

Verslag Python Challenge 1

Aiko Decaluwe, Viktor van Nieuwenhuize, Fien Dewit (Groep 4)
Opleiding: Tweede Bachelor Fysica en Sterrenkunde
Vak: Statistiek en Gegevensverwerking

Datum: 17 November 2020

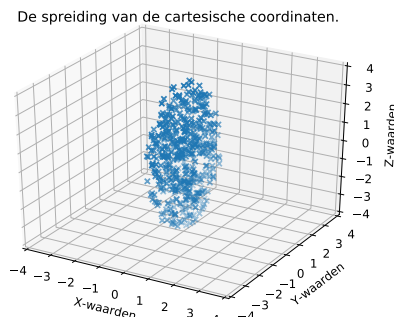
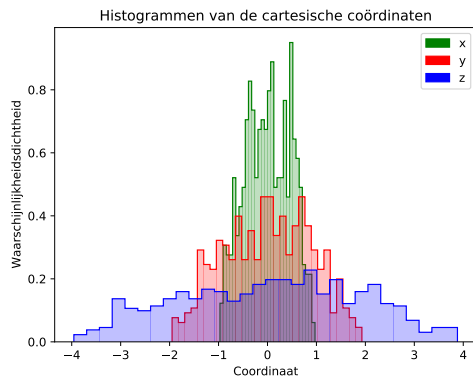
Inleiding

In dit verslag bekijken we hoe fouten propageren bij overgang van sferische naar cartesische coördinaten. De gegeven dataset met sferische coördinaten werd omgezet naar cartesische aan de hand van volgende formules:

$$\begin{cases} x &= r \cos \theta \cos \phi \\ y &= r \sin \theta \sin \phi \\ z &= r \cos \theta \end{cases}$$

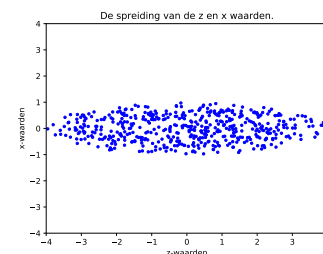
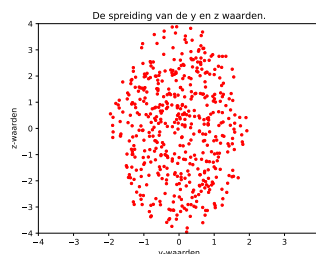
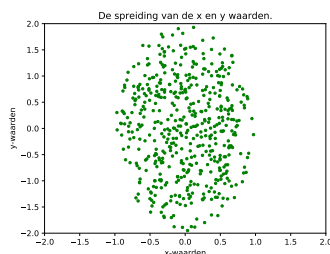
Ten slotte wordt ook de invloed van een systematische fout onderzocht.

Opdracht 1



(a) XYZ spreiding van de cartesische coördinaten als histogram (b) XYZ spreiding van de cartesische coördinaten als scatterplot

Figuur 1: 3D verdeling van de cartesische coördinaten.



(a) XY spreiding van de cartesische coördinaten (b) YZ spreiding van de cartesische coördinaten (c) ZX spreiding van de cartesische coördinaten

Figuur 2: 2D doorsnedes van de cartesische coördinaten.

We maken gebruik van de formules gegeven in de inleiding om van sferische coördinaten (r , θ en ϕ) over te gaan naar cartesische coördinaten (x , y , en z). De spreiding van de cartesische coördinaten wordt weergegeven in de figuren 1 en 2 hierboven.

Uit Figuur 1 (a) kunnen we de spreiding van de coördinaten zien. De x -, y - en z -coördinaten lijken allemaal vrijwel uniform verdeeld met een lichtjes lagere dichtheid aan de rand. De x -coördinaten liggen allemaal binnen het interval $[-1, 1]$, de y -coördinaten liggen allemaal binnen het interval $[-2, 2]$ en de z -coördinaten binnen het interval $[-4, 4]$.

