max(abs(eig(A)));
    spectral=[spectral spec];
    normnew=norm(A);
    norm2=[norm2 normnew];
end
plot([2.^(2:10)], spectral, 'r-');
hold on;
plot ([2.^(2:10)], norm2, '.-');
legend('spectral radius', '2-norm')
diff=[norm2-spectral]'
```

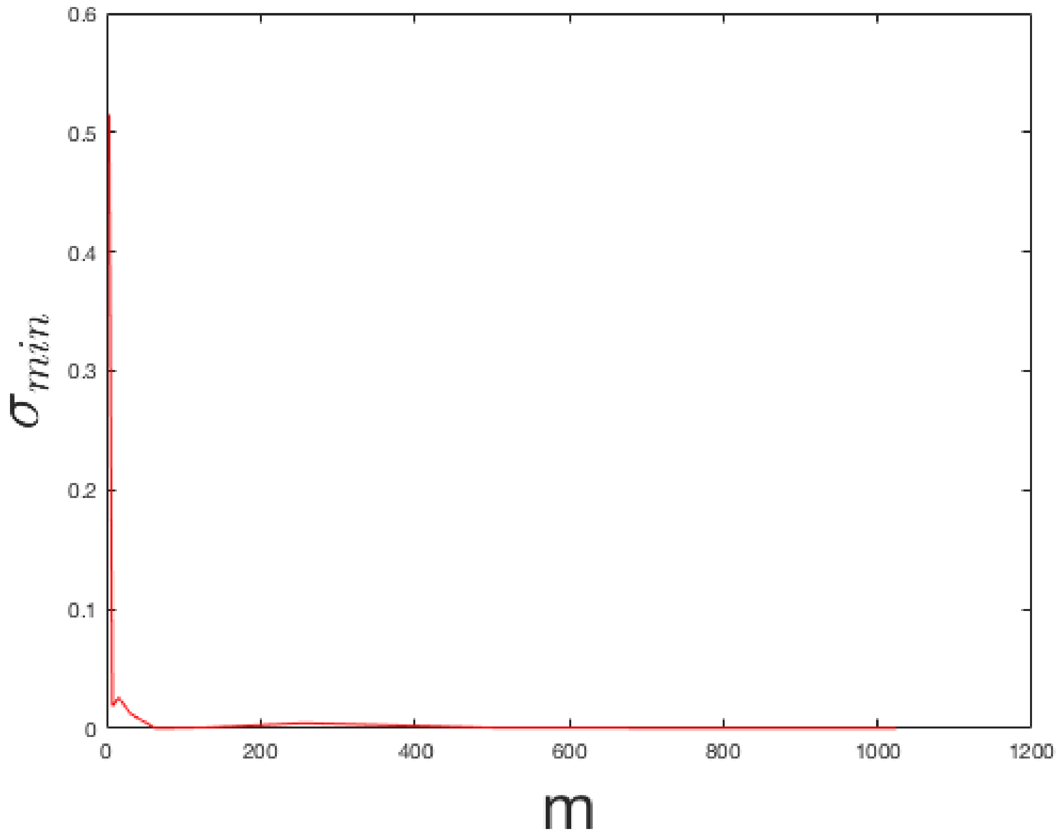


It shows that for a random matrix, the spectral radius does not approach to the 2-norm as $m \to \infty$.

(c)

