

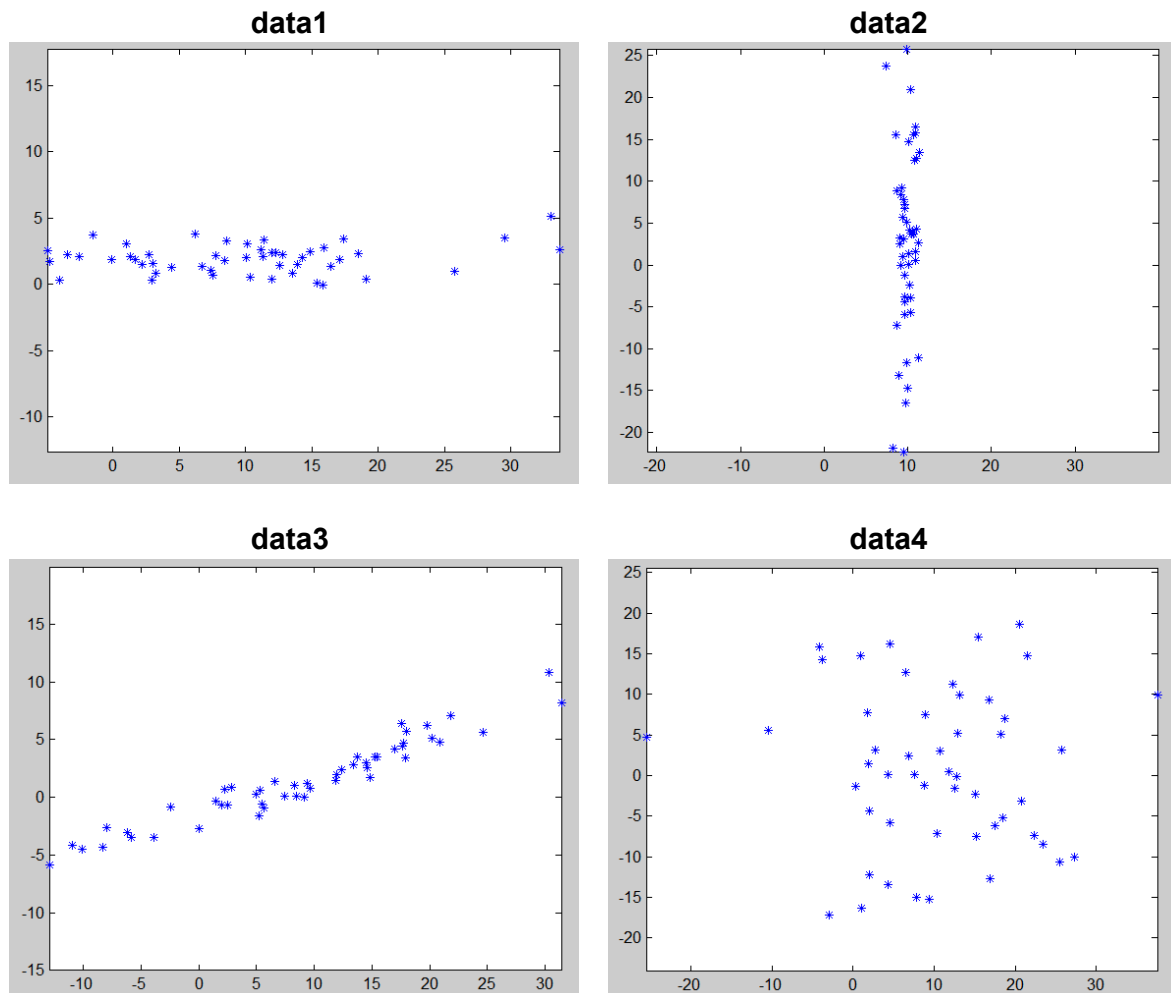
# Medizinische Bildverarbeitung

## Ausarbeitung der 1. Übung: PCA

Gruppenmitglieder: Felix Schuller (1025256) und Johannes Riedmann (0926649)

### 1. Kovarianzmatrix

b) Plot und Interpretation der unterschiedlichen C



Mittels *ourCov.m* berechnete Kovarianzmatrizen der einzelnen Datensätze:

ourCov(data1)		ourCov(data2)		ourCov(data3)		ourCov(data4)	
81.4252	1.7825	0.7404	1.2033	110.4927	36.1014	115.8434	-4.5221
1.7825	1.1887	1.2033	116.3126	36.1014	12.7527	-4.5221	96.5867

Die Struktur der Kovarianzmatrizen ist folgende:

$$C = \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(Y,X) & Var(Y) \end{pmatrix}$$

