Yannis Bähni
yannis.baehni@uzh.ch

SOLUTIONS SHEET 6

Disclaimer: I try to use now recisely the theory and notation provided in the lecture. But still one word to my last code `myspline.m`: Yes, it is really my code. I used the book *Stoer/Bulirsch: Numerische Mathematik I* which I appreciate a lot. Since I used theory from there, I cited it in my code. But I try to avoid this in the next exercise sheets.

Exercise 1. Define the *residual function* $\varrho : \mathbb{R}^n \to \mathbb{R}$ with parameters $P \in M_{mn}(\mathbb{R})$ and $y \in \mathbb{R}^m$ as $\varrho(a; y, P) := \|y - Pa\|_2^2$. "$\Rightarrow$": Assume $a_0 \in \mathbb{R}^n$ minimizes $\varrho(a; y, P)$. This implies $\nabla\varrho(a_0; y, P) = 0$ and thus $P^tPa_0 = P^ty$ (since the normal system *is precisely* the gradient of $\varrho(a; y, P)$). Hence a_0 is a solution of the normal system. "$\Leftarrow$": Let $a_0 \in \mathbb{R}^n$ be a solution of the normal system and $a \in \mathbb{R}^n$ arbitrary. Then

$$
\begin{aligned}
\|y - Pa\|_2^2 =& \|y - Pa + Pa_0 - Pa_0\|_2^2 \\
=& \|(y - Pa_0) - (Pa - Pa_0)\|_2^2 \\
=& \left((y - Pa_0) - (Pa - Pa_0)\right)^t \left((y - Pa_0) - (Pa - Pa_0)\right) \\
=& \left((y - Pa_0)^t - (Pa - Pa_0)^t\right) \left((y - Pa_0) - (Pa - Pa_0)\right) \\
=& \|y - Pa_0\|_2^2 + \|Pa - Pa_0\|_2^2 - (y^t - a_0^tP^t)(Pa - Pa_0) - (a^tP^t - a_0^tP^t)(y - Pa_0) \\
=& \|y - Pa_0\|_2^2 + \|Pa - Pa_0\|_2^2 - \left[y^tPa - y^tPa_0 - a_0^tP^tPa + a_0^tP^tPa_0\right] \\
& - \left[a^tP^ty - a^tP^tPa_0 - a_0^tP^ty + a_0^tP^tPa_0\right] \\
=& \|y - Pa_0\|_2^2 + \|Pa - Pa_0\|_2^2 - \left[a^tP^ty - a_0^tP^ty - a^tP^tPa_0 + a_0^tP^tPa_0\right] \\
& - \left[a^tP^ty - a^tP^tPa_0 - a_0^tP^ty + a_0^tP^tPa_0\right] \\
=& \|y - Pa_0\|_2^2 + \|Pa - Pa_0\|_2^2 - \left[a^t(P^ty - P^tPa_0) + a_0^t(P^tPa_0 - P^ty)\right] \\
& - \left[a^t(P^ty - P^tPa_0) + a_0^t(P^tPa_0 - P^ty)\right] \\
=& \|y - Pa_0\|_2^2 + \|Pa - Pa_0\|_2^2 \\
\geqslant& \|y - Pa_0\|_2^2
\end{aligned}
$$

Thus $\varrho(a_0; y, P) \leqslant \varrho(a; y, P)$ for any $a \in \mathbb{R}^n$. By definition a_0 minimizes ϱ. $\square$

Exercise 2. **a.** The coefficients are given by (approximately)

$$
\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} 4.0591 \\ 0.6140 \\ -2.5315 \\ 0.7058 \end{bmatrix} \tag{1}
$$

according to the solution of the normal system $P^tPa = P^ty$ with the built-in `linsolve`. The code can be found in listing 1.

b. The absolute error of the least square method can be seen below. In comparison with the range of the data values (from 2 to 100) the linear fitting is quite accurate (all around 1). Additionally a plot of the linear fit and the data can be found in figure 1.

$$
\begin{bmatrix} 0.0379 & 0.4918 & 1.1173 & 0.2839 & 0.6808 & 0.2651 & 0.0793 & 0.0057 & 0.6675 & 0.8475 \end{bmatrix}
$$