```
svds=[];
for j=2:10
    m=2.^j
    A = randn(m,m)/sqrt(m);
```

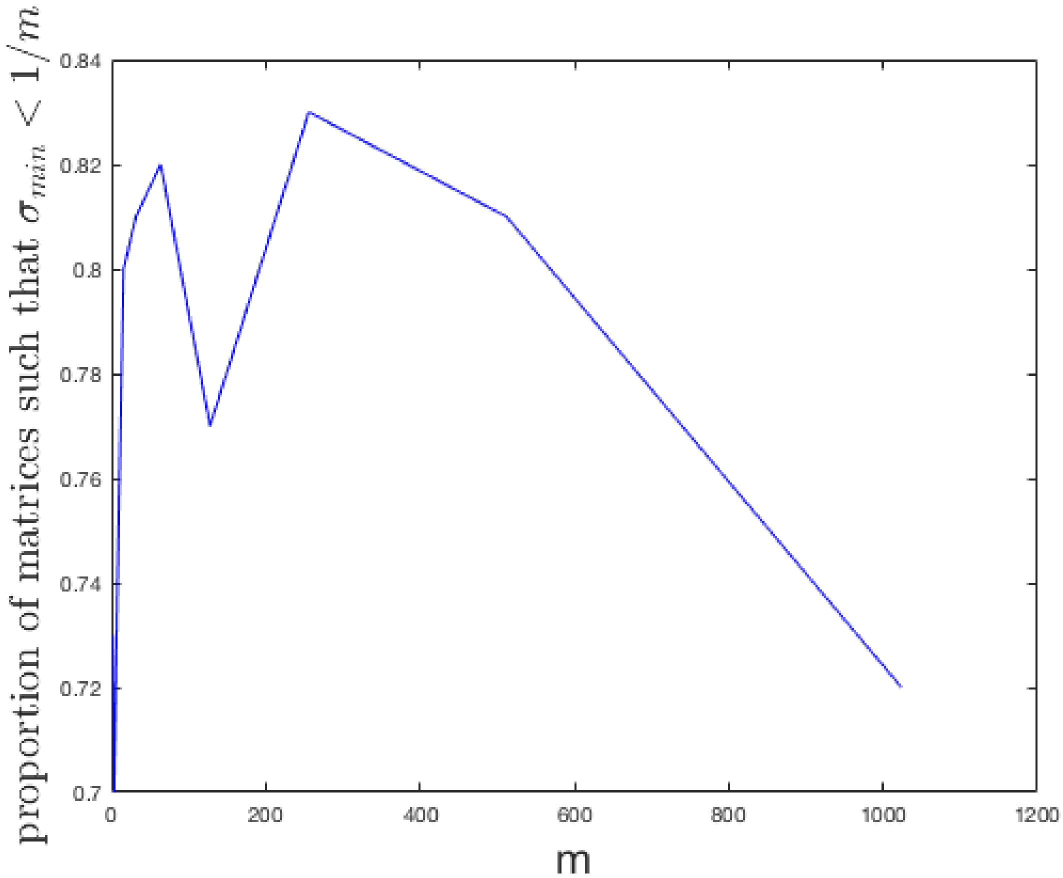
```
    singval=min(svd(A));
    svds=[svds singval]
end
figure;clf
plot([2.^(2:10)],svds, 'r')
xlabel('m','FontSize', 30)
ylabel('$\sigma_{min}$','Interpreter','latex','FontSize', 30)
```



It shows that σ_{min} decreases as the size of matrix increases.

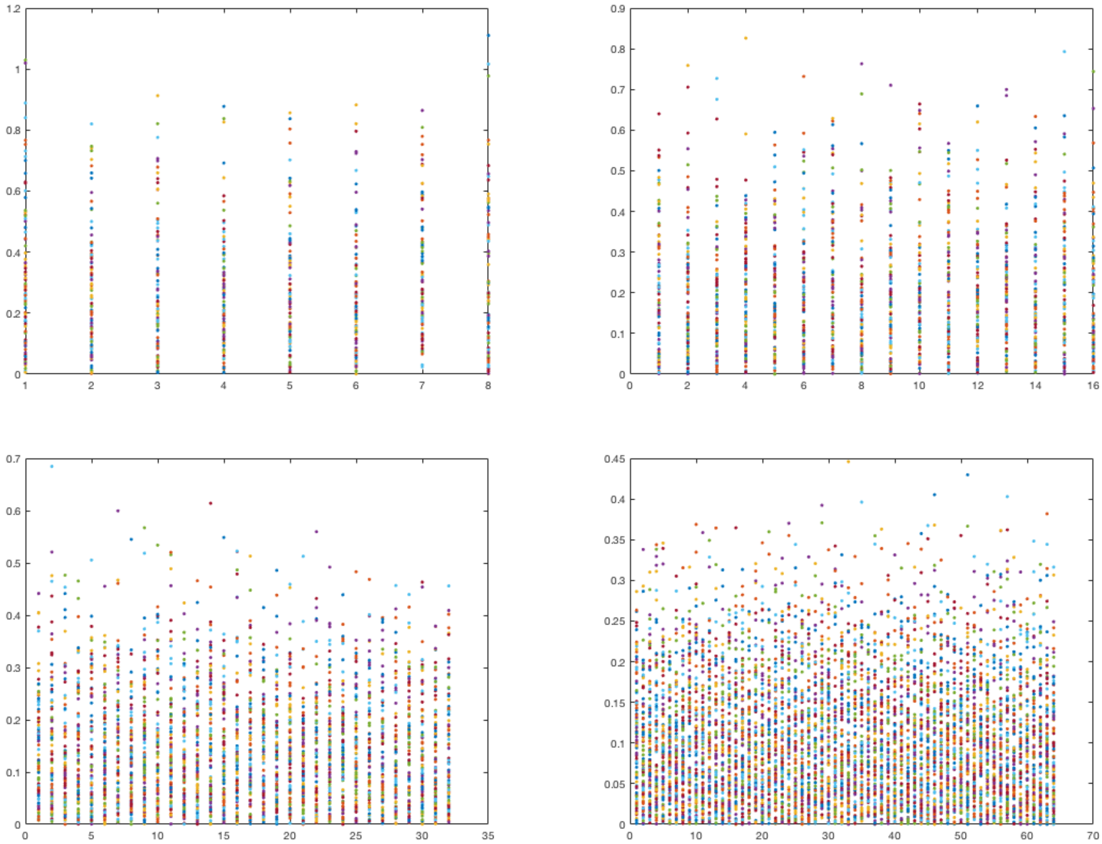
```
j=1:10;
prop=[];
for m=2.^j;
svds=zeros(100,1);
for i=1:100;
    A = randn(m,m)/sqrt(m);
    svds(i)=min(svd(A));
end
propnew=length(find(svds< 1/m))/100;
```

