Exercises

SC2_11 – Best Approximation in the $\|\cdot\|_2$ and Linear Least Squares.

1. Compare graphically the following best linear approximations in $\|\cdot\|_2$ to the function $f(x)=x^3$ on $[-1,+1]$:
   - $f_1(x)$: ***best approx.*** in the subspace $\Pi_1[-1,+1]=\text{span}\{1, x\}$, of at most 1st degree algebraic polynomials on the interval $[-1,+1]$;
   - $f_2(x)$: ***best approx.*** in the subspace $\boldsymbol{Q}_1[-1,+1]=\text{span}\{1, e^{ix}\}$, of at most 1st degree trigonometric polynomials on the interval $[-1,+1]$;
   - $f_3(x)$: ***best approx.*** in the subspace $\boldsymbol{P}_2[-1,+1]=\text{span}\{1, \cos(x), \sin(x)\}$, of at most 2nd degree trigonometric polynomials on the interval $[-1,+1]$;

   Comment on the results.
   Remember that, for a generic interval $[\alpha,\beta]$, the scalar product of f and g is defined as $\langle f,g\rangle=\int_\alpha^\beta \overline{f(x)}\,g(x)\,\mathrm{d}x$.

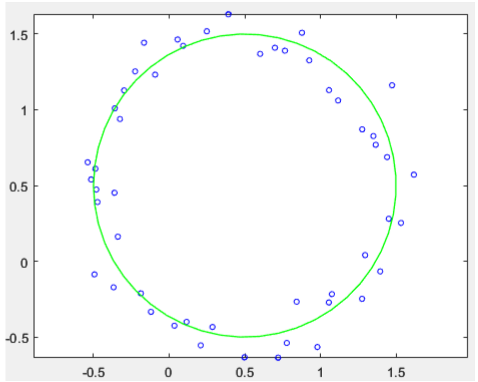
2. Find the *best linear approximations in* $\|\cdot\|_2$ to the function $f(x)=x^3$ in the subspaces $\boldsymbol{Q}_1[-\pi,+\pi]=\text{span}\{1, e^{ix}\}$, of at most 1st degree trigonometric polynomials, and $\boldsymbol{P}_2[-\pi,+\pi]=\text{span}\{1, \cos(x), \sin(x)\}$, of at most 2nd degree trigonometric polynomials, on the interval $[-\pi,+\pi]$. Compare the results with each other and with those of exercise 1.

3. By means of MATLAB *Symbolic Math Toolbox* check that, if X_{LS} represents the set of Least Squares (LS) solutions for an incompatible system, then $\min\|X_{LS}\|_2$ is reached by the LS solution returned by `pinv()`. The problem is given below.

```
u=[1 1 1 1]'; v=[1 -1 1 -1]'; A=u*v'; A=A(:,1:3); b=[1 0 1 0]';
disp([rank(A) rank([A b])])
      1     2          % A*x=b: incompatible system
disp(size(A))          % rank(A)=1 (non-max)
      4     3
```

4. Find in $\mathbb{R}^2$ a circle Γ, of equation: $x^2+y^2+\alpha x+\beta y+\gamma=0$, which is the *best fit* of a random sample of N data, obtained as in the following MATLAB box. What is the least 2-norm ($\min\|\cdot\|_2$) solution? Is it unique? If so, why?

```
% N=,100,150,200 number of samples
X0=0.5; Y0=0.5; r0=1; N=50;
t=linspace(-pi,pi,N)';
% exact Γ
Xi=X0+r0*cos(t);
Yi=Y0+r0*sin(t);
perc=0.15; % noise percentage
% perturbed Γ
xi=Xi+perc*(2*rand(N,1)-1);
yi=Yi+perc*(2*rand(N,1)-1);
plot(xi,yi,'ob'); axis equal;
hold on; plot(Xi,Yi,'r')
```



5. Solve, by means of the Linear Least Squares method, the following fitting problem (*wind tunnel experiment*):

   speed **v** (m/s) `v=[10 20 30 40 50 60 70 80]';`

   force **F** (N) `F=[25 70 380 550 610 1220 830 1450]';`

   whose (non-linear) mathematical model is the power function: $y = f(x) = a\,x^b,\ a,\ b \in \mathbb{R}$.

6. Find, by means of *Symbolic Math Toolbox*, the ***best approx.*** *w.r.t.* $\|\cdot\|_2$ of n real numbers $\{a_k\}_{k=1,\dots,n}$ (setting a value for n). What is the ***best approx.*** found?

7. Find the Least Squares line to the following data sets and comment on the obtained results:
   - A=$\{(1,1),(1,2),(1,3),(1,5)\}$ [hint: use the eq. $ax + by + c = 0$];
   - B=$\{(1,1),(-1,2),(1,3),(-1,5)\}$ [hint: use the eq. $y = mx + q$];

8. The file `salaries.csv` (downloadable from the course page on the eLearning platform) contains data concerning 398 NBA (National Basketball Association) players in the 2015-2016 season: `Name`, `Rebounds`, `Fouls`, `Points`, `Salary` (as per contract). In order to load file data and to store them in a matrix, use the following MATLAB code:

```
T=readtable('salaries.csv')
T =
  398×5 table
                 Name                 Rebounds    Fouls    Points      Salary
    _______________________________   ________    _____    ______    __________
    {'Aaron Brooks'               }     101        132       491        2.7e+06
    {'Aaron Gordon'               }     507        153       719     4.3513e+06
    {'Aaron Harrison'             }      15         10        18     8.7464e+05
    {'Adreian Payne'              }     111         77       132     2.0222e+06
    {'Al Horford'                 }     596        163      1249      2.654e+07
                  :                      :          :         :           :
M=table2array(T(:,2:end));
```

We want to find the ***best approx.*** *in* $\|\cdot\|_2$ of the data, by minimizing the following functional:

$$J_{LS} = \sum_{k=1}^{398} \left[c_1 (\text{Rebounds})_k + c_2 (\text{Fouls})_k + c_3 (\text{Points})_k - (\text{Salary})_k \right]^2$$

Compute the residue norm of the LS solutions and the elapsed time (`tic;` ... `T=toc`), and compare the results obtained by means of the following algorithms:

- MATLAB backslash operator (`\`).
- Resolution of Normal Equations.
- QR factorization.
- SVD factorization.