Yannis Bähni
yannis.baehni@uzh.ch

```
xi = .1:.1:1.;
fi = [1e+2,34,17,12,9,6,5,4,4,2];
phi1 = @(x) 1./x;
phi2 = @(x) 1./x.^2;
phi3 = @(x) exp(-(x-1));
phi4 = @(x) exp(-2 * (x-1));

M = [phi1(xi)',phi2(xi)',phi3(xi)',phi4(xi)'];
%(a)
sol = linsolve(M' * M, M' * fi');
disp('coefficients ak:');
disp(sol);

ftilde = @(x) sol(1) * phi1(x) + sol(2) * phi2(x) + sol(3) * phi3(x) + ...
    sol(4) * phi4(x);

%(b)
plot(xi,fi,'x','color','blue');
hold on;
x = linspace(.1,1.,250);
figure(1);
plot(x,ftilde(x),'color','red');
disp('Absolute error of LSTSQ:')
disp(abs(ftilde(xi) - fi));
ylim([2,100]);
xlabel('$x$', 'interpreter', 'latex', 'fontsize', 18);
ylabel('$f(x)$', 'interpreter', 'latex', 'fontsize', 18);
grid on;
l = legend('$(x_i,f_i)$','$\tilde{f}(x)$');
set(l, 'fontsize', 16, 'interpreter', 'latex','location','north');
fig = figure(1);
fig.PaperUnits = 'inches';
fig.PaperPosition = [0, 0, 12, 6];
saveas(fig, 'ex_2_b.jpg');
```

LISTING 1. `src/ex_2.m`.

Exercise 3.

a. The source code can be found in listing 2. Remark: Since the function is 1-periodic and the function `mydft` uses this periodicity one could have omnmitted the term $e^{-2\pi}$ in c_k since it vanishes with $e^{2\pi}$ in $\tilde{f}(x)$.

b. The plot can be found in figure 2. The data is nicely interpolated for $l = 6, 8, 10, 12$. Remark: For $l = 2$ I observe that the trigonometric polynomial does not interpolate the data. I already asked Celine, she did the same observation (to be clarified).

c. According to the lecture we expect the theoretical complexity $O(N^2)$. However as one can see in figure 3 this complexity is an overestimation. We observe a much lower complexity. This is probably due to the fact that we use special $N = 2^l$ or that the complexity is true for $N \to +\infty$.