```
prop=[prop propnew];
end
figure; clf;
plot(2.^(1:10), prop, 'b')
xlabel('m','FontSize', 20)
ylabel('proportion of matrices such that $\sigma_{min}<1/m$','Interpreter','latex','FontSi
```

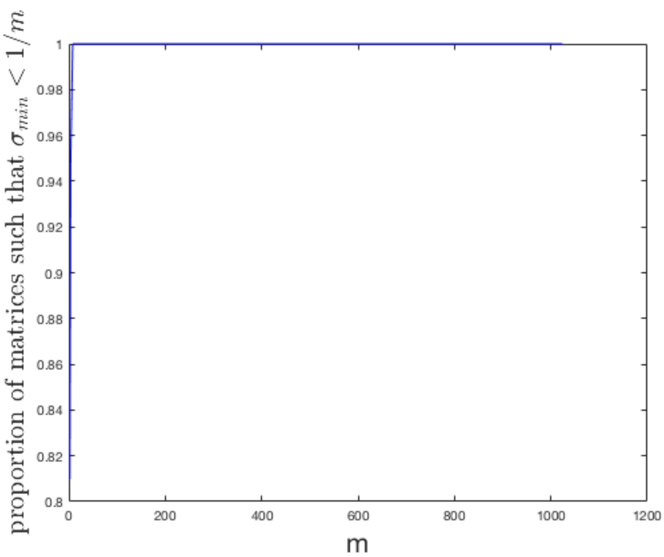


The plot shows that the proportions of random matrices such that $\sigma_{min} < 1/m$ decreases as m increases.

(d)



For (a)-(b), the results are not too different expect that the scatter plots of the eigenvalues are more uniformly distributed (different from the diagonally clustered plots). The result for (c) is very different, as the proportions of matrices satisfying the inequality become



to 1 as m increases.

2 Lecture 13 Conditioning and Stability (29/06/2018)

Exercise 2.1. *In IEEE single precision arithmetic, the gaps between adjacent numbers are* $2^{-23} \times 2^j$ *on the interval* $[2^j, 2^{j+1}]$. *The gap between an adjacent pair of nonzero IEEE double precision real numbers is* $2^{-52} \times 2^j$ *in this case. Thus, there are*

$$\frac{2^{-23} \times 2^j}{2^{-52} \times 2^j} - 1 = 2^{29} - 1$$

IEEE double precision numbers.

Exercise 2.2. *The elements of* F *are the numbers* 0 *together with all numbers of the form*

$$x = \pm(m/\beta^t)\beta^e$$

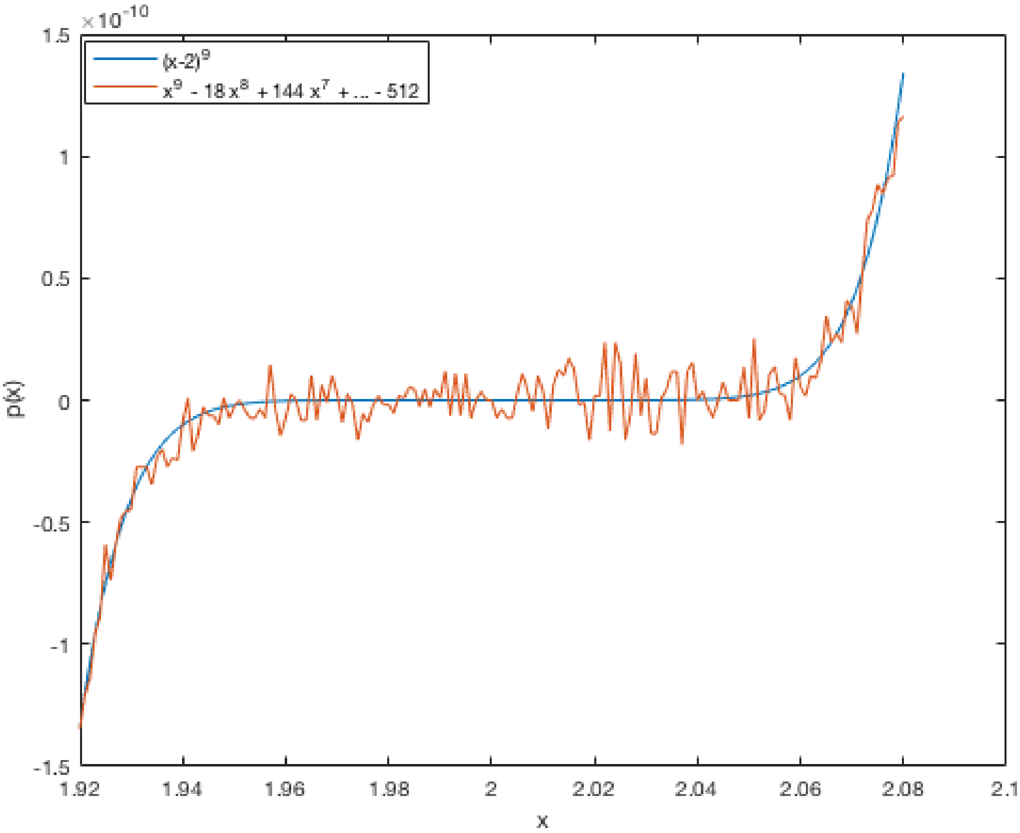
where m *is an integer in the range* $1 \leq m \leq \beta^t$ *and* e *is an arbitrary integer.*
(a)It is the number $2^t + 1$ *where* t *is the precision (* 24 *and* 53 *for IEEE single and double precision, respectively).*
(b) $n = 2^24 + 1$ *for IEEE single precision arithmetic while* $n = 2^53 + 1$ *for IEEE double precision arithmetic.*
(c) Construct 5 *consecutive integers* $[2^t - 2, 2^t - 1, 2^t, 2^t + 1, 2^t + 2]$ *and compute the differences between the adjacent numbers.*
If all the numbers were represented correctly all the differences should equal exactly 1. *The output of the following code demonstrates that this is not the case.*

