Prof. Ch. Schwab
B. Fitzpatrick
J. Zech

Numerische Mathematik I

ETH Zürich
D-MATH

Homework Problem Sheet 12

For some problems, parts of the solution are already given. Fill in the gaps and complete the proofs where you see a red band at the left margin.

Introduction. Householder, least squares, pseudoinverse

Problem 12.1 Householder Transformation and proof of [NMI, Theorem 5.20]

Let $0 \neq \mathbf{v} \in \mathbb{R}^n$. Show that the Householder Transformation

$$\mathbf{Q} := \mathbf{I}_n - \frac{2}{\mathbf{v}^\top \mathbf{v}} \mathbf{v}\mathbf{v}^\top,$$

where $\mathbf{I}_n$ denotes the identity matrix, has the following properties a) - d):

(12.1a) $\mathbf{Q}$ is symmetric, i. e. $\mathbf{Q}^\top = \mathbf{Q}$.

Solution:

$$\mathbf{Q}^\top =$$

(12.1b) $\mathbf{Q}$ is orthogonal, i. e. $\mathbf{Q}^\top = \mathbf{Q}^{-1}$.

Solution:

$$\mathbf{Q}^\top \mathbf{Q} =$$

(12.1c) $\mathbf{Q}$ is involutary, i. e. $\mathbf{Q}^2 = \mathbf{I}_n$.

Solution:

(12.1d) Let $0 \neq \mathbf{a} \in \mathbb{R}^n$ and $\vec{e_1}$ the first unit vector. Let $\alpha = \|\mathbf{a}\|_2$ or, in the case $\mathbf{a} = c\vec{e_1}$, $c \in \mathbb{R} \setminus \{0\}$, let $\alpha = -\|\mathbf{a}\|_2$. We then have

$$\mathbf{Q}\mathbf{a} = \alpha\vec{e_1},$$

for $\mathbf{v} = \mathbf{a} - \alpha\vec{e_1}$.

Solution: We want to show: For $\mathbf{v} = \mathbf{a} - \alpha\vec{e_1}$, $0 \neq \mathbf{a} \in \mathbb{R}^n$ and $\alpha = \pm\|\mathbf{a}\|_2$ we have:

$$\mathbf{Q} = \mathbf{I}_n - \frac{2}{\mathbf{v}^\top\mathbf{v}}\mathbf{v}\mathbf{v}^\top \Rightarrow \mathbf{Q}\mathbf{a} = \alpha\vec{e_1}.$$

We simply plug in:

(12.1e) Show the following identity occuring in the proof of [NMI, Theorem 5.20]:

$$F(\vec{x}^{(0)}) - F(\vec{x}^\star) - F'(\vec{x}^{(0)})(\vec{x}^{(0)} - \vec{x}^\star) = \int_0^1 \big(F'(\vec{x}^{(0)} + t(\vec{x}^\star - \vec{x}^{(0)})) - F'(\vec{x}^{(0)})\big)(\vec{x}^{(0)} - \vec{x}^\star)\,\mathrm{d}t\;.$$

Solution: We have $(F_j)_{j=1}^n = F : \mathbb{R}^n \supseteq D \to \mathbb{R}^n$, and hence $F'(\vec{x}) = (\partial F_j/\partial x_i)_{j,i=1}^n \in \mathbb{R}^{n\times n}$ for every $\vec{x} \in D$. For $j \in \{1,\dots,n\}$ denote by $G_j : [0,1] \to \mathbb{R}$ the function $G_j(t) := F_j(\vec{x}^{(0)} + t(\vec{x}^\star - \vec{x}^{(0)}))$. Then,

Problem 12.2 MATLAB: Linear Curve Fitting

We are given the points $(x_i, y_i)^\top, i = 1, \ldots, 8$, in the plane by

x	0.7	3.3	5.6	7.5	6.4	4.4	0.3	-1.1
y	4.0	4.7	4.0	1.3	-1.1	-3.0	-2.5	1.3

Calculate the radius r and the centre $M = (m_1, m_2)^\top$ of a circle in such a way that the circle pictures the points best, i. e. such that the sum of the squares of the distance in between the points and the circle becomes minimal.
The equation of a circle reads as follows

$$(x - m_1)^2 + (y - m_2)^2 = r^2.$$

(12.2a) Introduce a new unknown $c = r^2 - m_1^2 - m_2^2$ and restate the equation of the circle as a linear equation in three unknowns $\mathbf{z} = (m_1, m_2, c)^\top$.