In zowel Figuur 1 (b) als in Figuur 2 kan men zien dat de correlatie tussen xy en yz vrij klein is, maar de correlatie tussen x en z lijkt een stuk groter te zijn.

Opdracht 2

Vervolgens werd er aangenomen dat de fout op r , θ en ϕ gelijk zijn aan $\sigma_r = 0.001$, $\sigma_\theta = 0.003$ en $\sigma_\phi = 0.005$ respectievelijk. Omdat deze fouten doorheen de coördinatentransformatie gepropageerd zijn, en er dus een correlatie ontstond tussen x , y en z , werd er nadien een algemene formule toegepast voor de foutenpropagatie. Deze is gelijk aan

$$C_{X,ij} = \sum_{kl} \frac{\partial X_i}{\partial R_k} \frac{\partial X_j}{\partial R_l} C_{R,kl}$$

Hierin is C_R de covariantiematrix voor de sferische coördinaten en staat X_i voor x , y en z en R_i voor r , θ en ϕ . Om het programmeerwerk wat eenvoudiger te maken werd alles berekend door de vergelijking in matrices te gieten en op deze manier $C_{X,ij}$ te berekenen als matrixvermenigvuldiging. Deze ziet er uit als volgt:

$$C_{X,ij} = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \phi} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \phi} \end{bmatrix} \cdot \begin{bmatrix} \sigma_r^2 & 0 & 0 \\ 0 & \sigma_\theta^2 & 0 \\ 0 & 0 & \sigma_\phi^2 \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \phi} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \phi} \end{bmatrix}^\top$$

Aangezien er 504 sferische coördinaten in de dataset zijn, verkregen we ook 504 covariantiematrices in deze vorm, één voor elke coördinaat.

Opdracht 3

Voor het plotten werden de matrices die we berekend hebben in opdracht 2 gebruikt. De diagonaalelementen van $C_{X,ij}$ zijn de varianties. De fouten op x , y en z zijn dan:

$$\begin{aligned} \sigma_x^2 &= \sqrt{C_{X,11}} = \sqrt{\text{cov}(x, x)} \\ \sigma_\theta^2 &= \sqrt{C_{X,22}} = \sqrt{\text{cov}(y, y)} \\ \sigma_\phi^2 &= \sqrt{C_{X,33}} = \sqrt{\text{cov}(z, z)} \end{aligned}$$

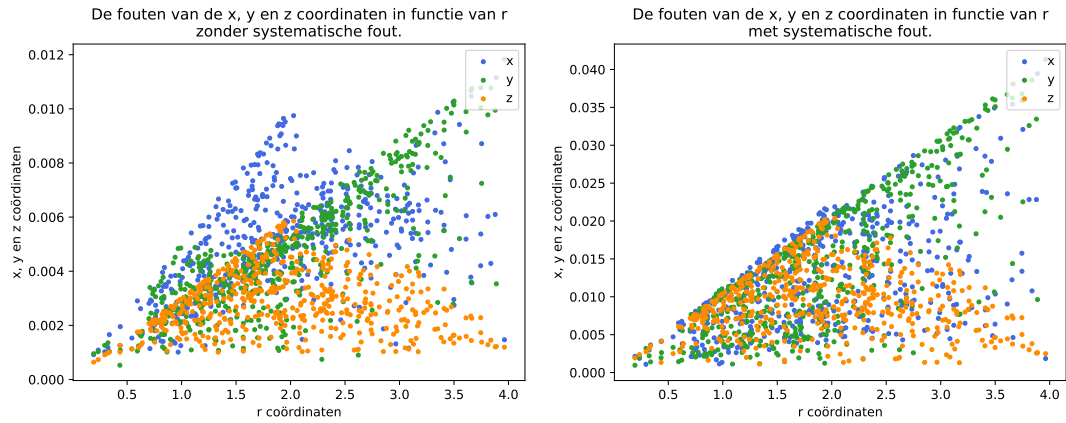
Vervolgens kunnen de correlaties bepaald worden via:

$$\begin{aligned} \rho_{xy} &= \frac{C_{X,12}}{\sqrt{C_{X,11} \cdot C_{X,22}}} = \frac{C_{X,12}}{\sqrt{C_{X,11} \cdot C_{X,22}}} = \frac{\text{cov}(x, y)}{\sqrt{\sigma_x^2 \cdot \sigma_y^2}} \\ \rho_{yz} &= \frac{C_{X,23}}{\sqrt{C_{X,22} \cdot C_{X,33}}} = \frac{C_{X,32}}{\sqrt{C_{X,22} \cdot C_{X,33}}} = \frac{\text{cov}(y, z)}{\sqrt{\sigma_y^2 \cdot \sigma_z^2}} \\ \rho_{zx} &= \frac{C_{X,31}}{\sqrt{C_{X,33} \cdot C_{X,11}}} = \frac{C_{X,13}}{\sqrt{C_{X,33} \cdot C_{X,11}}} = \frac{\text{cov}(z, x)}{\sqrt{\sigma_z^2 \cdot \sigma_x^2}} \end{aligned}$$

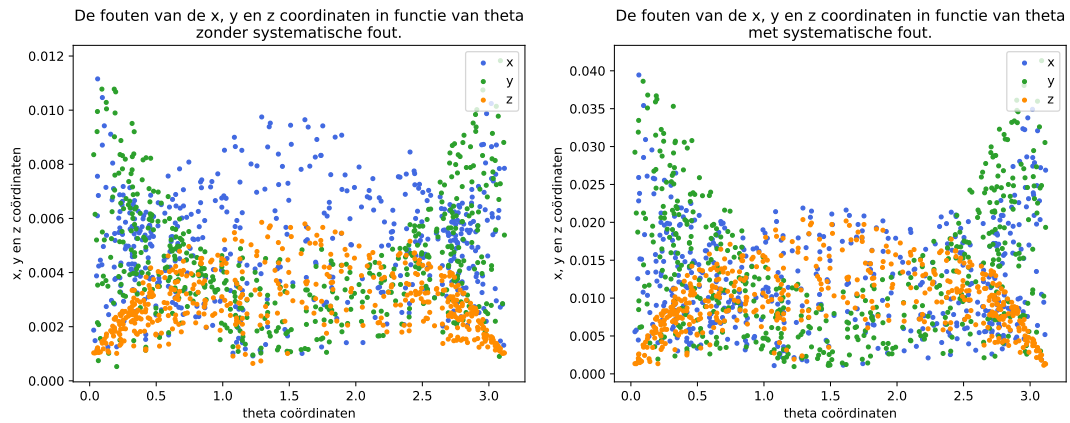
Hierna werden zowel de fouten als de correlaties uitgezet in functie van de sferische coördinaten. Telkens werden alle datapunten voor x , y en z op hetzelfde plot geplaatst. Dit is te zien aan de linkerkanten van figuren 3,4 en 5.

Opdracht 4

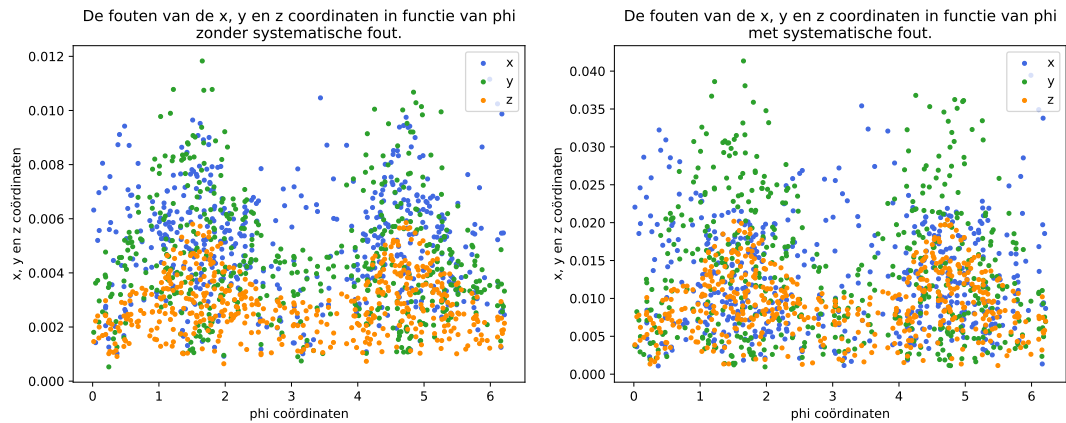
In opdracht 4 werden de berekeningen uit opdracht 3 herhaald, deze keer bevat de covariantiematrix van de sferische coördinaten een systematische fout $S = \pm 0.01$.