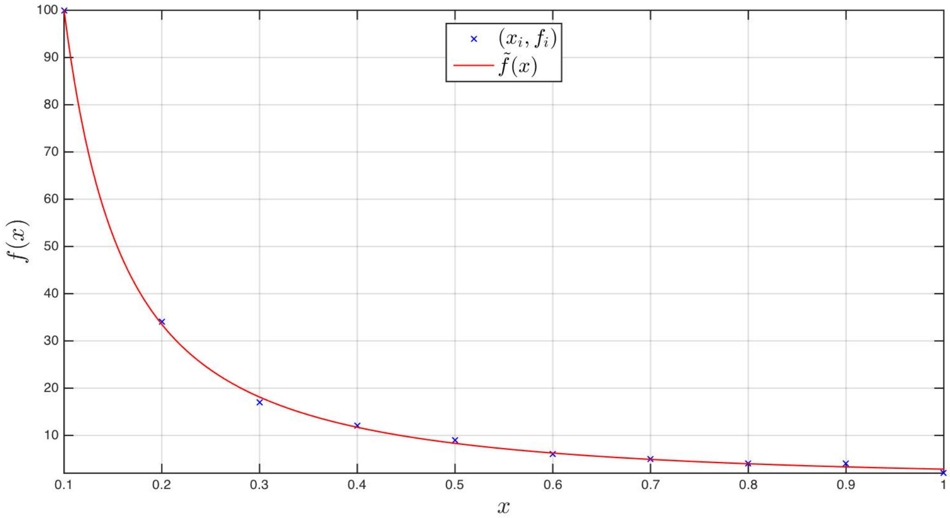

FIGURE 1. Plot of the linear fit and the data points.

```
function [ fval ] = mydft( xi,fi,x )
n = length(xi);
if mod(n - 1,2) == 0
    M = (n - 1)/2;
    mu = 0;
else
    M = (n - 2)/2;
    mu = 1;
end

index = -(M + mu):1:(M + mu);
c = zeros(1,length(index));
for k = 1:length(index)
    for j = 1:n
        c(k) = c(k) + fi(j) * exp(-2 * pi * i * index(k) * xi(j));
    end
end
fval = zeros(1,length(x));
for j = 1:length(x)
    for k = 1:length(index)
        fval(j) = fval(j) + 1/n * c(k) * exp(2 * pi * i * index(k) * x(j));
    end
end
```

LISTING 2. `src/mydft.m`

d. The plot can be found in figure 4. If we define $\varepsilon_k := \max_{x\in[0,1]} |f_{2^k} - f|$ and consider the plot which is a double logarithmic plot, we may conclude that the convergence

Yannis Bähni
yannis.baehni@uzh.ch

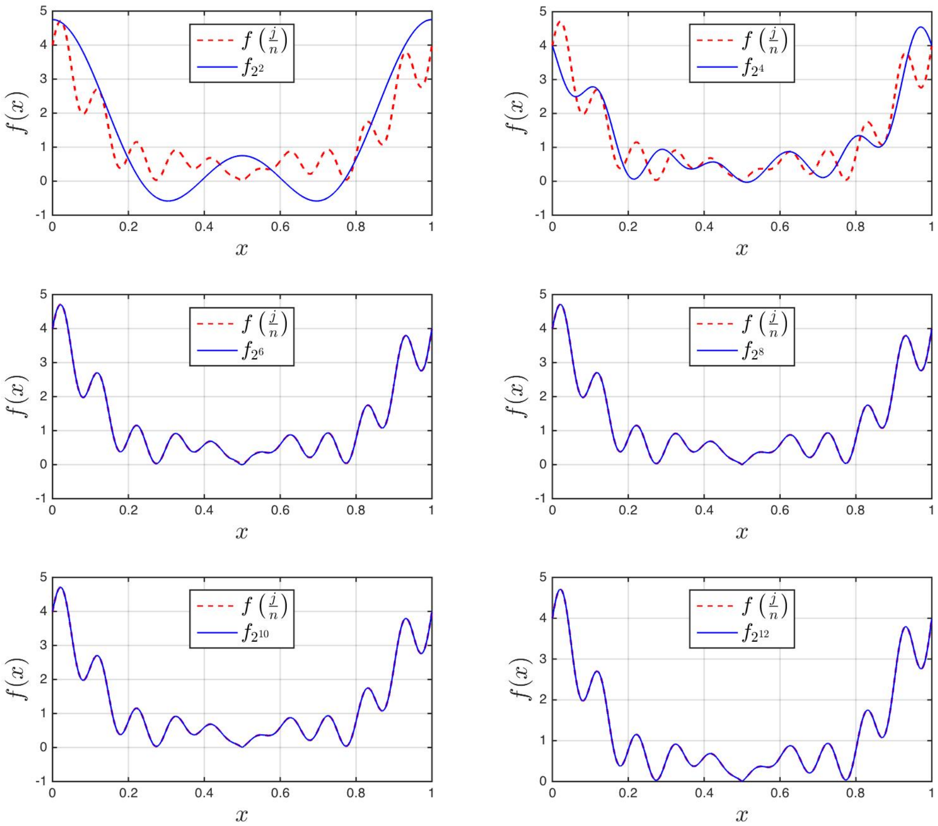

FIGURE 2. Plot of the interpolated data for several numbers of nodes.

k	$\left\lvert\log\left(\frac{\varepsilon_{k+1}}{\varepsilon_k}\right) / \log\left(\frac{h_{k+1}}{h_k}\right)\right\rvert$
2	0.1552
3	0.0748
4	4.7737
5	1.3060
6	1.0836
7	1.0417
8	1.0178
9	1.0679
10	1.0042
11	28.8160

TABLE 1. Convergence rate of the DFT method for $h := (n/2^l)_{l=2}^{12}$.

rate is of the type Ch_k^b for $b \in \mathbb{N}$ and $h_k := n/2^k$. As one can tell from table 1 the convergence rate of the DFT method is approximately h. Thus linear convergence.