Solution:

(12.2b) Rewrite the equation of the circle as an over-determined system of linear equations $\mathbf{Az} = \mathbf{b}$ (where $\mathbf{A} \in \mathbb{R}^{8\times3}$).

Solution: The system $\mathbf{Az} = \mathbf{b}$ is of the following form:

(12.2c) Use the QR decomposition in order to solve the linear curve fitting problem $\min_{\mathbf{z}} \|\mathbf{Az} - \mathbf{b}\|_2$ (in order to do this see [NMI, Code. 6.20], the `for` loop can be replaced by the MATLAB-function `qr`). State r and M of the circle you found.

Solution:

(12.2d) Display the points and the calculated circle in the same plot.

Solution:

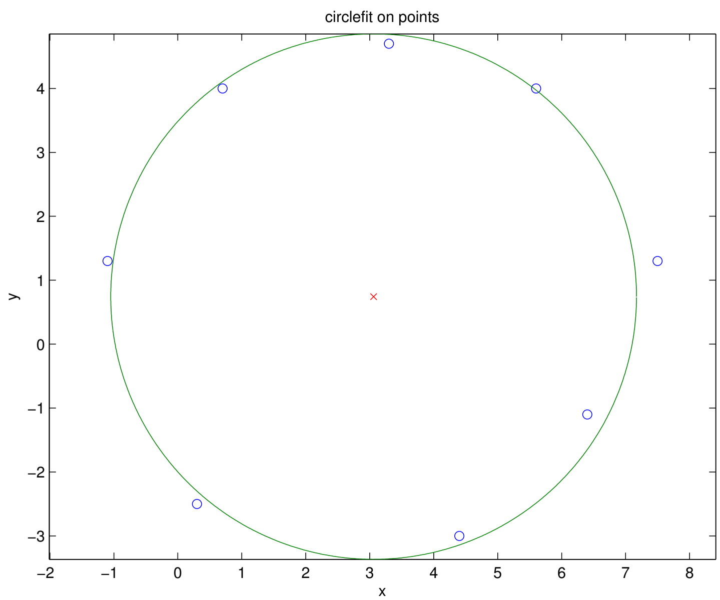

Figure 12.1: Circlefit for given points

Problem 12.3 The Pseudo inverse

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ with rank r. Show that the pseudo inverse $\mathbf{A}^\dagger$ of $\mathbf{A}$ has the following properties:

(12.3a) $\mathbf{A}^\dagger \mathbf{A} = \mathbf{I}_n$ if $r = n \leq m$.

Solution: By the definition of the pseudo inverse from [NMI, Sec. 6.6] we have that $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ and $\mathbf{A}^\dagger = \mathbf{V}\boldsymbol{\Sigma}^\dagger\mathbf{U}^\top$. Therefore

(12.3b) $\mathbf{A}^\dagger = \mathbf{A}^{-1}$ if $r = n = m$.

Solution:

(12.3c) $\mathbf{A}^\dagger = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top$ if $r = n$.

Solution: Note that $\mathbf{A}^\top\mathbf{A}$ must have the same rank as the matrix $\mathbf{A}$, which is n. Therefore $\mathbf{A}^\top\mathbf{A}$ is invertible and we can write $\mathbf{A}^\dagger = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top \Leftrightarrow \mathbf{A}^\top\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\top$. We have

$$\mathbf{A}^\top\mathbf{A}\mathbf{A}^\dagger =$$

(12.3d) $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$ and $\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger$.

