

# Reduced-Rank-Regression

Daniel Herbst

10. Mai 2021

## Zusammenfassung

Nachdem in der vorigen Präsentation das multivariate Regressionsmodell mit festen Inputvariablen vorgestellt wurde, möchten wir uns nun dem Fall zufälliger Inputvariablen zuwenden. Zudem werden wir eine Verallgemeinerung des multivariaten Regressionsmodells, das *Reduced-Rank-Regressionmodell (RRR)*, betrachten, bei dem wir zusätzlich den Rang der Regressionskoeffizientenmatrix einschränken. Dieses Verfahren verallgemeinert unter anderem auch einige Techniken der multivariaten Statistik, etwa die Hauptkomponentenanalyse zur Dimensionsreduktion. Im Wesentlichen folgt der Vortrag [Iz08, Kapitel 6.3].

## 1 Einleitung

Unsere Ausgangssituation ist die Folgende:

$$\mathbf{X} := (X_1, \dots, X_r)^\top \quad \text{und} \quad \mathbf{Y} := (Y_1, \dots, Y_s)^\top$$

seien Zufallsvektoren mit gemeinsamer Verteilung  $\mathbb{P}^{(\mathbf{X}, \mathbf{Y})}$ , wobei  $r, s \in \mathbb{N}$  mit  $s \leq r$ , und ferner definieren wir folgende Schreibweisen für die entsprechenden Erwartungswerte und Kovarianzmatrizen (deren Existenz wir voraussetzen):

$$\boldsymbol{\mu}_{\mathbf{X}} := \mathbb{E}\mathbf{X}, \quad \boldsymbol{\mu}_{\mathbf{Y}} := \mathbb{E}\mathbf{Y} \quad \text{und} \quad \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} & \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}} \\ \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} & \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Y}} \end{pmatrix} := \Sigma\left(\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}\right).$$

Abgesehen von  $\mathbb{P}^{\mathbf{X}} \ll \lambda^r$  und  $\mathbb{P}^{\mathbf{Y}} \ll \lambda^s$ , d.h. der Stetigkeit der Zufallsvektoren  $\mathbf{X}$  und  $\mathbf{Y}$ , möchten wir zunächst keine weiteren Annahmen über die Verteilungen von  $\mathbf{X}$  und  $\mathbf{Y}$  treffen.

## 2 Klassisches multivariates Regressionsmodell mit zufälliger Inputvariable

Das klassische multivariate Regressionsmodell mit zufälliger Inputvariable ist von folgender Form: Für  $\mathbf{X}$  und  $\mathbf{Y}$  gelte

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Theta}\mathbf{X} + \mathcal{E}, \tag{2.1}$$

wobei  $\boldsymbol{\mu} \in \mathbb{R}^s$  und  $\boldsymbol{\Theta} \in \mathbb{R}^{s \times r}$  unbekannte Parameter seien sowie  $\mathcal{E}$  ein (nicht beobachtbarer)  $s$ -dimensionaler zufälliger Fehler ist mit Erwartungswert  $\mathbb{E}\mathcal{E} = 0$  und Kovarianzmatrix  $\boldsymbol{\Sigma}_{\mathcal{E}\mathcal{E}}$ . Zudem werden  $\mathbf{X}$  und  $\mathcal{E}$  als unkorreliert vorausgesetzt.

Wir möchten nun optimale  $\boldsymbol{\mu}$  und  $\boldsymbol{\Theta}$  finden in dem Sinne, dass die symmetrische, positiv semidefinite Matrix

$$W(\boldsymbol{\mu}, \boldsymbol{\Theta}) := \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})(\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})^\top] \in \mathbb{R}^{s \times s}$$

in gewisser Weise minimiert wird. Hierfür verwenden wir folgende Halbordnung, die man auf dem Raum der symmetrischen  $s \times s$ -Matrizen definieren kann:

**Definition 2.1** (Löwner-Halbordnung). Sei  $V := \{\mathbf{A} \in \mathbb{R}^{s \times s} \mid \mathbf{A}^\top = \mathbf{A}\}$  der Raum der symmetrischen  $s \times s$ -Matrizen. Setze für  $\mathbf{A}, \mathbf{B} \in V$ :

$$\mathbf{A} \leq_L \mathbf{B} :\Leftrightarrow \mathbf{B} - \mathbf{A} \text{ positiv semidefinit.}$$

Es lässt sich leicht nachprüfen, dass diese Relation eine Halbordnung auf  $V$ , die sogenannte Löwner-Halbordnung, definiert.