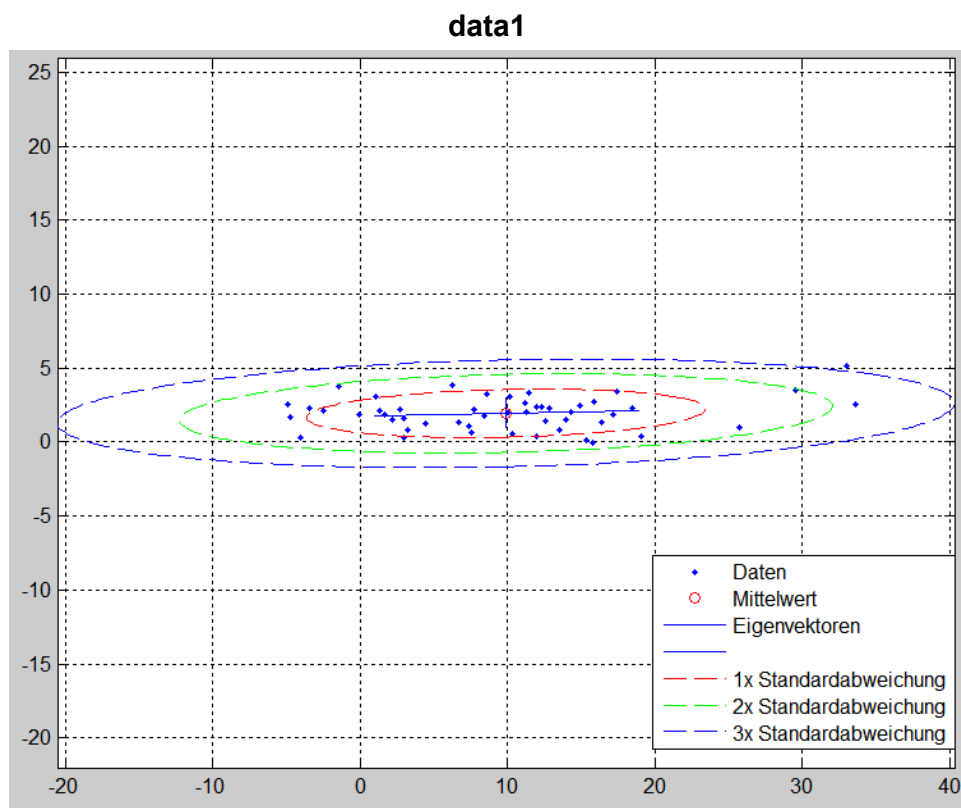
Var(X bzw. Y) ... Varianz von X bzw. Y

Cov(X, Y bzw. Y, X) ... Kovarianzen

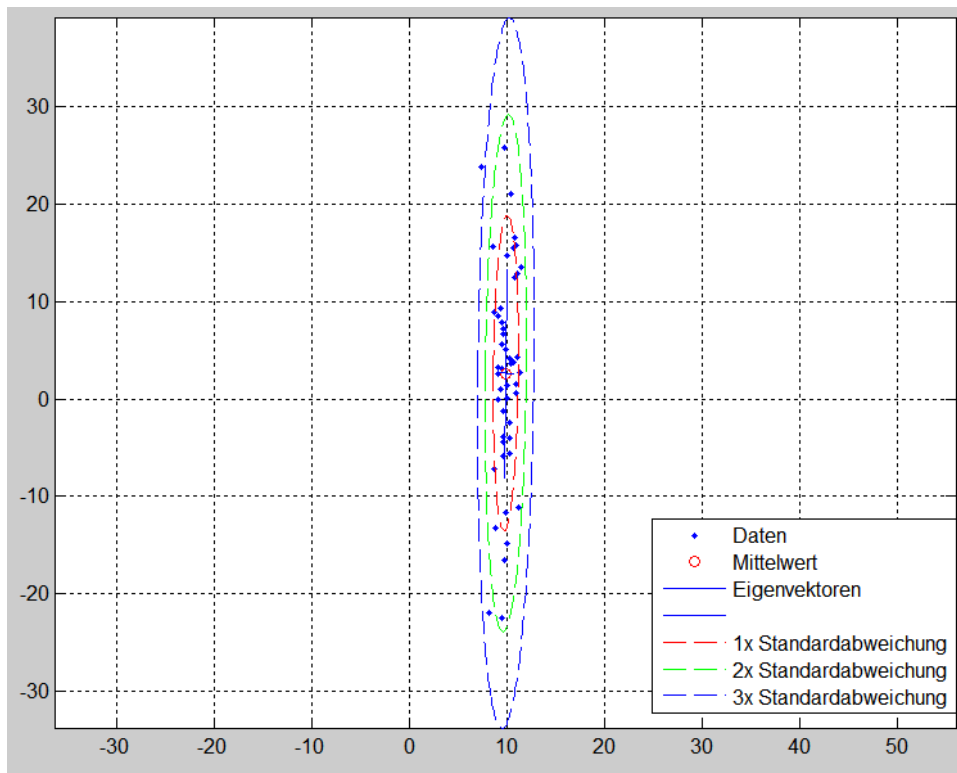
C zeigt für data1 eine hohe Var(X) an, was beim Plotten der Daten gut nachvollziehbar ist, da diese eher entlang der x-Achse gestreut sind. Bei data2 zeigt sich selbiges Verhalten, nur diesmal für die y-Achse, entsprechend der hohen Var(Y). data3 zeigt eine größere Varianz entlang der X-Achse als entlang der Y-Achse, was durch die große horizontale, jedoch niedrige vertikale Ausbreitung der Punkte im Plot sichtbar wird. data4 verfügt über eine hohe Streuung entlang beider Achsen und eine negative Kovarianz der x/y-Werte, (x steigt, y fällt und umgekehrt) diese ist am Plot etwas schwieriger zu erkennen, jedoch ein großes Streuverhalten entlang beider Achsen ist sofort erkennbar.

