Week 7 Tutorial 7

(7.1) *Least square solution of a linear system:* We wish to solve a system of linear equations

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ with rk $\boldsymbol{A} = m$, $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{b} \in \mathbb{R}^m$. In many applications

- it is often not possible to calculate the inverse of $\boldsymbol{A}$,
- or we have more equations than unknowns, that is $m > n$, and therefore can not expect a unique solution.

In the later case one can calculate the least square solution, that is the vector $\boldsymbol{x}$ that minimises the squared Euclidean: distance

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2$$

(we multiply with $\frac{1}{2}$ to make the gradient nicer). Show that

1. the gradient of $f = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2$ is given by

$$\nabla_{\boldsymbol{x}} f = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$$

2. Calculate the critical point

$$\boldsymbol{x}^* = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}.$$

   Is this point unique? Explain why $\boldsymbol{A}^T\boldsymbol{A}$ is invertible.

Solution: We can write the squared Euclidean distance as

$$f(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = \frac{1}{2}\left(\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{b} + \boldsymbol{b}^T\boldsymbol{b}\right)$$

We note that $\boldsymbol{A}^T\boldsymbol{A}$ is always symmetric, therefore we can use the example 2.5 in the lecture notes and conclude that

$$\nabla f = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$$

The critical point can be found by setting $\nabla f = \boldsymbol{0}$, giving the stated solution.
The solution is unique since the problem is quadratic (we see that the Hessian is the matrix $\boldsymbol{A}^T\boldsymbol{A}$ which is always positive definite. Furthermore $\boldsymbol{A}^T\boldsymbol{A}$ is a square matrix in $\mathbb{R}^{m\times m}$ which has rank m. This follows that if $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = 0$ then

$$0 = \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{A}\boldsymbol{x}) = \langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}\rangle = \|\boldsymbol{A}\boldsymbol{x}\|^2.$$

Which implies that $\boldsymbol{A}\boldsymbol{x} = 0$ and since $\boldsymbol{A}$ has full rank, we conclude that $\boldsymbol{x} = \boldsymbol{0}$.

(7.2) *Norms:* Let $f : [0,1] \to \mathbb{R}$ be continuous. Show that $\|f\|_\infty = \sup_x |f(x)|$ is indeed a norm.

Solution:

1. If $\|f\|_\infty = 0$, then $\sup|f(x)| = 0$ which implies that $f \equiv 0$.
2. $\|\alpha f\|_\infty = \sup|\alpha f(x)| = |\alpha| \sup|f| = |\alpha| \|f\|_\infty$.
3. $\|f+g\|_\infty = \sup_x |f(x)+g(x)| \leq \sup_x |f(x)| + |g(x)| = \|f\|_\infty + \|g\|_\infty$.

(7.3) *Interpolation I:* Given $n+1$ data points (x_i, y_i) with $x_i, y_i \in \mathbb{R}$ we wish to find a polynomial, which fits these points exactly.
One can for example consider a polynomial of degree n (then we have the same number of equations and unknowns):

$$p_n(x) = a_0 + a_1 x + a_2 x^2 + \ldots a_n x^n.$$

Its coefficients are determined by setting $y_i = p(x_i)$ for $i = 1, \ldots n+1$.

1. Show that the coefficients $\mathbf{a} = (a_0, a_1, \ldots a_n)$ can be found y solving the system $\boldsymbol{V}\boldsymbol{a} = \boldsymbol{y}$, where $\boldsymbol{V}$ is the so called Vandermonde matrix

$$\boldsymbol{V} = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 & \ldots & x_1^n \\ 1 & x_1 & x_2^2 & x_2^3 & \ldots & x_2^n \\ \vdots & \vdots & \vdots & \vdots & & \\ 1 & x_n & x_n^2 & x_n^3 & \ldots & x_n^n \end{pmatrix}$$

   and $\boldsymbol{y} = (y_1, \ldots y_n)$.

2. The determinant of $\boldsymbol{V}$ is given by det $\boldsymbol{V} = \Pi_{0 \leq i,j \leq n}(x_j - x_i)$. Why can this become problematic?