Mit dieser Halbordnung können wir nun den folgenden Satz formulieren:

**Satz 2.2.** Seien  $\mathbf{X}$  und  $\mathbf{Y}$  wie in der Einleitung, so ist

$$\underset{(\boldsymbol{\mu}, \boldsymbol{\Theta})}{\operatorname{argmin}} W(\boldsymbol{\mu}, \boldsymbol{\Theta}) = (\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\Theta}\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}) =: (\boldsymbol{\mu}_{\min}, \boldsymbol{\Theta}_{\min}),$$

wobei bezüglich  $\leq_L$  minimiert wird. Zudem ist

$$W(\boldsymbol{\mu}_{\min}, \boldsymbol{\Theta}_{\min}) = \boldsymbol{\Sigma}_{\mathbf{YY}} - \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\Sigma}_{\mathbf{XY}}.$$

Ferner gilt für  $\mathcal{E} := \mathbf{Y} - \boldsymbol{\mu}_{\min} - \boldsymbol{\Theta}_{\min}\mathbf{X}$ , dass

$$\mathbb{E}\mathcal{E} = \mathbf{0} \quad \text{und} \quad \mathbb{E}[\mathcal{E}(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})^\top] = \mathbf{0},$$

d.h.  $\mathbf{X}$  und  $\mathcal{E}$  sind unkorreliert.

**Bemerkungen.**

- Dass  $(\boldsymbol{\mu}_{\min}, \boldsymbol{\Theta}_{\min})$   $W(\boldsymbol{\mu}, \boldsymbol{\Theta})$  minimiert, soll hier natürlich **nicht** wie etwa in der Mengenlehre bedeuten, dass  $W(\boldsymbol{\mu}_{\min}, \boldsymbol{\Theta}_{\min})$  ein minimales Element der Menge  $\mathcal{S} := \{W(\boldsymbol{\mu}, \boldsymbol{\Theta}) \mid \boldsymbol{\mu} \in \mathbb{R}^s, \boldsymbol{\Theta} \in \mathbb{R}^{s \times r}\}$  ist, sondern die stärkere Bedingung, dass

$$W(\boldsymbol{\mu}_{\min}, \boldsymbol{\Theta}_{\min}) \leq_L W(\boldsymbol{\mu}, \boldsymbol{\Theta}) \quad \forall \boldsymbol{\mu} \in \mathbb{R}^s, \boldsymbol{\Theta} \in \mathbb{R}^{s \times r}.$$

Insbesondere kann es nur eine Minimalstelle geben.

- Die Stetigkeit von  $\mathbf{X}$  und  $\mathbf{Y}$  sichert, dass  $\boldsymbol{\Sigma}_{\mathbf{XX}}$  (und  $\boldsymbol{\Sigma}_{\mathbf{YY}}$ ) invertierbar sind: Denn wäre  $\boldsymbol{\Sigma}_{\mathbf{XX}}$  singulär, so gäbe es ein  $\mathbf{v} \in \mathbb{R}^s \setminus \{0\}$  mit

$$\mathbf{v}^\top \boldsymbol{\Sigma}_{\mathbf{XX}} \mathbf{v} = \mathbb{E}[\|\mathbf{v}^\top (\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})\|_2^2] = 0,$$

d.h.

$$\mathbb{P}(\mathbf{v}^\top (\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}) = 0) = 1$$

und die Verteilung von  $\mathbf{X}$  ist auf einer Hyperebene konzentriert (analog für  $\mathbf{Y}$ ).

