Corrigé 12

Exercice 1

1.a) On a

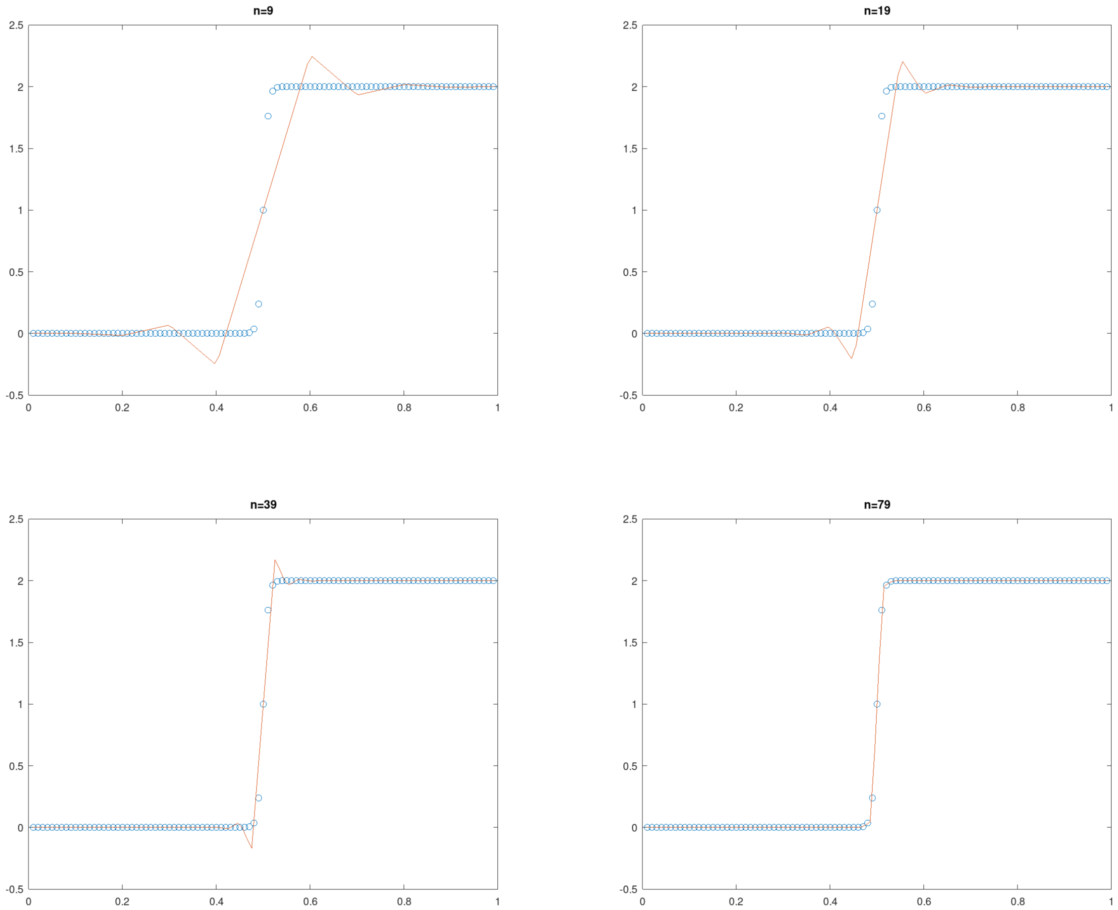
$$\frac{1}{2}||A\vec{\alpha}-\vec{b}||^2 = \frac{1}{2}\sum_{j=1}^{m}\left((A\vec{\alpha}-\vec{b})_j\right)^2 = \frac{1}{2}\sum_{j=1}^{m}\left(\sum_{i=1}^{n} A_{ji}\alpha_i - b_j\right)^2.$$

En définissant $A_{ji} = \varphi_i(x_j)$ et $b_j = y_j$ avec $j = 1, \dots, m$ et $i = 1, \dots, n$, on obtient $f(\vec{\alpha}) = \frac{1}{2}||A\vec{\alpha}-\vec{b}||^2$.
Dans le cours on a vu que $\nabla f(\vec{\alpha}) = A^T(A\vec{\alpha}-\vec{b})$ et donc $\vec{\alpha}^*$ satisfait la relation requise.