## 2. PCA

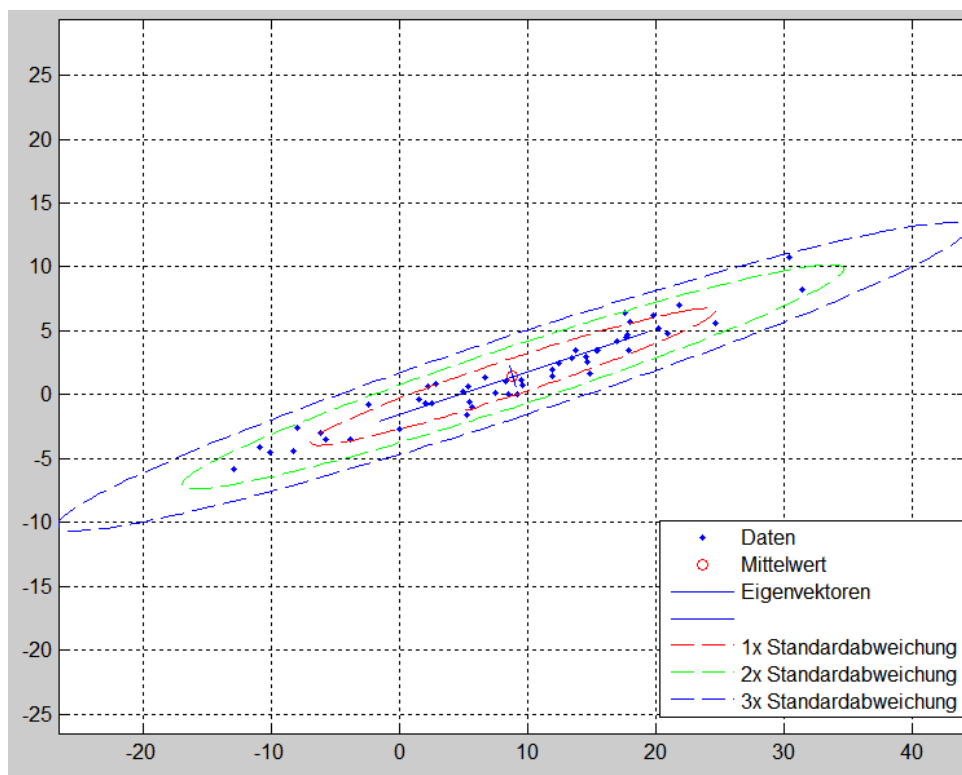
a. Plot der Daten in *daten.mat* mittels *plot2DPCA.m*:

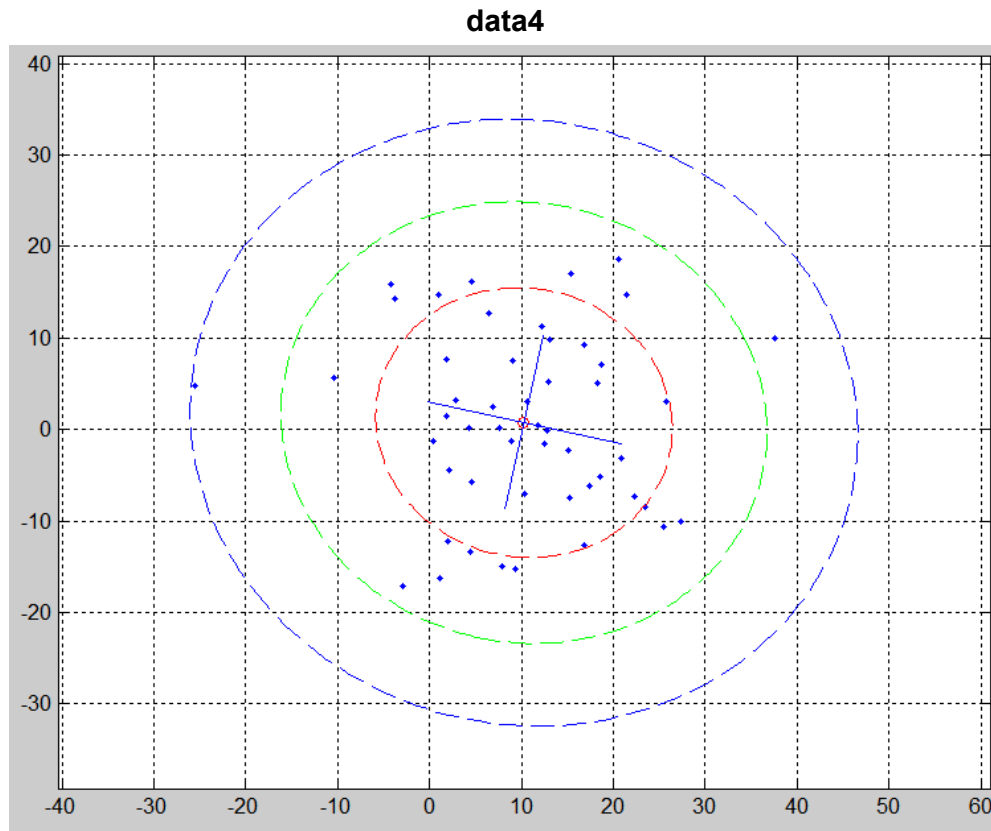


**data2**



**data3**





- b. Was geben die Eigenvektoren an? Wo sieht man das im Plot?
- c. Was geben die Eigenwerte an? Wo sieht man das im Plot? In welcher Relation stehen sie zur Gesamtvarianz?