- Mit dem Satz von Courant-Fischer [Iz08, Seite 52 f.] lässt sich zeigen, dass, falls  $\lambda_j(\mathbf{A})$  der  $j$ -größte Eigenwert von  $\mathbf{A} \in \mathbb{R}^{s \times s}$  ist, dann

$$\mathbf{A} \leq_L \mathbf{B} \implies \lambda_j(\mathbf{A}) \leq \lambda_j(\mathbf{B}), \quad j = 1, \dots, s.$$

Damit minimiert  $(\boldsymbol{\mu}_{\min}, \boldsymbol{\Theta}_{\min})$  auch

$$\text{tr}(W(\boldsymbol{\mu}, \boldsymbol{\Theta})) = \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})^\top (\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})], \quad (2.2)$$

was häufig als Minimierungsbedingung bei der multivariaten Regression genutzt wird. Die hier vorgestellte Variante ist also allgemeiner.

*Beweis.* Mit  $\mathbf{X}_c := \mathbf{X} - \boldsymbol{\mu}_\mathbf{X}$ ,  $\mathbf{Y}_c := \mathbf{Y} - \boldsymbol{\mu}_\mathbf{Y}$  gilt für alle  $\boldsymbol{\Theta} \in \mathbb{R}^{s \times r}$ ,  $\boldsymbol{\mu} \in \mathbb{R}^s$ :

$$\begin{aligned} W(\boldsymbol{\mu}, \boldsymbol{\Theta}) &= \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})(\mathbf{Y} - \boldsymbol{\mu} - \boldsymbol{\Theta}\mathbf{X})^\top] \\ &= \mathbb{E}[(\mathbf{Y}_c - \boldsymbol{\Theta}\mathbf{X}_c + (\boldsymbol{\mu}_\mathbf{Y} - \boldsymbol{\Theta}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu}))(\mathbf{Y}_c - \boldsymbol{\Theta}\mathbf{X}_c + (\boldsymbol{\mu}_\mathbf{Y} - \boldsymbol{\Theta}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu}))^\top] \\ &= \mathbb{E}[\mathbf{Y}_c \mathbf{Y}_c^\top - \mathbf{Y}_c \mathbf{X}_c^\top \boldsymbol{\Theta}^\top + \mathbf{Y}_c (\boldsymbol{\mu}_\mathbf{Y} - \boldsymbol{\Theta}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu})^\top \\ &\quad - \boldsymbol{\Theta}\mathbf{X}_c \mathbf{Y}_c^\top + \boldsymbol{\Theta}\mathbf{X}_c \mathbf{X}_c^\top \boldsymbol{\Theta}^\top - \boldsymbol{\Theta}\mathbf{X}_c (\boldsymbol{\mu}_\mathbf{Y} - \boldsymbol{\Theta}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu})^\top \\ &\quad + (\boldsymbol{\mu}_\mathbf{Y} - \boldsymbol{\Theta}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu}) \mathbf{Y}_c^\top - (\boldsymbol{\mu}_\mathbf{Y} - \boldsymbol{\Theta}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu}) \mathbf{X}_c^\top \boldsymbol{\Theta}^\top \\ &\quad + (\boldsymbol{\mu}_\mathbf{Y} - \boldsymbol{\Theta}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu})(\boldsymbol{\mu}_\mathbf{Y} - \boldsymbol{\Theta}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu})^\top] \\ &= \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Y}} - \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Theta}^\top - \boldsymbol{\Theta} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}} + \boldsymbol{\Theta} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} \boldsymbol{\Theta}^\top \\ &\quad + (\boldsymbol{\mu}_\mathbf{Y} - \boldsymbol{\Theta}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu})(\boldsymbol{\mu}_\mathbf{Y} - \boldsymbol{\Theta}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu})^\top \\ &= \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Y}} - \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}} \\ &\quad + \underbrace{(\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1/2} - \boldsymbol{\Theta} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{1/2})(\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1/2} - \boldsymbol{\Theta} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{1/2})^\top}_{=: A(\boldsymbol{\Theta}) \quad \text{positiv semidefinit}} \\ &\quad + \underbrace{(\boldsymbol{\mu}_\mathbf{Y} - \boldsymbol{\Theta}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu})(\boldsymbol{\mu}_\mathbf{Y} - \boldsymbol{\Theta}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu})^\top}_{=: B(\boldsymbol{\mu}, \boldsymbol{\Theta}) \quad \text{positiv semidefinit}} \\ &\geq \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Y}} - \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}}, \end{aligned}$$

