1. Laat zien dat $A^\dagger = (A^T A)^{-1} A^T$ als $A \in \mathrm{R}^{m\times n}$ met $m \geq n$ en $\mathrm{rang}(A) = n$. $A^\dagger$ is de pseudoinverse van A. Eerst ben ik natuurlijk eens moeten gaan opzoeken wat een pseudoinverse is. In Golub & Van Loan kwamen ze af met volgende definitie: $A^\dagger$ is de matrix zodat

$$AA^\dagger A = A$$
$$A^\dagger A A^\dagger = A^\dagger$$
$$\left(A^\dagger A\right)^T = A^\dagger A$$
$$\left(AA^\dagger\right)^T = AA^\dagger$$

Het bewijs zelf loopt van een leien dakje. We gaan van start met het gegeven:

$$AA^\dagger A = A$$

Beide leden transponeren is natuurlijk toegestaan en geeft:

$$A^T (AA^\dagger)^T = A^T$$

Omdat $(AA^\dagger)^T = AA^\dagger$ kunnen we schrijven:

$$A^T AA^\dagger = A^T$$

en omdat $A^T A$ een volle rang heeft kan men beide leden met het inverse vermenigvuldigen. Dit geeft:

$$A^\dagger = (A^T A)^{-1} A^T$$

---

2. Om aan te tonen dat $\mathrm{rang}(\nabla f(\mathbf{x})) = n$ volstaat het te zeggen dat $||\nabla f(\mathbf{x})\mathbf{y}||_2 \geq \beta ||\mathbf{y}||_2$

---

3. De Taylorontwikkeling van f in een omgeving van een punt $\mathbf{y}$ waarbij f van $\mathrm{R}^m \to \mathrm{R}^n$ is mij natuurlijk ook weer totaal onbekend. Hiervoor heb ik het boek 'Iterative solutions of nonlineair equations in several variables' bij de hand genomen. Op pagina 80 staat een taylorontwikkeling waarbij de rest–term als integraal geschreven wordt.

$$f(\mathbf{x}) - f(\mathbf{y}) - \nabla f(\mathbf{y})(\mathbf{x} - \mathbf{y}) \leq \int_0^1 (1-t)\nabla^{(2)} f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x})^2 dt$$

Hierop een middelwaardestelling loslaten geeft het gevraagde.

4a. Als $\mathbf{a}$ het minimum is van φ en $A_\mathbf{a} = \nabla f(\mathbf{a})$ dan is $A_\mathbf{a}^T f(\mathbf{a}) = \mathbf{0}$.

$$\begin{pmatrix} \frac{\partial f_1(\mathbf{a})}{\partial x_1} & \cdots & \frac{\partial f_m(\mathbf{a})}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial f_1(\mathbf{a})}{\partial x_n} & \cdots & \frac{\partial f_m(\mathbf{a})}{\partial x_n} \end{pmatrix} \begin{pmatrix} f_1(\mathbf{a}) \\ \vdots \\ f_m(\mathbf{a}) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Meer bepaald komt dit neer op bewijzen dat $\forall i \quad \frac{\partial f_1(\mathbf{a})}{\partial x_i}.f_1(\mathbf{a}) + \ldots + \frac{\partial f_m(\mathbf{a})}{\partial x_i}.f_m(\mathbf{a}) = 0$. Of anders gezegd dat $< \frac{\partial f(\mathbf{a})}{\partial x_i}, f(\mathbf{a}) >= 0$
Omdat $\mathbf{a}$ het minimum is van φ geldt

$$\left.\frac{\partial\varphi(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\mathbf{a}} = \sum_{j=1}^{m} 2.f_j(\mathbf{a}).\frac{\partial f_j(\mathbf{a})}{\partial x_i} = \mathbf{0}$$

Deze laatste som is 0 als en slechts als $<\frac{\partial f(\mathbf{a})}{\partial x_i}, f(\mathbf{a})>= 0$ en bijgevolg is bewezen dat $A_{\mathbf{a}}^T f(\mathbf{a}) = 0$.

4b. $A_k^T\left(f(\mathbf{x}_k) + A_k(\mathbf{x}_{k+1} - \mathbf{x}_k)\right) = \mathbf{0}$ en wel omdat $\mathbf{x}_{k+1} - \mathbf{x}_k = -\mathbf{s}_k = -A_k^{\dagger} f(\mathbf{x}_k)$. Zodoende wordt dit

$$A_k^T\left(f(\mathbf{x}_k) - A_k(A_k^T A_k)^{-1} A_k^T f(\mathbf{x}_k)\right) = \mathbf{0}$$