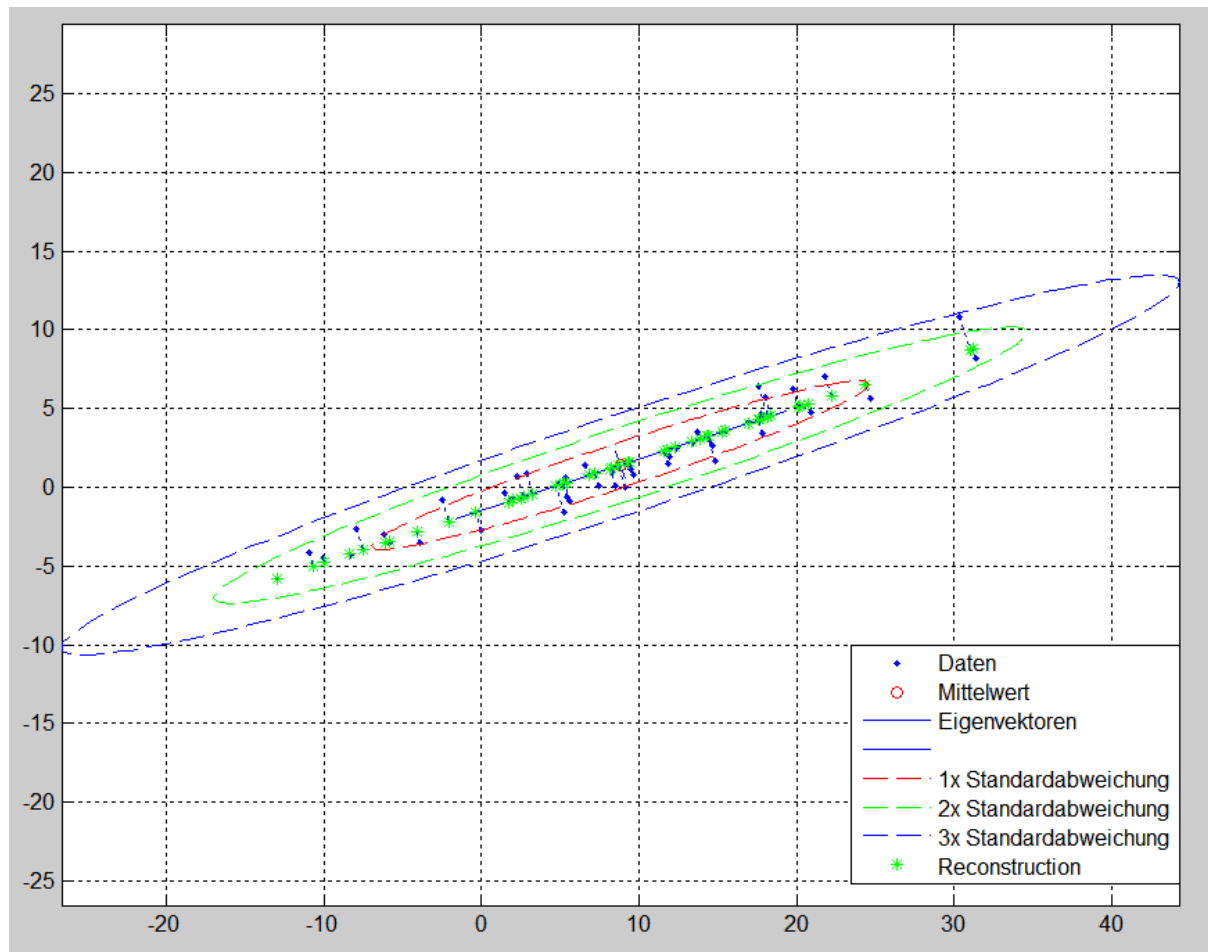
Die Eigenvektoren charakterisieren die Richtung des Streuverhaltens bzw. der Gesamtvarianz der Daten, und werden auf möglichst geringen orthogonalen Abstand zu den Messwerten angepasst. Zur Beschreibung der Gesamtvarianz fallen verschiedenen Eigenvektoren unterschiedlich signifikant aus. Wie signifikant nun ein Eigenvektor ist, wird anhand seines Eigenwertes bestimmt: Im Plot sieht man dies an den an die Streuungen „angepassten“ Eigenvektoren die nach Signifikanz (durch Multiplikation mit der Wurzel der Eigenwerte) skaliert wurden.

- d. Welchen Einfluss hat ein fehlender Mittelwertabzug (bei D) auf die Berechnung?

Ein fehlender Mittelwertabzug würde dazu führen, dass die Daten nicht zentriert sind.