Exercise 4. **a.** The plot can be found in figure 5.

Yannis Bähni
yannis.baehni@uzh.ch

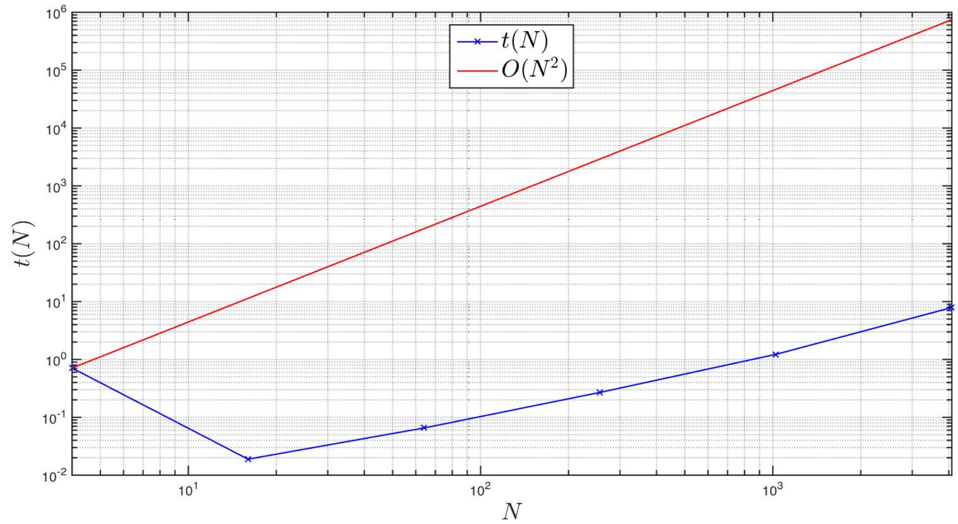

FIGURE 3. Plot of the (time) complexity of the trigonometric interpolation.

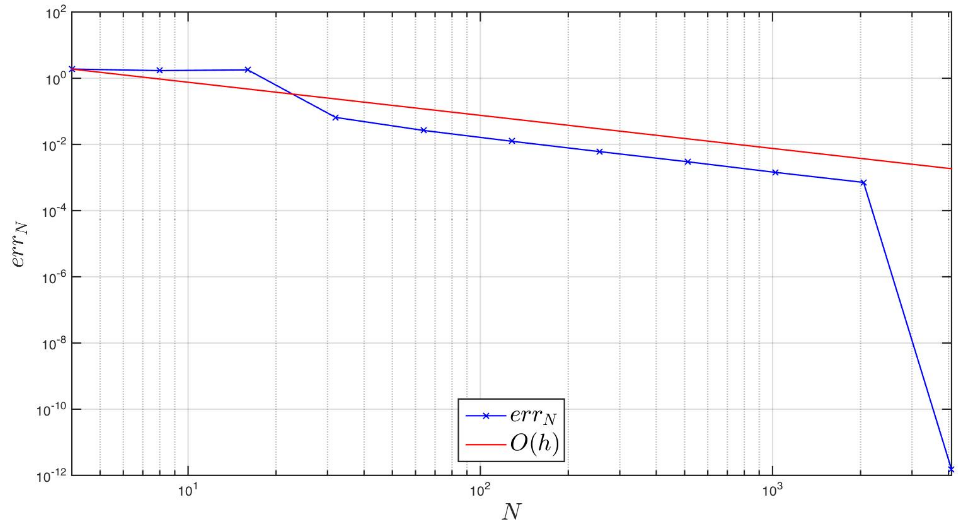

FIGURE 4. Plot of the error of the trigonometric interpolation for several N.

b. The code can be found in listing 3.
c. The plot can be found in figure 6.