Solution: Using the solution for subproblem (12.3a) we have

Problem 12.4 Linear Regression in 1D

(12.4a) Let sampling points $-\infty < x_1 < \ldots < x_n < \infty$ and weights $\{y_k\}_{k=1}^n$ be given. Calculate $a, b \in \mathbb{R}$, in such a way that

$$E(a,b) = \sum_{k=1}^{n} ((a + x_k b) - y_k)^2$$

is minimized.
HINT: $E(a,b)$ is of the form $E(\mathbf{z}) = \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2^2$ with $\mathbf{z} = (a,b)^\top$.

Problem 12.5 Least squares

We consider least squares problems of the form

i) Find the minimizer $\mathbf{x} \in \mathbb{R}^n$ of $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$ with $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $m > n$ such that $\operatorname{rank}(\mathbf{A}) = n$.

ii) Find the minimizer $\mathbf{x} \in \mathbb{R}^l$ of $\|\mathbf{B}\mathbf{x} - \mathbf{y}\|_2$ with minimal 2-norm, where $\mathbf{y} \in \mathbb{R}^k$, $\mathbf{B} \in \mathbb{R}^{k \times l}$ and $k, l \in \mathbb{N}$.

(12.5a) Show that the residual $\mathbf{r}$ of the solution $\mathbf{x}$ of the least squares problem i), is orthogonal to the columns of $\mathbf{A}$. Use this to deduce the normal equation.

(12.5b) Let $\mathbf{Q}$, $\mathbf{R}$ be the QR decomposition of the matrix $\mathbf{A}$, with $\mathbf{A} = \mathbf{Q}\mathbf{R}$, where

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_0 \\ \mathbf{0} \end{pmatrix}$$

with $\mathbf{R}_0 \in \mathbb{R}^{n \times n}$. Let the vector $\widetilde{\mathbf{b}} = \mathbf{Q}^\top \mathbf{b}$ be decomposed into $\widetilde{\mathbf{b}} = (\widetilde{\mathbf{b}}_1^\top, \widetilde{\mathbf{b}}_2^\top)^\top$, $\widetilde{\mathbf{b}}_1 \in \mathbb{R}^n$, $\widetilde{\mathbf{b}}_2 \in \mathbb{R}^{m-n}$. Show that the solution $\mathbf{x}$ of the least squares problem i) is equal to the solution of the equation system $\mathbf{R}_0 \mathbf{x} = \widetilde{\mathbf{b}}_1$.

(12.5c) We want to figure out the coefficients a, c of the function $f(t) = at(1 + c\mathrm{e}^{-t})$, in such a way that f approximates the data (t_i, f_i) as well as possible in the sense of least squares. The data is given by

t_i	0.2	0.4	0.6	0.8	1.0
f_i	2.3	3.0	2.9	2.0	1.1

In order to do this, reformulate the problem such that it is of the form i), by setting up $\mathbf{A}$, $\mathbf{b}$, $\mathbf{x}$ schematically without calculating the explicit values.

(12.5d) Define the pseudoinverse $\mathbf{B}^+$ of $\mathbf{B}$. Prove that $\mathbf{x} := \mathbf{B}^+\mathbf{y}$ solves the least squares problem ii).

(12.5e) Denote by $\mathbf{I} \in \mathbb{R}^{k\times k}$ the identity matrix. Show that

$$\mathbf{B}^+ = \lim_{\substack{\varepsilon>0\\ \varepsilon\to 0}} \mathbf{B}^\top(\mathbf{B}\mathbf{B}^\top + \varepsilon\mathbf{I})^{-1}.$$

Why does the inverse exist for $\varepsilon > 0$?

Published on May 18, 2016.

To be submitted on May 31, 2016.

MATLAB: Submit all files in the online system. Include the files that generate the plots. Label all your plots. Include commands to run your functions. Comment on your results.

References

[NMI] Lecture Notes for the course "Numerische Mathematik I".

Last modified on May 17, 2016