### 3. Unterraum-Projektion

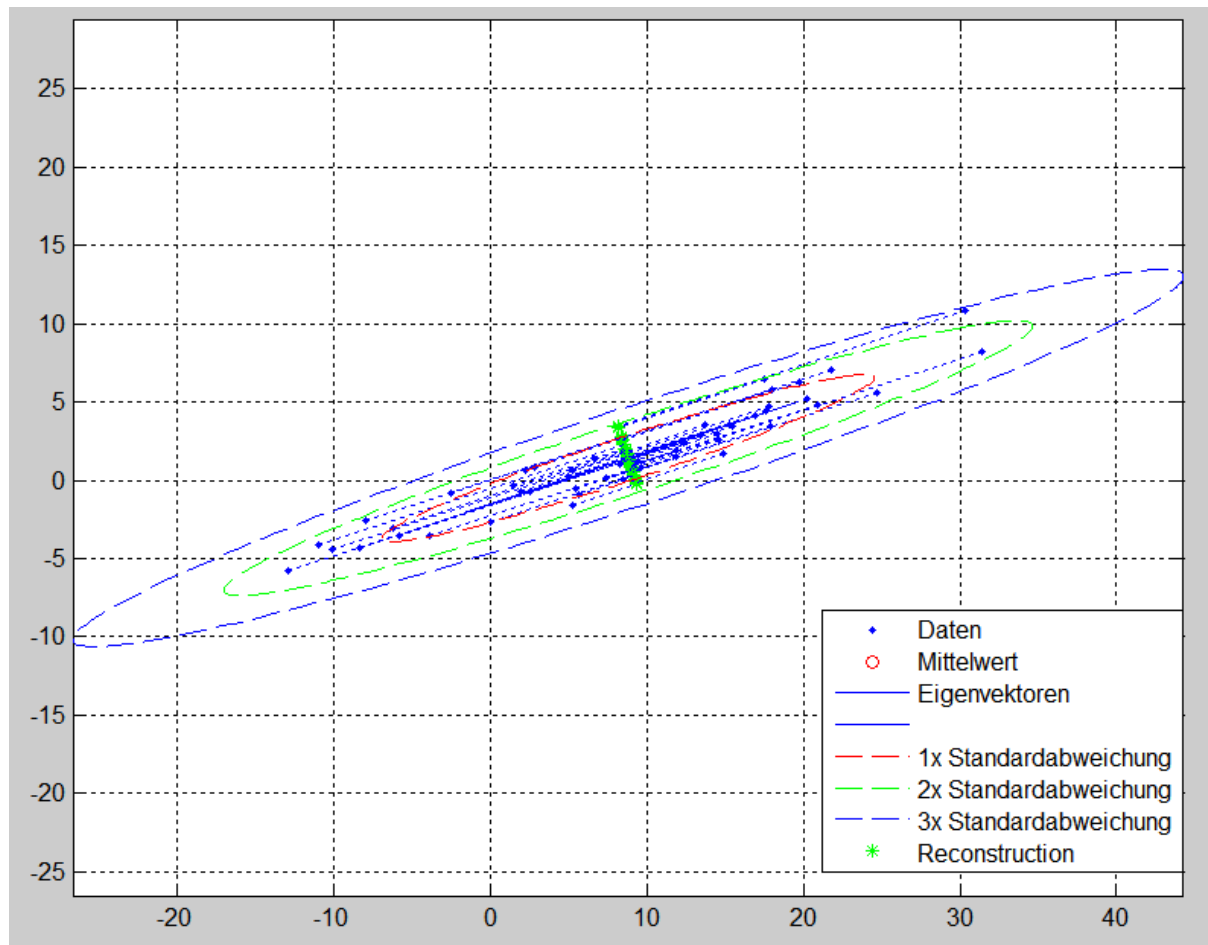
#### a. Projektion der Daten von data3 auf den Hauptvektor



Durch die Reduktion von 2 Eigenvektoren (Haupt- und Nebenvektor) auf nur den Hauptvektor geht bei der Projektion auf den Hauptvektor eine Dimension verloren. D.h. die Daten sind danach nur mehr 1-dimensional. Durch die Rekonstruktion erhält man wieder 2D-Daten, die nun jedoch allesamt auf dem Hauptvektor (Eigenvektor mit dem größten Eigenwert) liegen.

Der durchschnittliche Fehler der Daten zwischen Rekonstruktions- und Originaldaten beläuft sich auf 0.2270 für die x-Werte und 0.6893 für die y-Werte.

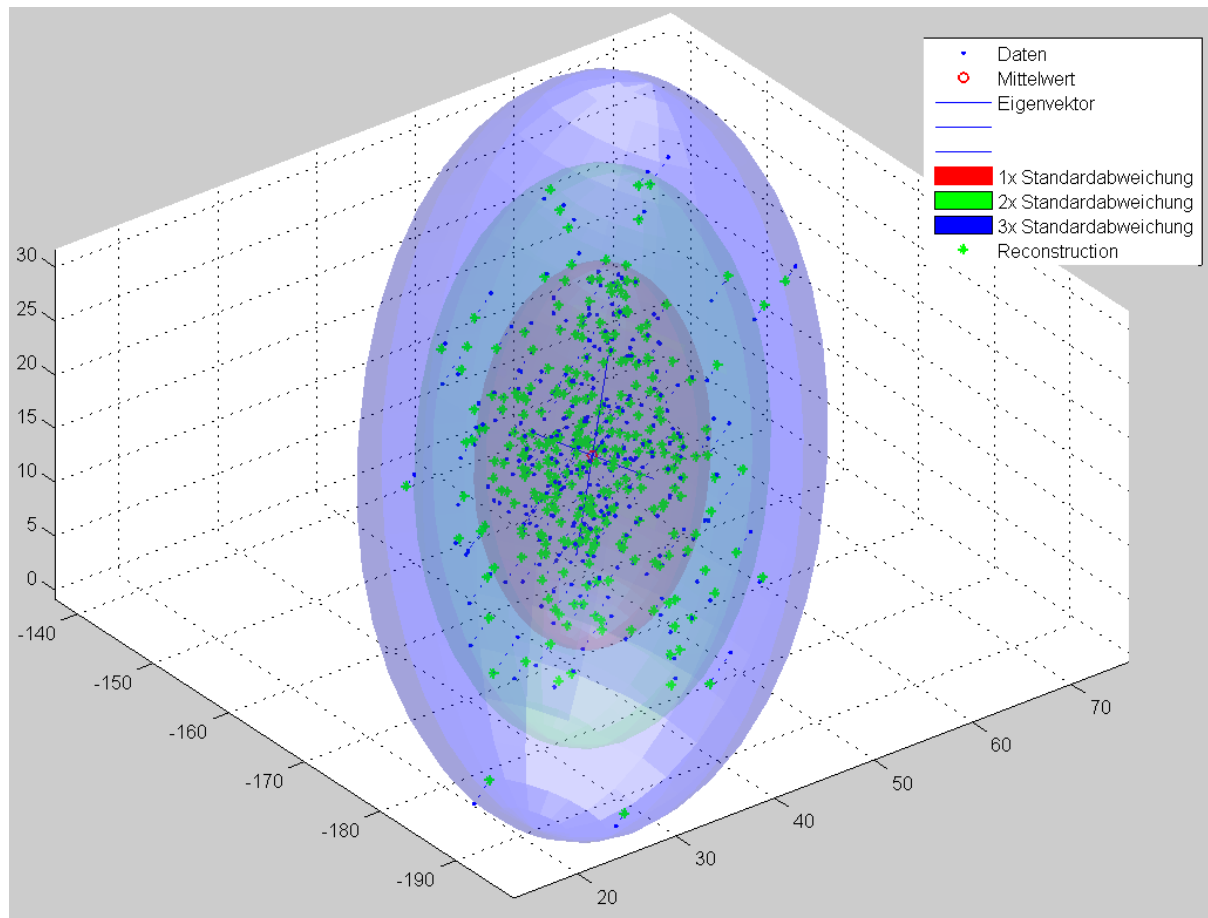
b. Projektion der Daten von data3 auf den Nebenvektor