```
t = 53;
n = 2^t + (-2:2)
diff(n)
n =
  1.0e+015 *
    9.0072    9.0072    9.0072    9.0072    9.0072
ans =
     1     1     0     2
```

Exercise 2.3. `x = (1.920:.001:2.080)';`

```
p1 = (x-2).^9;
p2 = x.^9 - 18*x.^8 + 144*x.^7 - 672*x.^6 + 2016*x.^5 - 4032*x.^4 + 5376*x.^3 - 4608*x.^2

figure(2);clf;
plot(x,[p1 p2]);
xlabel('x'); ylabel('p(x)');
legend('(x-2)^9', 'x^9 - 18 x^8 + 144 x^7 + ... - 512')
```



Exercise 2.4. *Recall that by Newton's method,*

$$x_{k+1} = x_k - \frac{p(x_k)}{p'(x_k)}.$$

(a) Note that

$$p(x) = x^5 - 2x^4 - 3x^3 + 3x^2 - 2x - 1,$$
$$p'(x) = 5x^4 - 8x^3 - 9x^2 + 6x - 2.$$

Since $x_0 = 0,$, *we have*

$$x_1 = (0 - \frac{p(x_0)}{p'(x_0)}(1+\varepsilon))(1+\varepsilon) = -0.5000000000000000.$$

$$x_2 = (x_1 - \frac{p(x_1)}{p'(x_1)}(1+\varepsilon))(1+\varepsilon) = -0.336842105263158.$$

$$x_3 = (x_2 - \frac{p(x_2)}{p'(x_2)}(1+\varepsilon))(1+\varepsilon) - = 0.315728448396289.$$

$$x_4 = (x_3 - \frac{p(x_3)}{p'(x_3)}(1+\varepsilon))(1+\varepsilon) = -0.315301162703277.$$

$$x_5 = (x_4 - \frac{p(x_4)}{p'(x_4)}(1+\varepsilon))(1+\varepsilon) = -0.315300986459363.$$

$$x_6 = (x_5 - \frac{p(x_5)}{p'(x_5)}(1+\varepsilon))(1+\varepsilon) = -0.315300986459333.$$

Note that we also expect rounding errors when we calculate $p(x_k)$ *and* $p'(x_k)$ *each time.* 16 *digits are correct in each of these numbers.*

```
syms p(x)
p(x) = x^5-2*x^4-3*x^3+3*x^2-2*x-1;
dp = diff(p,x);

x=zeros(7,1);
for i=1:6;
    x(i+1)=x(i)-p(x(i))/dp(x(i));
end
```