Figuur 3: *Vergelijking van de fouten in functie van r. Zonder systematische fout links, met rechts.*



Figuur 4: *Vergelijking van de fouten in functie van θ . Zonder systematische fout links, met rechts.*



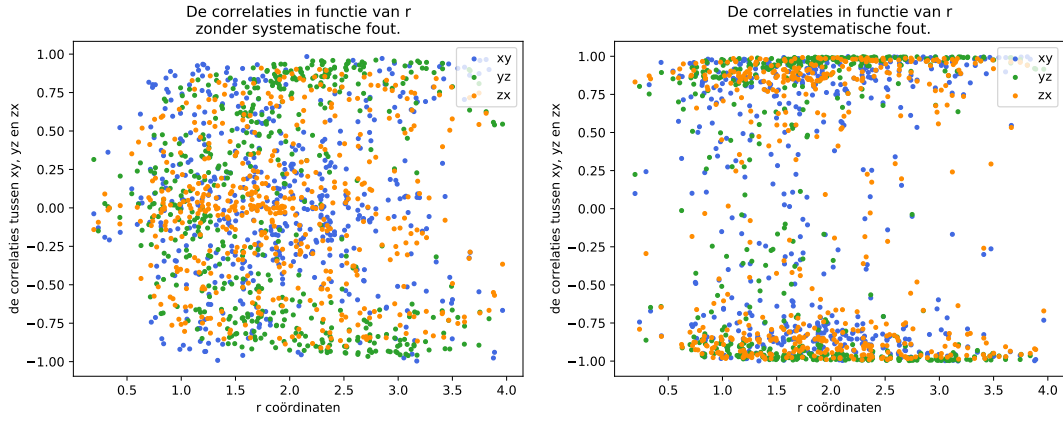
Figuur 5: *Vergelijking van de fouten in functie van ϕ . Zonder systematische fout links, met rechts.*

Het grootste verschil tussen de fouten met en zonder systematische fout, is dat het interval waarin deze liggen veel groter is indien er een systematische fout meespeelt. Met ander woorden, de fouten op x , y en z zijn veel groter indien er een systematische fout optreedt.

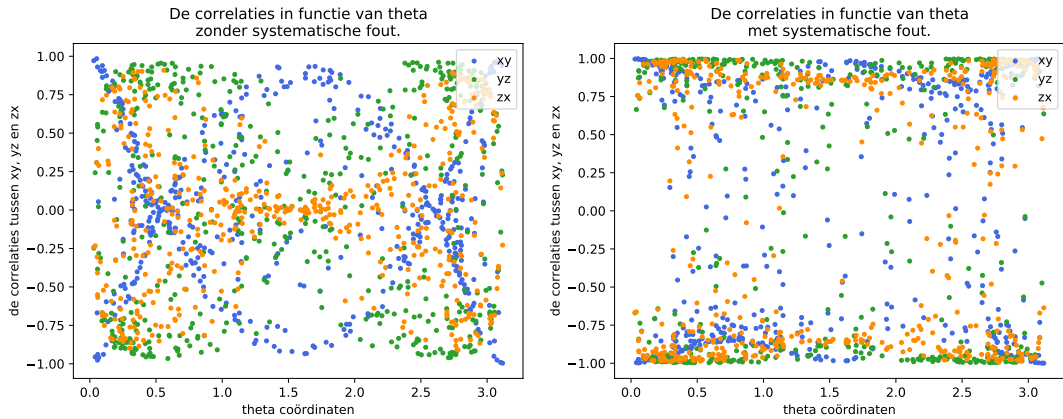
Verder is er in figuur 3 zichtbaar dat alle waarden begrensd worden door een lineaire functie en dat de helling van deze functie verschillend is indien er niet gelet wordt op de systematische fout (linkerkant). Indien er wel een systematische fout is, worden de 3 verdelingen door quasi dezelfde rechte begrensd.