Da der durchschnittliche Fehler bei einer Projektion auf den Nebenvektor um ein Vielfaches höher ist (in diesem Fall 8.4626 für x-Werte und 2.7868 für y-Werte, statt 0.2270 und 0.6893 beim Hauptvektor) empfiehlt es sich den Hauptvektor zur Darstellung der Datenmatrix zu wählen. Zur Erinnerung, der Hauptvektor ist im 2D-Fall jener Vektor mit dem höchsten Eigenwert.

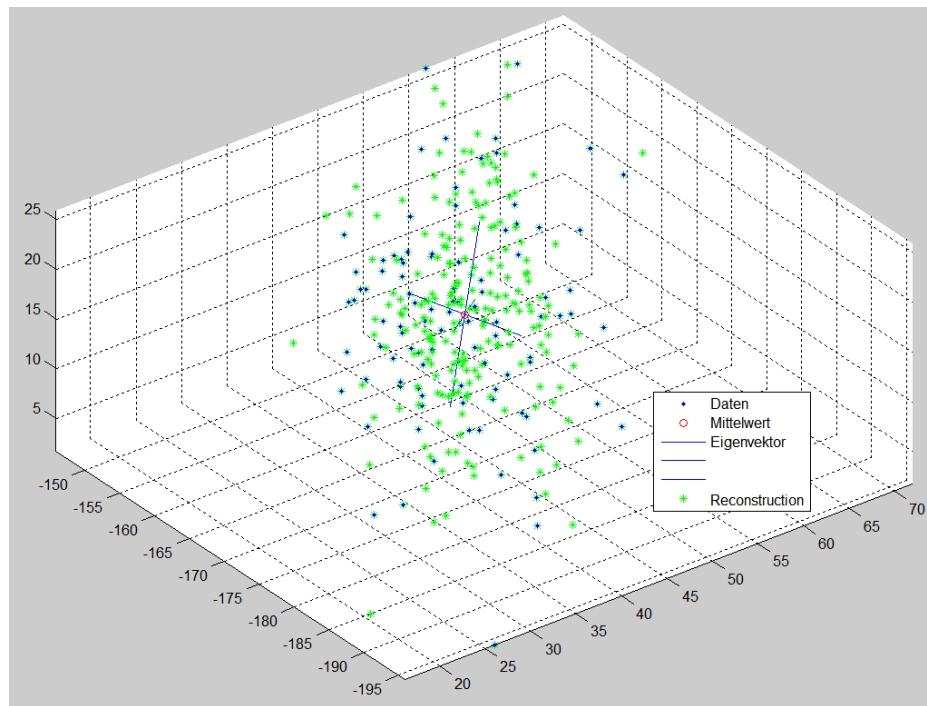
## 4. Untersuchungen in 3D

### a. Berechnung der PCA und Analyse der Daten

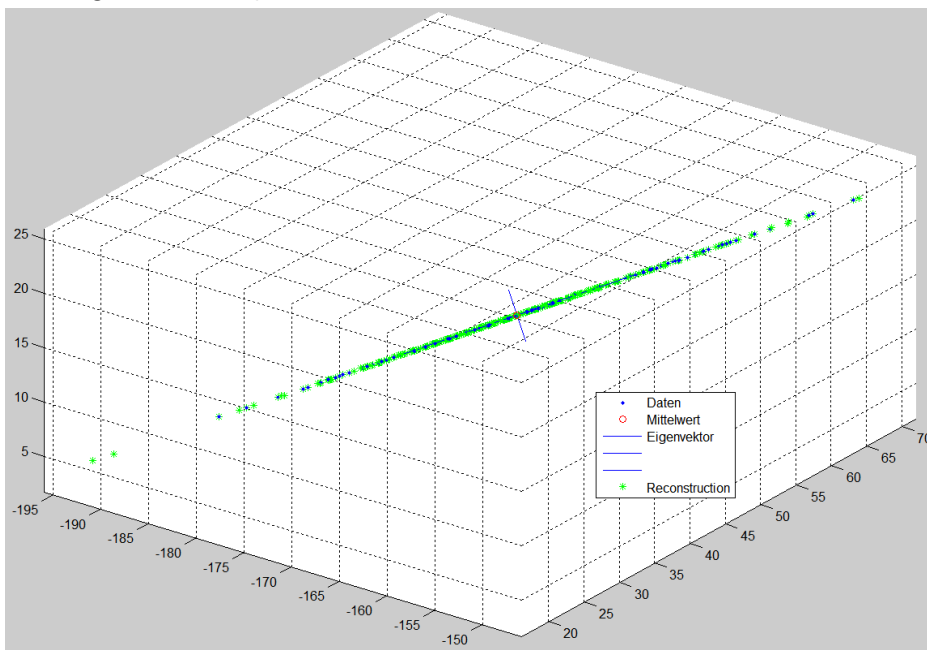


Die Varianzen in x-, y- und z-Richtung zeigen sich in der räumlichen Streuung der Datenpunkte im Plot. Je größer diese Varianzen in Richtung einer Achse (Eigenvektoren im Plot), umso größer sind die Eigenwerte der entsprechenden Achse und umso länger sind die Eigenvektoren, da diese mit dem entsprechenden Eigenwert skaliert werden. Je größer die Varianzen, umso größer sind auch die Standardabweichungen, die im Plot durch Ellipsoide dargestellt sind.