APPENDIX A. DERIVATION OF THE NORMAL EQUATIONS

I will give a more detailed outline of the derivation of the *normal system* (as an exercise for me). Let $P \in M_{mn}(\mathbb{R})$ and $y \in \mathbb{R}^m$. We are looking for the *optimal solution* $a \in \mathbb{R}^n$ for $m > n$, this means that a minimizes the *residuum* $\|y - Pa\|_2^2$ in euclidean standard norm. The residuum can be interpreted as a function $\varrho : \mathbb{R}^n \to \mathbb{R}$ with parameters $y \in \mathbb{R}^m$ and $P \in M_{mn}(\mathbb{R})$. We thus

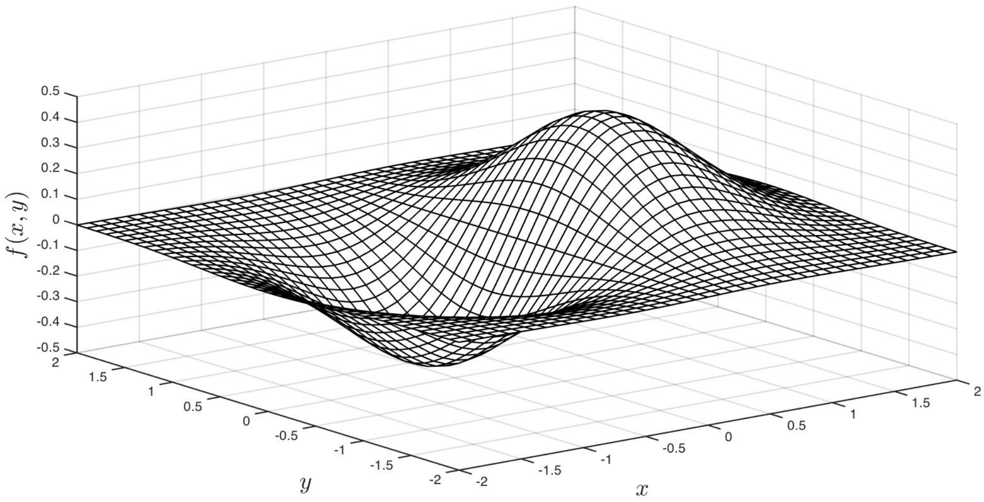

FIGURE 5. Plot of the function $f(x) = xe^{-x^2-y^2}$ on $[-2,2] \times [-2,2]$ using a grid with 1681 points.

```
function [ Z ] = mylagrange2d( X,Y,Xi,Yi,Zij )
Z = zeros(length(Y),length(X));
for k = 1:length(X)
    for l = 1:length(Y)
        for i = 1:length(Xi)
            for j = 1:length(Yi)
                Z(k,l) = Z(k,l) + Zij(i,j) * lagrange_basis(X(k),Xi,i) ...
                    * lagrange_basis(Y(l),Yi,j);
            end
        end
    end
end
end

function [ B ] = lagrange_basis( x,xi,j )
    B = 1;
    for k = 1:length(xi)
       if (k ~= j)
          B = B * (x - xi(k))/(xi(j) - xi(k));
       end
    end
end
```

LISTING 3. src/mylagrange2d.m

write $\varrho(a; y, P)$ instead of $\varrho(a)$. Since the function ϱ is affine linear we may consider the gradient $\nabla\varrho(a; y, P)$ at $a \in \mathbb{R}^n$ of ϱ