De A_k^T binnen brengen geeft

$$A_k^T f(\mathbf{x}_k) - A_k^T A_k(A_k^T A_k)^{-1} A_k^T f(\mathbf{x}_k) = \mathbf{0}$$

hierin is $A_k^T A_k(A_k^T A_k)^{-1} = I$ en bijgevolg staat er een trivialiteit. Met name

$$A_k^T f(\mathbf{x}_k) - A_k^T f(\mathbf{x}_k) = \mathbf{0}$$

4c.

$$\begin{aligned} A_k^T A_k(\mathbf{x}_{k+1} - \mathbf{a}) =& (A_{\mathbf{a}}^T - A_k^T) f(\mathbf{a}) + A_k^T \mathbf{r}_k \\ & \text{vanwege (8)} \\ =& -A_k^T f(\mathbf{a}) + A_k^T \mathbf{r}_k \end{aligned}$$

Dus $A_k(\mathbf{x}_{k+1} - \mathbf{a}) + f(\mathbf{a}) = \mathbf{r}_k$ en gehoopt wordt dat

$$||\mathbf{r}_k||_2 \leq \frac{1}{2}\gamma||\mathbf{x}_k - \mathbf{a}||_2^2$$

$$\begin{aligned} A_k^T \mathbf{r}_k &= A_k^T\left(A_k(x_{k+1} - a) + f(\mathbf{a})\right) \\ &= A_k^T\left(A_k(x_{k+1} - a) + f(\mathbf{a}) - A_k\mathbf{x}_k + A_k\mathbf{x}_k + f(\mathbf{x}_k) - f(\mathbf{x}_k)\right) \\ &= A_k^T\left(A_k(x_{k+1} - x_k) + f(\mathbf{x}_k)\right) + A_k^T\left(f(\mathbf{a}) - f(\mathbf{x}_k) - A_k(\mathbf{a} - \mathbf{x}_k)\right) \end{aligned}$$

Hieruit en uit (5) volgt dat

$$||\mathbf{r}_k||_2 = ||f(\mathbf{a}) - f(\mathbf{x}_k) - A_k(\mathbf{a} - \mathbf{x}_k)||_2 \leq \frac{1}{2}\gamma||\mathbf{a} - \mathbf{x}_k||_2^2$$

Bij stom toeval hetgeen bewezen moest worden.

5.
Dan moet nu bewezen worden dat $\frac{d}{dt}\varphi(\mathbf{x} - t\mathbf{s}_k)\Big|_t = 0 < 0$. Maar eerst is er een would–be lemma nodig:

$$AA^{\dagger} = I$$

want

$$
\begin{aligned}
P &= AA^{\dagger} \\
&= A.(A^TA)^{-1}A^T \\
PA &= A.(A^TA)^{-1}A^TA \\
&= A
\end{aligned}
$$

en bijgevolg is $P = I$. Dan nu eens kijken naar de afgeleide (lijkt sterk op het bewijs uit de cursus, p.62), noem $\psi(\tau) = \varphi(\mathbf{x} - \tau\mathbf{x})$

$$
\begin{aligned}
\psi'(0) &= \left.\frac{d}{dt}\varphi(\mathbf{x} - t\mathbf{s})\right|_{t=0} \\
&= \left.\frac{d}{dt}\sum_{i=1}^{n} f_i^2(\mathbf{x} - t\mathbf{s})\right|_{t=0} \\
&= -\sum_{j=1}^{n} \frac{\partial\varphi(\mathbf{x})}{\partial x_j}\mathbf{s}_j \\
&= -\sum_{j=1}^{n}\sum i = 1^n 2f_i(\mathbf{x})\frac{\partial f_i(\mathbf{x})}{\partial x_j}\mathbf{s}_j \\
&= -2f(\mathbf{x})^T\nabla f(\mathbf{x})\mathbf{s} \\
&= -2f(\mathbf{x})^T\nabla f(\mathbf{x})(\nabla f(\mathbf{x}))^{\dagger}f(\mathbf{x}) \\
&= -2||f(\mathbf{x})||_2^2 \\
&= -2\psi(0)
\end{aligned}
$$

Het nuttige aan bovenstaande is dat hiermee de lijnminimalisatie berekend kan worden.