b. Unterraum-Projektion auf die ersten beiden Eigenvektoren



Deutlich erkennbare Ebene (Verlust einer räumlichen Dimension nach Projektion auf die ersten beiden Eigenvektoren)

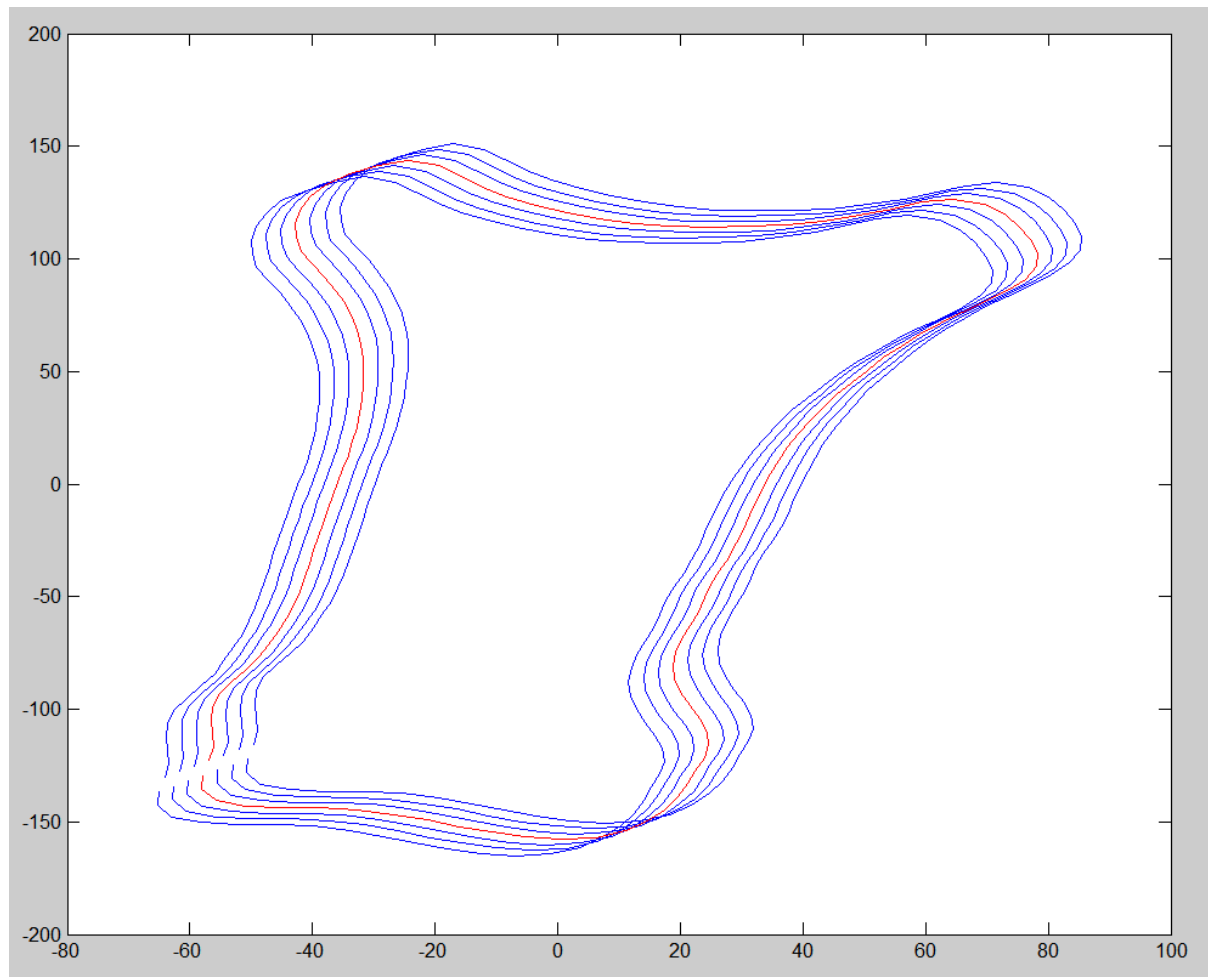


Nach der Projektion auf die ersten beiden Eigenvektoren erhalten wir 2D-Daten, d.h. die Information die bei der Rekonstruktion verloren geht ist eine räumliche Dimension der Daten. Der Informationsverlust hält sich jedoch in Grenzen, da nur die unbedeutendste Information (Eigenvektor mit dem geringsten Eigenwert) verloren geht.