1.b) Le fichier est completé de la manière suivante :

```
function neurls(N,M)
% N : nb of neurons
% M : nb of interpolation points
% mat: the rectangular matrix corresponding to the least square problem
% val: the values at interpolation points
% weight : the weights of the neural network
  xinterp=zeros(M,1);
  val=zeros(M,1);
  for j=1:M
    xinterp(j)=j/(M+1);
    val(j)=1+tanh(100*(xinterp(j)-0.5));
  end

  mat=zeros(M,N);

  for i=1:N
    for j=1:M
      mat(j,i) = basis(N,i,xinterp(j));
    end
  end
  weight=(mat'*mat)\(mat'*val)

% plot results
  xx=zeros(100,1);
  yy=zeros(100,1);
  for k=1:100
    xx(k)=k/101;
    yy(k)=sumbasis(N,weight,xx(k));
  end
  axis([0 1 -1 1]);
  plot(xinterp,val,'o',xx,yy);
  refresh();
end
function sumbasis=sumbasis(N,weight,x)
  sumbasis=0;
  for i=1:N
    sumbasis=sumbasis+weight(i)*basis(N,i,x);
  end
end

function basis=basis(N,i,x)
  basis=max(0,x-i/(N+1));
end
```

1.c) On va comparer $f(x)$ et $\sum_{i=1}^{n} \alpha_i^* \varphi_i(x)$ à l'aide des plots suivantes



On observe que plus grand n mieux $f(x)$ est approximé par $\sum_{i=1}^{n} \alpha_i^* \varphi_i(x)$.
Remarque : Cet exercice s'inspire par un résau de n neurone à une couche avec des fonctions d'activation de type relu ($relu(x) = \max(0, x)$).

Exercice 2

2.a) En utilisant les propriétés de cosinus on a

$$\lambda_n = 2(1 - \cos(\frac{n\pi}{n+1})) = 2(1 - \cos(\pi - \frac{\pi}{n+1})) = 2(1 + \cos(\frac{\pi}{n+1})).$$

On utilise la formule de Taylor à l'ordre 2 autour de zéro :

$$\cos(x) = 1 - \frac{x^2}{2!} + o(x^2)$$

et on obtient

$$\cos(\frac{\pi}{n+1}) = 1 - \frac{(\frac{\pi}{n+1})^2}{2!} + o((\frac{\pi}{n+1})^2) = 1 - \mathcal{O}(\frac{1}{n^2}).$$

On a donc que

$$\lambda_n = 2(2 - \mathcal{O}(\frac{1}{n^2})) = 4 - \mathcal{O}(\frac{1}{n^2}).$$

D'après le théorème de cours on a convergence si

$$\alpha < \frac{2}{\lambda_n} = \frac{2}{4 - \mathcal{O}(\frac{1}{n^2})} = \frac{1}{2} + \mathcal{O}(\frac{1}{n^2}).$$

Donc la méthode converge si $\alpha \leqslant \frac{1}{2}$.

2.b) La méthode du gradient à pas optimal est implementée de la manière suivante :

```
function [i,x]=gradopt(N)
I=speye(N,N);
diag = 2*I;
subd=-sparse(2:N,1:N-1,1,N,N);
A=diag+subd+subd';
b=ones(N,1);
x=zeros(N,1);
for i=1:100000
  p=b-A*x;
  alpha= dot(p,p)/dot(p',A*p);
  x=x+alpha*p;
  if (norm(p)<1.e-6)
    break
  end
end
```

2.c) On complète le tableau à l'aide des deux programmes. On observe que pour les deux méthodes si on multiplie n par 2, le nombre d'itérations est multiplié par 4 et donc le nombre d'itérations est bien $\mathcal{O}(n^2)$. On observe aussi que la méthode du gradient à pas constant ne converge pas avec $\alpha = 0.6$ et que avec α optimal le nombre d'itérations est plus petite que avec les autres valeurs de α.

n	$\alpha = 0.1$	$\alpha = 0.49$	$\alpha = 0.6$	α *optimal*
9	1511	304	-	293
19	6169	1254	-	1235
39	25223	5142	-	5109
79	103120	21040	-	20983