Uppgift 2: Minstakvadratmetoden och CO_2 mätningar

Här ska vi analysera mätningar av halten av koldioxid i atmosfären med minstakvadratmetoden. För detta ska du själv hämta data från en öppen databas. Vi rekommenderar att ni fokuserar på data från de senaste 60-70 åren som beskriver den data brukar kallas *Keelingkurvan*. Det finns även annan data över kortare eller längre perioder, men denna data inte kan approximeras med enkla polynom och passar därför inte lika bra in i här uppgiften. Om ni väljer att använda denna form av data så får ni en extrauppgift; ni ska hitta en funktions-bas som bättre beskriver er data. Den mätdata ni ska använda kan ni hämta från till exempel https://scrippsco2.ucsd.edu. Här kan ni hitta filer på ”csv” format, vilka kan läsas med funktionen *readtable*, som returnerar ett table-objekt. Ett exempel på hur *readtable* kan användas är om man vill ta fram kolumnerna ”Date1” och ”CO2” från en fil *monthly_in_situ_co2_mlo.csv*:

```
TABLE=readtable('monthly_in_situ_co2_mlo.csv'); % Skapar table-objekt
t=[TABLE.Date1];  % Hämtar kolumnen Date1 från TABLE och
                  % gör om till en vektor mha ”[…]”.
y=[TABLE.CO2];    % Hämtar kolumnen CO2 från TABLE och
                  % gör om till en vektor mha ”[…]”.
```

Notera att det allmänt kan vara krångligt att räkna med datum (tex, hur många dagar är det mellan 21 november 2013 och 1 maj 2023), men i vissa av csv filerna finns det datum på decimalform som kan vara användbara!

Eran uppgift är:

A. Att hämta data från CO_2 mätningar från en öppen databas. Notera att denna data ska laddas upp på Canvas under uppgiften ”CO2 Data till projekt 2”
B. Läs in datafilen i Matlab och finn en tidsvektor, $\mathbf{t}$, samt en datavektor, $\mathbf{y}$, med resultat från CO_2 mätningar. Notera att om ni använder ett annat format än csv har ni själva ansvar för att läsa och processa den data ni använder. Ni måste dessutom gå igenom filen själva för att se till att den data ni använder motsvarar just en tidsvektor, samt en relevant datavektor.
C. Skriv ut vilken data ni har i $\mathbf{t}$-vektorn och $\mathbf{y}$ -vektorn. Se till att all värden i vektorerna motsvarar mätdata. Notera att koldioxidhalten alltid måste vara ett positivt reellt tal och tidsvektorn måste innehålla rimliga datum. I koden ni fått har jag lagt in ett par rader för att hjälpa er ta bort vektor-element som inte motsvarar riktig data.
D. Nedan ska ni göra minsta numeriska kvadratanpassningar. Dessa är inte exakta och kan vara känsliga för avrundningsfel. Framför allt om kolumnerna i matrisen ni skapar är olika stora. För att undvika dessa problem ska vi skapa en normaliserad tids vektor $\mathbf{s}$, som är en funktion av tiden. In ni t.ex. använder en tidsvektor med årtal kan ni använda, $s(t) = (t - 1950)/50$, så att $s = 0$ motsvara år 1950 och $s = 1$ motsvara år 2000. Därmed har en ett par vektorer, $\mathbf{y}$ och $\mathbf{s}$, som tillsammans beskriver punkter längs funktionen, $y(s)$. Ni har även ett par vektorer, $\mathbf{y}$ och $\mathbf{t}$, som tillsammans beskriver punkter längs funktionen, $y(t)$.
E. Nu ska vi ställa upp de matriser och vektorer som vi behöver för att finna en rät linje som är en minsta kvadratanpassningar till vår mätdata, $y(s)$. Låt $\boldsymbol{\beta}$ vara en vektor som beskriver parametrarna i den räta linjens ekvation. Skapa matrisen X_1 och vektorn $\mathbf{b}_1$, så

att den vektor, $\boldsymbol{\beta}$, som minimerar $(X_1\boldsymbol{\beta} - \boldsymbol{b}_1)$ kommer att representerar en minsta-kvadrat approximation av våra mätdata, $y(s)$.

F. Skapa en minstakvadrat anpassning av $y(s)$ till en rät linje. Rita både mätdata och anpassningen i samma graf med den vanliga tiden, **t**, på x-axeln.
VIKTIGT: Här är det inte tillåtet att använda färdiga funktioner för kurvanpassning eller optimering. Anpassningen ska istället skapas genom att först skapa ett ekvationssystem med lika många ekvationer som obekanta. Därefter får ni använda funktionerna *mldivide*, eller *inv* för att lösa ekvationssystemet.

G. Skapa en minstakvadrat anpassning av $y(s)$ till ett andragradspolynom som funktion av den normaliserade tids-variabeln, **s**. Rita både mätdata och anpassningen i samma graf med den vanliga tiden, **t**, på x-axeln.
VIKTIGT: Här är det inte tillåtet att använda färdiga funktioner för kurvanpassning eller optimering. Anpassningen ska istället skapas genom att först skapa ett ekvationssystem med lika många ekvationer som obekanta. Därefter får ni använda funktionerna *mldivide*, eller *inv* för att lösa ekvationssystemet.

H. Skapa en minstakvadrat anpassning av $y(s)$ till ett tredjegradspolynom som funktion av den normaliserade tids-variabeln, **s**. Rita både mätdata och anpassningen i samma graf med den vanliga tiden, **t**, på x-axeln.
VIKTIGT: Här är det inte tillåtet att använda färdiga funktioner för kurvanpassning eller optimering. Anpassningen ska istället skapas genom att först skapa ett ekvationssystem med lika många ekvationer som obekanta. Därefter får ni använda funktionerna *mldivide*, eller *inv* för att lösa ekvationssystemet.

I. (*Bara för er som inte använder CO_2 data från de senaste 60-70 åren*)
Om dessa anpassningar inte ger en rimlig uppskattning av er CO_2 data så får ni föreslå en annan funktions-bas som ger en bättre beskrivning av er data (kanske lägga till en exponential- eller en sinus-funktion). Använd den nya basen och gör en anpassning, samt visa resultat i en graf.

Frågor:

1. Beskriv med egna ord hur de tre kurvorna beskriver mätdata. Framför allt, blir lösningen lite eller mycket bättre när vi går från en rät linje till en andragradsfunktion? Blir den mycket bättre när vi går från en andragradsfunktion till en tredjegradsfunktion?
2. Kan man använda dessa anpassningar för att uppskatta utsläppen om 2 år? Motivera ditt svar.
3. Kan man använda dessa anpassningar för att uppskatta utsläppen om 6 månader? Motivera ditt svar.
4. Kan man använda dessa anpassningar för att uppskatta utsläppen om 50 år? Motivera ditt svar.