7a. Toon aan dat $\mathbf{s} := (A^TA + \mu^2 I_n)^{-1}A^T\mathbf{b}$ met $\mu > 0$ de oplossing is van het kleinste-kwadraten-probleem

$$
\min_{x\in\mathrm{R}^n} ||M\mathbf{x} - \tilde{\mathbf{b}}||_2, \quad M := \begin{pmatrix} A \\ \mu I_n \end{pmatrix} \in \mathrm{R}^{(n+m)\times n} \quad \text{en} \quad \tilde{\mathbf{b}} := \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix} \in \mathrm{R}^{n+m}
$$

Zoals gegeven op p.52 in de cursus minimaliseert $\mathbf{x}$ $||A\mathbf{x} - \mathbf{b}||$ als en slechts als $\mathbf{x}$ een oplossing is van $A^TA\mathbf{x} = A^T\mathbf{b}$ dus $\mathbf{s}$ minimaliseert $||M\mathbf{x} - \tilde{\mathbf{b}}||$ als en slechts als $\mathbf{s}$ een oplossing is van $M^TM\mathbf{x} = M^T\tilde{\mathbf{b}}$.

$$
\begin{aligned}
\begin{pmatrix} A^T & \mu I_n \end{pmatrix}\begin{pmatrix} A \\ \mu I_n \end{pmatrix}\mathbf{x} &= \begin{pmatrix} A^T & \mu I_n \end{pmatrix}\begin{pmatrix} b \\ 0 \end{pmatrix} \\
(A^TA + \mu^2 I_n^2)\mathbf{x} &= A^T\mathbf{b} \\
\mathbf{x} &= A^T\mathbf{b}(A^TA + \mu^2 I_n)^{-1}
\end{aligned}
$$

deze laatste $\mathbf{x}$ minimalisert dus $||M\mathbf{x} - \tilde{\mathbf{b}}||$. En daarmee is bewezen wat bewezen moest worden.

7b. Hoe kan nu de QR–ontbinding van M efficiënt berekend worden ? In het geval dat de QR–ontbinding van A gekend is:

$$
M = \begin{pmatrix} A \\ \mu I_n \end{pmatrix} = \begin{pmatrix} QR \\ \mu I_n \end{pmatrix} = \begin{pmatrix} Q & 0 \\ 0 & I_n \end{pmatrix}\begin{pmatrix} R \\ \mu I_n \end{pmatrix}
$$

Nu kunnen we zeer snel de QR–ontbinding van $\begin{pmatrix} R & \mu I_n \end{pmatrix}^T$ berekenen door middel van Givensrotations. Dit omdat de matrix verschrikkelijk ijl is. We krijgen dan.

$$M = \begin{pmatrix} Q & 0 \\ 0 & I_n \end{pmatrix} Q'R'$$

Als de QR–ontbinding van A nog niet gekend is nemen we best de Givens-rotations benadering omdat de matrix betrekkelijk ijl is in den beginne.
Als er een QBV^T ontbinding ter beschikking is kan hetzelfde als hierboven gedaan worden. Maar de winst zit hem in het oplossen van de bovendriehoekmatrix. Deze bevat nu veel meer 0 en kan bijgevolg rapper opgelost worden (in deze oefening zal dat dus geen verschil maken omdat $n = 3$) !

Om nu het gegeven probleem op te lossen moeten we de gewogen functionaal

$$J(\mathbf{x}) : \quad \mathrm{R}^3 \to \mathrm{R}^m \quad := \sum_m \frac{(p_k - \pi_k)^2}{\pi_k(1-\pi_k)}$$

minimaliseren. Om dit te kunnen doen op de gegeven methode moet er een f gevonden worden zodat $J(\mathbf{x}) = ||f(\mathbf{x})||_2^2$. Merk op dat p_k eigenlijk een functie is van $\mathrm{R}^3 \to \mathrm{R}^m$ en dat π_k beshouwd kan worden als een element van R^m.

$$p_i(A, B, \gamma) := 1 - e^{e^{i\lambda}\frac{b}{\lambda}(e^{\lambda/2} - e^{-\lambda/2}) + A}$$

f wordt uiteindelijk iets als

$$f(\mathbf{x})_i : \quad \mathrm{R}^3 \to \mathrm{R} \quad := \frac{p_i(\mathbf{x}) - \pi_i}{\sqrt{\pi_i(1-\pi_i)}}$$

Nu moet nog de jacobiaan $\nabla f(\mathbf{x})$ berekend kunnen worden.

$$\nabla f(\mathbf{x})_{ij} := \frac{\partial f(\mathbf{x})_i}{\partial x_j}$$

Dit geeft de volgende matrix

$$\nabla f(A, B, \lambda) = \begin{pmatrix} \frac{\partial f(A,B,\lambda)_1}{\partial A} & \frac{\partial f(A,B,\lambda)_1}{\partial B} & \frac{\partial f(A,B,\lambda)_1}{\partial \lambda} \\ \vdots & \vdots & \vdots \\ \frac{\partial f(A,B,\lambda)_m}{\partial A} & \frac{\partial f(A,B,\lambda)_m}{\partial B} & \frac{\partial f(A,B,\lambda)_m}{\partial \lambda} \end{pmatrix}$$

Laat ons deze partiëele afgeleiden eens berekenen.