Wanneer we dit grondiger onderzoeken, blijkt dat er in de plotjes met systematische fout veel meer overlap is tussen de waarden voor x en z als in de plotjes zonder systematische fout. De systematische fout verandert dus niet alleen de grootte van de fouten maar ook zijn verdeling.

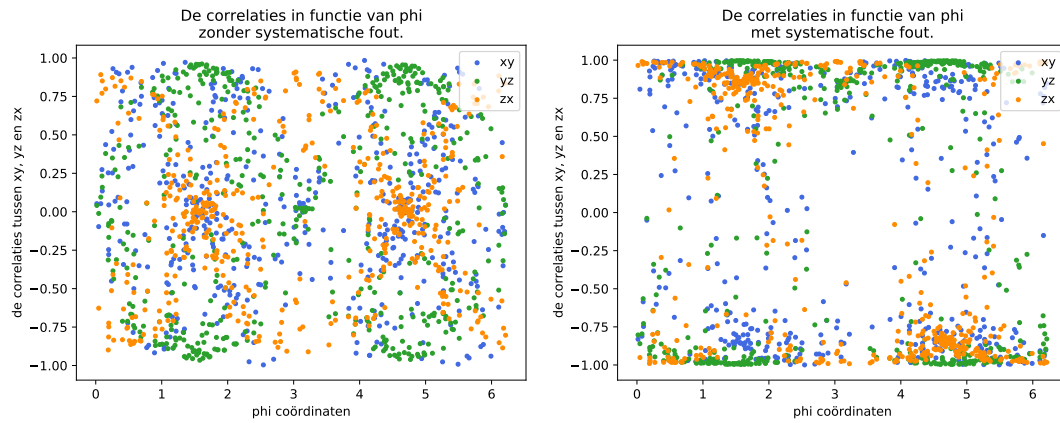
Verder lijken de plotjes wel vrij sterk op elkaar. Het grootste verschil is dat de fouten vergroten, en dat de verdeling van σ_x merkbaar verandert zoals hierboven beschreven werd.



Figuur 6: *Vergelijking van de correlaties in functie van r . Zonder systematische fout links, met rechts.*



Figuur 7: *Vergelijking van de correlaties in functie van θ . Zonder systematische fout links, met rechts.*



Figuur 8: *Vergelijking van de correlaties in functie van ϕ . Zonder systematische fout links, met rechts.*

In Figuur 6, 7 en 8 zijn de correlaties tussen x en y , y en z en z en x terug te vinden, telkens uitgezet in functie van een sferische coördinaat.

Het meest opmerkelijke verschil dat te zien is tussen de correlaties met en zonder systematische fout, is dat het lijkt alsof alle waarden in de plotjes met systematische fout naar de uitersten zijn geduwd. De dichtheid van waarden rond 0 is opmerkelijk kleiner en de meeste waarden liggen rond -1 of 1 . Het is wel duidelijk dat de vorm van de verdeling samengedrukt is, maar verder is deze vorm niet opmerkelijk veranderd.

Hieruit kunnen we concluderen dat de coördinaten veel sterker correleren wanneer er een systematische fout in het spel is.