mit Gleichheit falls  $A(\boldsymbol{\Theta}) = B(\boldsymbol{\mu}, \boldsymbol{\Theta}) = 0$ , was der Fall ist, wenn

$$\boldsymbol{\mu} = \boldsymbol{\mu}_\mathbf{Y} - \boldsymbol{\Theta}\boldsymbol{\mu}_\mathbf{X} =: \boldsymbol{\mu}_{\min} \quad \text{und} \quad \boldsymbol{\Theta} = \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} =: \boldsymbol{\Theta}_{\min}.$$

Man erhält nun

$$\begin{aligned} \mathcal{E} &= \mathbf{Y} - \boldsymbol{\mu}_{\min} - \boldsymbol{\Theta}_{\min} \mathbf{X} \\ &= \mathbf{Y} - (\boldsymbol{\mu}_\mathbf{Y} - (\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) \boldsymbol{\mu}_\mathbf{X}) - (\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{X} \\ &= \mathbf{Y}_c - \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \mathbf{X}_c \end{aligned}$$

und es gilt offensichtlich  $\mathbb{E}\mathcal{E} = \mathbf{0}$  sowie

$$\mathbb{E}[\mathcal{E} \mathbf{X}_c^\top] = \mathbb{E}[(\mathbf{Y}_c - \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \mathbf{X}_c) \mathbf{X}_c^\top] = \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} - \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} = \mathbf{0},$$

d.h.  $\mathbf{X}$  und  $\mathcal{E}$  sind unkorreliert. □

### 3 Reduced-Rank-Regressionsmodell

Wir betrachten nun das folgende Regressionsmodell

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{C}\mathbf{X} + \mathcal{E}, \quad (3.1)$$

bei dem  $\boldsymbol{\mu} \in \mathbb{R}^s$  und  $\mathbf{C} \in \mathbb{R}^{s \times r}$  unbekannte Parameter sind sowie  $\mathcal{E}$  ein nicht beobachtbarer, zufälliger Fehler mit  $\mathbb{E}\mathcal{E} = 0$  und Kovarianzmatrix  $\boldsymbol{\Sigma}_{\mathcal{E}\mathcal{E}}$ , wobei  $\mathbf{X}$  und  $\mathcal{E}$  unkorreliert seien.

Im Gegensatz zu (2.1), wo wir über  $\boldsymbol{\Theta}$  keine weiteren Annahmen treffen wollten, soll es nun zusätzlich ein gewisses  $t$  geben mit

$$\text{rk } \mathbf{C} \leq t \leq s.$$

In diesem Fall gibt es Matrizen  $\mathbf{A} \in \mathbb{R}^{s \times t}$  und  $\mathbf{B} \in \mathbb{R}^{t \times r}$  mit  $\mathbf{C} = \mathbf{A}\mathbf{B}$ , d.h. das Modell (3.1) lässt sich nun schreiben als

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{B}\mathbf{X} + \mathcal{E}, \quad (3.2)$$

wobei  $\boldsymbol{\mu}$ ,  $\mathbf{A}$  und  $\mathbf{B}$  nun die unbekannten Parameter sind.

#### Ein Kleinste-Quadrate-Kriterium

Auch hier möchten wir für  $\mathbf{X}$  und  $\mathbf{Y}$  wie in Satz 2.2  $\boldsymbol{\mu} \in \mathbb{R}^s$ ,  $\mathbf{A} \in \mathbb{R}^{s \times t}$  und  $\mathbf{B} \in \mathbb{R}^{t \times r}$  finden, die  $\mathbf{Y} - \boldsymbol{\mu} - \mathbf{A}\mathbf{B}\mathbf{X}$  auf eine gewisse Art und Weise minimieren, gehen nun aber etwas anders vor: Für eine symmetrische, positiv definite Gewichtsmatrix  $\boldsymbol{\Gamma} \in \mathbb{R}^{s \times s}$  wollen wir nun

$$W_{t,\boldsymbol{\Gamma}}(\boldsymbol{\mu}, \mathbf{A}, \mathbf{B}) := \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \mathbf{A}\mathbf{B}\mathbf{X})^\top \boldsymbol{\Gamma} (\mathbf{Y} - \boldsymbol{\mu} - \mathbf{A}\mathbf{B}\mathbf{X})]$$

über  $\boldsymbol{\mu}$ ,  $\mathbf{A}$  und  $\mathbf{B}$  minimieren.