$$\frac{\partial f(A, B, \lambda)_i}{\partial A} = \frac{e^{e^{i\lambda}\frac{B}{\lambda}(e^{-\lambda/2} - e^{\lambda/2}) - A}}{\sqrt{\pi_i(1-\pi_i)}}$$

$$\frac{\partial f(A, B, \lambda)_i}{\partial B} = \frac{e^{e^{i\lambda}\frac{B}{\lambda}(e^{-\lambda/2} - e^{\lambda/2}) - A + i\lambda}(e^{-\lambda/2} - e^{\lambda/2})}{\lambda\sqrt{\pi_i(1-\pi_i)}}$$

$$\frac{\partial f(A, B, \lambda)_i}{\partial \lambda} = \frac{Be^{e^{i\lambda}\frac{B}{\lambda}(e^{-\lambda/2} - e^{\lambda/2}) - A + i\lambda}(e^{-\lambda/2}(2i\lambda + \lambda - 2) - e^{\lambda/2}(2i\lambda - \lambda - 2))}{2\lambda^2\sqrt{\pi_i(1-\pi_i)}}$$

De paraboolbenadering voor de lijnminimalisatiestrategie (b) is gegeven als volgt

$$p(t) = \psi(0) - 2t\psi(0) + t^2(\psi(t) + \psi(0))$$

Deze komt met ψ overeen in $\psi(0)$, $\psi(t)$ en $\psi'(0)$. Zoals gewenst met andere woorden. Het minimum van deze parabool wordt aangetroffen in

$$t^* = \frac{\psi(0)}{\psi(0) + \psi(t)}$$

Opmerkingen.

- Het is me opgevallen dat δ_c bepaalt hoe hard hij rekening houdt met de richtingsafgeleiden. Als deze vast gekozen wordt en te groot durft deze methode wel eens te ver te springen en uiteindelijk te divergeren.
- $m_c(\mathbf{x}_c) = ||f(\mathbf{x}_c) + \nabla f(\mathbf{x}_c)(\mathbf{x}_c - \mathbf{x}_c)||_2^2 = ||f(\mathbf{x}_c)||_2^2 = \varphi(\mathbf{x}_c)$
- Experiment 4 is, zoals aangeraden, het eerst uitgevoerd en deze klopt inderdaad. Na 1 slag zit hij plat op het juiste antwoord.
- De bepaling van μ gebeurt opnieuw door een iteratie. Deze is te vinden in `CalcMu`.
- Ik heb de gewogen benadering gedaan en het resultaat is op het eerste gezicht gewoonweg fout. Doch na een halve minuut denken keek ik eens naar de 'betrouwbaarheid' van de meting. Met name de factor $\pi_k(1 - \pi_k)$ had ik over het hoofd gezien.
- Ik heb voor de zekerheid ook een niet-gewogen benadering toegevoegd en deze lijkt - grafisch dan - beter te kloppen.
- Bij het toevoegen van normaal verdeelde ruis met verschillende varianties komt er uit de bus dat de functionaalwaarde van de dichtste benadering de wortel is van de toegevoegde ruis. M.a.w als de variantie van de ruis 10^{-11} is dan heeft de gewogen beste benadering een norm $3,16931.10^{-6}$. Zie tabelletje

10^{-5}	$0,100611$
10^{-8}	$0,00316955$
10^{-11}	$0,000100288$
10^{-14}	$3,16931.10^{-6}$

- laat ons eens kijken naar het toevoegen van ruis en niet gewogen benaderen. Dat de normen er een factor 100 naast zitten is normaal omdat ik met een vector van 100 groot werk.

10^{-5}	$0,0313603$
10^{-8}	$0,000992079$
10^{-11}	$3,13727.10^{-5}$
10^{-14}	$9,92091.10^{-7}$

Files:

- `MT-11.DAT` 10^{-11} normaal verdeelde ruis, niet gewogen
- `MT-8.DAT` 10^{-8} normaal verdeelde ruis, niet gewogen
- `MT-5.DAT` 10^{-5} normaal verdeelde ruis, niet gewogen
- `MT-14.DAT` 10^{-14} normaal verdeelde ruis, niet gewogen
- `MT-11g.DAT` 10^{-11} normaal verdeelde ruis, wel gewogen
- `MT-8g.DAT` 10^{-8} normaal verdeelde ruis, wel gewogen
- `MT-5g.DAT` 10^{-5} normaal verdeelde ruis, wel gewogen
- `MT-14g.DAT` 10^{-14} normaal verdeelde ruis, wel gewogen
- `MT-G.XLS` grafiek van de gewogen benadering
- `MT.XLS` grafiek van de niet gewogen benadering

References en hulpmiddelen:

- Ortega and Rheinboldt, "Iterative solutions of Nonlinear Equations in Several Variables". (1970)
- Gene H. Golub, Charles F. Van Loan, "Matrix Computations", (1989), ISBN 0-8018-3772-3