Yannis Bähni
yannis.baehni@uzh.ch

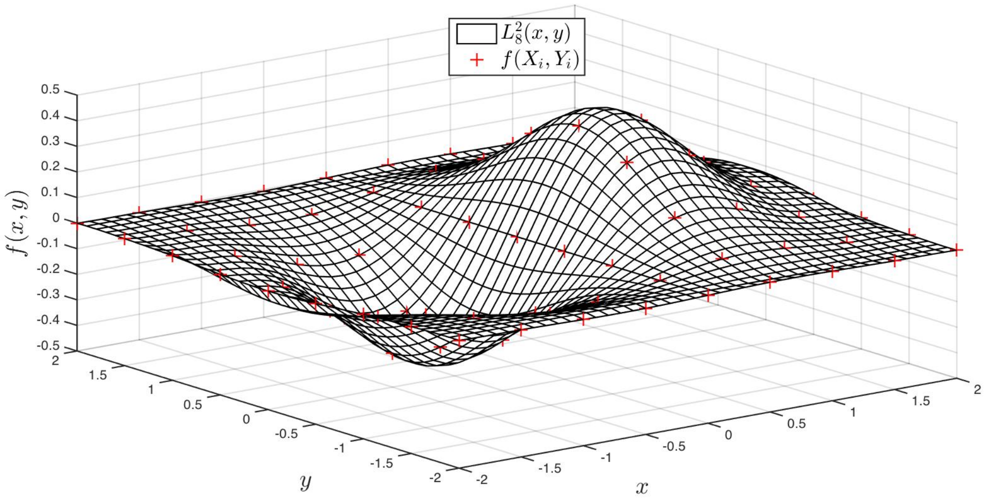

FIGURE 6. Plot of the Lagrange interpolant of the function $f(x) = xe^{-x^2-y^2}$ on $[-2,2]\times[-2,2]$ using an interpolation grid with 81 points.

$$\begin{aligned}
\nabla\varrho(a;y,P) &= \begin{pmatrix} \frac{\partial\varrho}{\partial a_1}(a;y,P) & \ldots & \frac{\partial\varrho}{\partial a_n}(a;y,P)^t \end{pmatrix} \\
&= 2\begin{pmatrix} \sum_{i=1}^m \left(\sum_{j=1}^n p_{ij}a_j\right) p_{i1} \\ \vdots \\ \sum_{i=1}^m \left(\sum_{j=1}^n p_{ij}a_j\right) p_{in} \end{pmatrix} - 2\begin{pmatrix} \sum_{i=1}^m p_{i1}y_i \\ \vdots \\ \sum_{i=1}^m p_{in}y_i \end{pmatrix} \\
&= 2P^tPa - 2P^ta
\end{aligned}$$

By $\varrho(a;y,P) = \|y-Pa\|_2^2 = (y-Pa)^t(y-Pa) = (y^t-a^tP^t)(y-Pa) = \|y\|_2^2 - y^tPa - a^tP^ty + \|Pa\|_2^2$ and

$$\begin{aligned}
\frac{\partial\varrho}{\partial a_k}(a;y,P) &= \frac{\partial}{\partial a_k}\left(\|y\|_2^2 - y^tPa - a^tP^ty + \|Pa\|_2^2\right) \\
&= \frac{\partial}{\partial a_k}\|Pa\|_2^2 - \frac{\partial}{\partial a_k}y^tPa - \frac{\partial}{\partial a_k}a^tP^ty \\
&= \frac{\partial}{\partial a_k}\left(\sum_{i=1}^m\left(\sum_{j=1}^n a_{ij}a_j\right)^2\right) - \frac{\partial}{\partial a_k}\left(\sum_{i=1}^m\left(\sum_{j=1}^n a_{ij}a_j\right)y_i\right) - \frac{\partial}{\partial a_k}\left(\sum_{j=1}^n\left(\sum_{i=1}^m a_{ij}y_i\right)a_j\right) \\
&= 2\sum_{i=1}^m\left(\sum_{j=1}^n a_{ij}a_j\right)a_{ik} - 2\sum_{i=1}^m a_{ik}y_i
\end{aligned}$$

for $k = 1,\ldots,n$. Now by Analysis in several variables we have for $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, U open, and differentiable in $a \in U$ the implication if $\nabla f(a) \neq 0$ then a is no local extrema. Logically equivalent is the implication (contraposition) if $a \in U$ is a local extrema of f and f is differentiable at a then $\nabla f(a) = 0$. Hence the condition $\nabla f(a) = 0$ is *necessary* for a being a local extrema of the function f. Hence if $a \in \mathbb{R}^n$ should minimize $\varrho(a; y, P)$ it must hold that $\nabla \varrho(a; y, P) = 0$ which yields the *normal system* $P^t P a = P^t y$ for a.