## KLEINSTE KWADRATEN

1) De volgende tabel geeft de evolutie van het bevolkingscijfer van de USA weer:

| | | | | | | |
|---|---|---|---|---|---|---|
| $t$ | 1900 | 1910 | 1920 | 1930 | 1940 | 1950 |
| $y$ | 75.995 | 91.972 | 105.711 | 123.203 | 131.669 | 150.697 |
| $t$ | 1960 | 1970 | 1980 | 1990 | 2000 | |
| $y$ | 179.323 | 203.212 | 226.505 | 249.633 | 281.422 | |

Bedoeling is om een voorspelling voor het bevolkingscijfer in 2020 te maken.

- Fit een veelterm van graad 3 volgens het principe van de kleinste kwadraten door de gegevens.
- Fit een veelterm van graad 8 volgens het principe van de kleinste kwadraten. Wat stel je vast tussen 2010 en 2020?
- Herbereken de veelterm van graad 3, nu gebruik makend van de basisfuncties $((t-1950)/50)^i$, $i = 0, \ldots, 3$. Vergelijk de conditionering met die waar de basisfuncties $t^i$ gebruikt werden.

2) Gegeven zijn de lineair onafhankelijke basisfuncties

$$\phi_{3i}(x) = \binom{3}{i} x^i (1-x)^{3-i} \qquad i = 0, 1, 2, 3.$$

Bereken de optimale lineaire combinatie

$$\phi(x) = \lambda_0 \phi_{3,0}(x) + \lambda_1 \phi_{3,1}(x) + \lambda_2 \phi_{3,2}(x) + \lambda_3 \phi_{3,3}(x)$$

die de fout

$$\sqrt{\sum_{j=0}^{19} (\phi(x_j) - y_j)^2}$$

minimaal maakt voor de data gegeven in onderstaande tabel. Geef ook het conditiegetal van de rechthoekige matrix in het kleinste kwadratenprobleem. Plot de berekende $\phi(x)$ samen met de datapunten.

| $j$ | $x_j$ | $y_j$ | $j$ | $x_j$ | $y_j$ |
|---|---|---|---|---|---|
| 0 | 0.0 | $-0.80$ | 10 | 3.6 | 0.74 |
| 1 | 0.6 | $-0.34$ | 11 | 4.7 | $-0.82$ |
| 2 | 1.5 | 0.59 | 12 | 5.2 | $-1.27$ |
| 3 | 1.7 | 0.59 | 13 | 5.7 | $-0.92$ |
| 4 | 1.9 | 0.23 | 14 | 5.8 | $-0.92$ |
| 5 | 2.1 | 0.10 | 15 | 6.0 | $-1.04$ |
| 6 | 2.3 | 0.28 | 16 | 6.4 | $-0.79$ |
| 7 | 2.6 | 1.03 | 17 | 6.9 | $-0.06$ |
| 8 | 2.8 | 1.50 | 18 | 7.6 | 1.00 |
| 9 | 3.0 | 1.44 | 19 | 8.0 | 0.00 |

3) We gebruiken dezelfde data als in de vorige vraag maar vervangen de functies $\phi_{3,i}(x)$ enerzijds door de veeltermen

$$T_i(x) = 2xT_{i-1}(x) - T_{i-2}(x) \qquad i = 0, \ldots, 3$$
$$T_0(x) = 1 \qquad T_1(x) = x$$

en anderzijds door de monomiale basis $x^i, i = 0, \ldots, 3$. Bereken optimale modellen van de vorm

$$\sum_{i=0}^{3} \tau_i T_i(x)$$

en

$$\sum_{i=0}^{3} \mu_i x^i$$

volgens het kleinste-kwadratenprincipe. Vergelijk de conditiegetallen en bereken de residu vectoren.

4) Aan het National Institute of Standards (NIST) te Washington werden een aantal standaardisatietesten uitgevoerd waarvan eentje de dataset opleverde die je kan downloaden van `www.itl.nist.gov/div898/strd/lls/data/Filip.shtml`. Het gaat om 82 observaties van een grootheid $y(x)$ die zich redelijk laat benaderen door een veeltermmodel van graad 10 (wees voorzichtig met het correct inlezen van de $x$- en $y$-waarden). Bereken de coefficienten van dat model.

5) Gegeven zijn de datapunten

| $x_i$ | 1.1 | 1.6 | 11.4 | 4.1 | 5.3 | 17.5 | 9.4 | 11.5 | 12.1 |
|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 7.9 | 24.8 | -28.8 | 42.6 | 29.6 | -34.6 | -3.1 | -28.7 | -39.6 |

Fit een veelterm van graad $k$ door de datapunten waarbij je $k$ varieert van 1 tot 5 (voor $k = 8$ bekom je de interpolerende veelterm). Plot voor elke $k$ de data en het model en druk tevens de waarde van de norm van het residu $r$ af (niet het kwadraat van die norm):

$$r^2 = \sum_{i=0}^{8} \left( y_i - \sum_{j=0}^{k} a_j x_i^j \right)^2 .$$

Waak ook over de conditionering van het probleem, m.a.w. druk het conditiegetal af. Welk polynomiaal model zou je verkiezen?