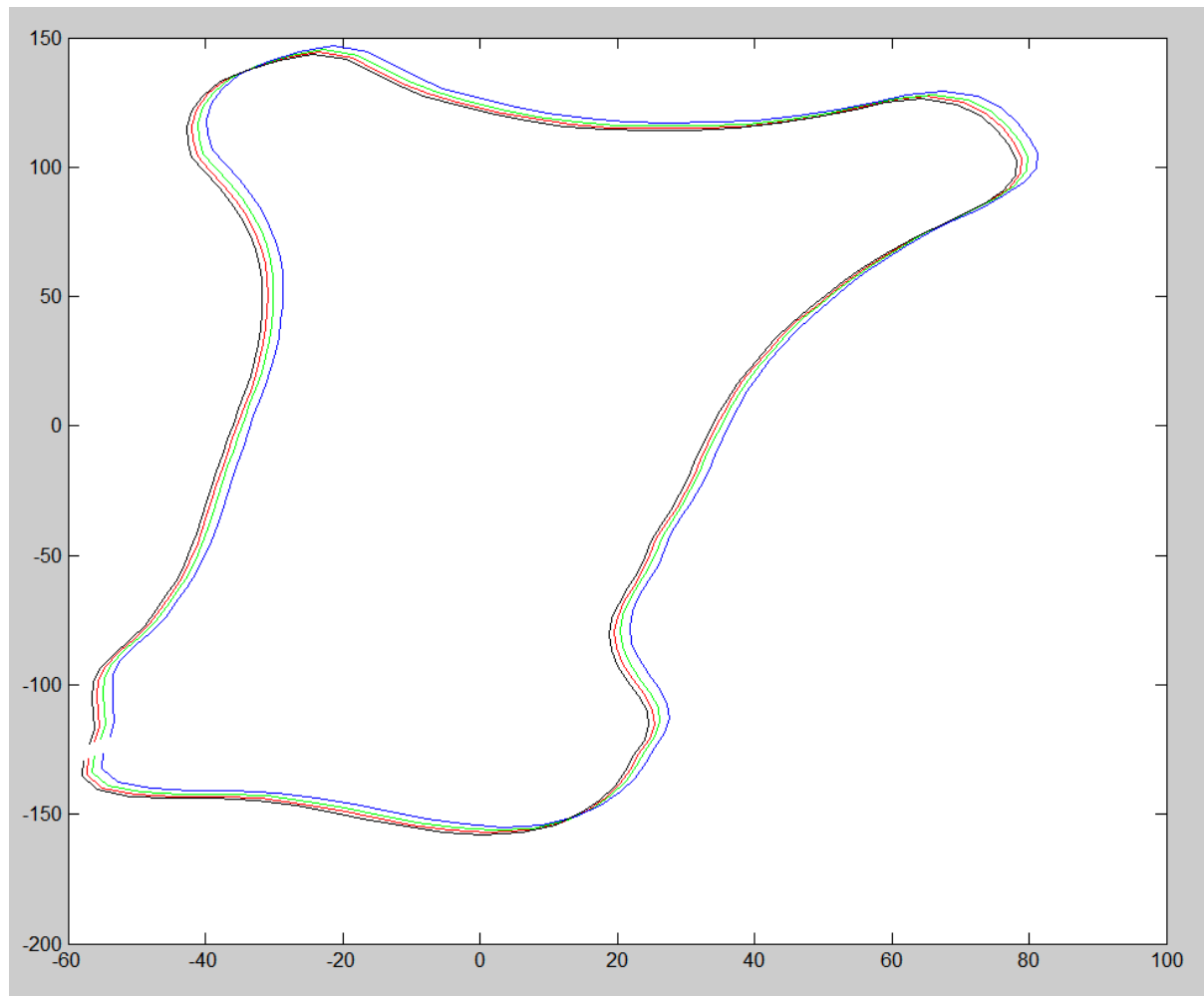
## 5. Shape Modell

### b) Unterschiedliche Modes interpretieren



Am Plot ist sehr leicht zu erkennen, dass die unterschiedlichen Shapes am manchen Stellen weiter auseinander liegen und an anderen Stellen sich sogar mit dem Mean-Shape überlappen. An den breiteren Stellen ist die Abweichung zum Mean-Shape größer, d.h. diese Datenpunkten besitzen eine größere Varianz. Die Datenpunkte an den engeren Stellen (nahezu-Überlappung mit Mean-Shape) haben daher eine kleinere Varianz. Aus dieser Beobachtung lässt sich direkt schließen, dass der Knochen über die Stellen, an denen die Shapes näher beieinander liegen, leichter zu identifizieren und klassifizieren ist als über jene Stellen, an denen die Shapes weit auseinander liegen.

c) Aus randomisiertem Parametervektor generierte Shapes interpretieren



(Shape mit 80% der Gesamtvarianz = GELB)

Shape mit 90% der Gesamtvarianz = ROT

Shape mit 95% der Gesamtvarianz = GRÜN

Shape mit 100% der Gesamtvarianz = BLAU

Mean-Shape = SCHWARZ

Die Summe aller Eigenwerte ergibt die Gesamtvarianz. Nun summieren wir so lange die Eigenwerte auf, bis wir 80, 90, 95 und 100% der Gesamtvarianz erreichen. Die 80 und 90% wurden jeweils durch die Summe der ersten drei Eigenwerte erreicht (nach zweitem Eigenwert knapp unter 80%, darüber knapp über 90%), deshalb ist das Shape mit 80% der Gesamtvarianz im Plot nicht vorhanden. Für 95% der Gesamtvarianz ist die Summe von 5 Eigenwerten notwendig und 100% erreichen wir nach 13 Eigenwerten.

Der Plot zeigt, dass je mehr Eigenwerte (mit entsprechender Anzahl an Parametern) für die Generierung der Shapes verwendet werden, umso wahrscheinlicher ist eine Konvergenz mit dem Mean-Shape.