Solution: Since we require that $p_n(x_i) = y_i$ we obtain n equation of the form

$$a_0 + a_1 x_i + a_2 x_i^2 + \ldots a_n x_i^2 = y_i$$

for $i = 1, \ldots n$. These equations can be written exactly in the above form.
The condition of the Vandermonde matrix may deteriorate for a large number of points. In particular if points get very close, the determinant can quite small. Furthermore finding the interpolation polynomial requires that the solution of a linear system, which is also computationally costly.

(7.4) *Interpolation II* Lagrangian interpolation is an attractive alternative to the Vandermonde interpolation, because it does not involve solving a system to find the interpolating polynomial. It is based on writing the polynomial in a different way, for example instead of writing $y = x^2 - 3x + 2$ we can also write $y = (x-2)(x-3)$.
Given a data set $(x_i, y_i) \in \mathbb{R}^2$, with $0 \leq i \leq n$. Then the Lagrange basis polynomials are given by

$$\ell_i(x) = \frac{\Pi_{k \neq i}(x - x_k)}{\Pi_{k \neq i}(x_i - x_k)} \quad i = 0, \ldots n$$

- Show that

$$\ell_i(x_j) = \begin{cases} 1 & j = i \\ 0 & j \neq i, \end{cases}$$

  and that the Lagrange interpolating polynomial through those data points

$$p_n(x) = \sum_{k=0}^{n} y_k \ell_k(x)$$

  satisfies $p_n(x_i) = y_i$ for every $i = 0, \dots n$.

- Consider the function $f(x) = \frac{1}{x}$. Given the function values at $x_0 = 2$, $x_1 = 2.5$ and $x_2 = 4$, that is given the data pairs $(2, \frac{1}{2})$, $(\frac{5}{2}, \frac{2}{5})$ and $(4, \frac{1}{4})$, calculate the corresponding interpolating Lagrange basis polynomials $\{\ell_0, \ell_1, \ell_2\}$ and the interpolating Lagrange polynomial

$$p_2(x) = \sum_{k=0}^{2} y_k \ell_k(x).$$

  Sketch the solution and the function f.

Solution: For the basis polynomials we see that the numerator and the denominator are the same for $i = j$ and that for $i \neq j$ the terms in the numerator are 0, showing the first statement. This implies the second statement (since ℓ either takes the value 0 or 1). The Lagrange basis polynomials are

$$\begin{aligned}
\ell_0(x) &= \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} = (x - 6.5)x + 10 \\
\ell_1(x) &= \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} = \frac{1}{3}((-4x + 24)x - 32) \\
\ell_2(x) &= \frac{1}{3}((x - 4.5)x + 5)
\end{aligned}$$

The respective Lagrange interpolating polynomial is then

$$p_2(x) = \sum_{k=0}^{2} y_k \ell_k(x) = (0.05x - 0.425)x + 1.15.$$

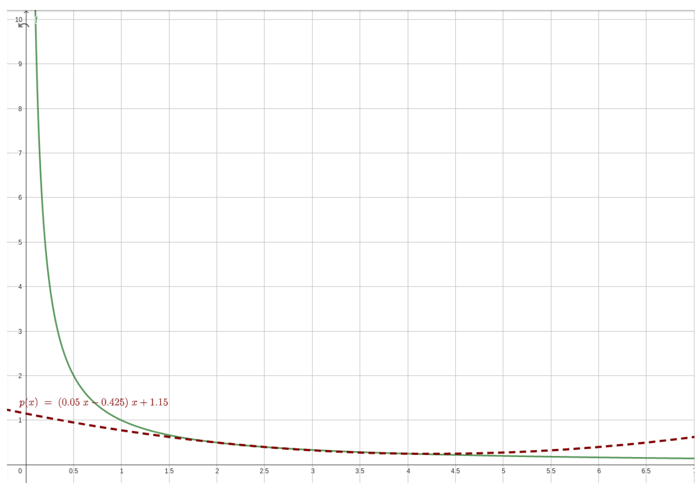

Figure 1: Lagrange interpolating polynomial vs. function f