Man beachte, dass es sich hierbei um eine Verallgemeinerung des Kriteriums (2.2) handelt, bei der noch die Gewichtsmatrix hinzugefügt wurde. Durch Variation dieser lassen sich mit der Reduced-Rank-Regression einige Techniken der multivariaten Statistik verallgemeinern, wie wir später sehen werden.

**Satz 3.1.** Seien  $\mathbf{X}$  und  $\mathbf{Y}$  wie in der Einleitung und  $1 \leq t \leq s$ . Dann ist

$$\underset{(\boldsymbol{\mu}, \mathbf{A}, \mathbf{B})}{\text{argmin}} W_{t,\boldsymbol{\Gamma}}(\boldsymbol{\mu}, \mathbf{A}, \mathbf{B}) = (\boldsymbol{\mu}_{\min}^{(t)}, \mathbf{A}_{\min}^{(t)}, \mathbf{B}_{\min}^{(t)})$$

mit

$$\begin{aligned} \mathbf{A}_{\min}^{(t)} &:= \boldsymbol{\Gamma}^{-1/2} \mathbf{U}_t && \in \mathbb{R}^{s \times t} \\ \mathbf{B}_{\min}^{(t)} &:= \mathbf{U}_t^\top \boldsymbol{\Gamma}^{1/2} \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} && \in \mathbb{R}^{t \times r} \\ \boldsymbol{\mu}_{\min}^{(t)} &:= \boldsymbol{\mu}_{\mathbf{Y}} - \mathbf{A}_{\min}^{(t)} \mathbf{B}_{\min}^{(t)} \boldsymbol{\mu}_{\mathbf{X}} && \in \mathbb{R}^s, \end{aligned}$$

wobei  $\mathbf{U}_t := (\mathbf{u}_1, \dots, \mathbf{u}_t) \in \mathbb{R}^{s \times t}$  und  $\mathbf{u}_1, \dots, \mathbf{u}_t$  orthonormale Eigenvektoren zu den Eigenwerten  $\lambda_1 \geq \dots \geq \lambda_t \geq 0$  von  $\boldsymbol{\Gamma}^{1/2} \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}} \boldsymbol{\Gamma}^{1/2}$  seien. Ferner gilt

$$W_{t,\boldsymbol{\Gamma}}(\boldsymbol{\mu}_{\min}^{(t)}, \mathbf{A}_{\min}^{(t)}, \mathbf{B}_{\min}^{(t)}) = \text{tr}(\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Y}}) \text{tr}(\boldsymbol{\Gamma}) - \sum_{i=1}^t \lambda_i.$$

Für den Beweis benötigen wir zunächst einen Hilfssatz aus der Linearen Algebra:

**Lemma 3.2** (Satz von Eckart-Young). Seien  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$  und  $b := \text{rk } \mathbf{B} \leq \text{rk } \mathbf{A} =: r$ , und sei  $\lambda_j(\mathbf{C})$  für eine reelle symmetrische Matrix  $\mathbf{C}$  deren  $j$ -größter Eigenwert. Dann gilt:

$$\lambda_j((\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})^\top) \geq \lambda_{j+b}(\mathbf{A}\mathbf{A}^\top)$$

mit Gleichheit für

$$\mathbf{A}_b := \sum_{i=1}^b \lambda_i^{1/2} \mathbf{u}_i \mathbf{v}_i^\top,$$

wobei  $\lambda_i := \lambda_i(\mathbf{A}\mathbf{A}^\top)$  und  $\mathbf{u}_i$  bzw.  $\mathbf{v}_i$  jeweils orthonormale Eigenvektoren von  $\mathbf{A}\mathbf{A}^\top$  bzw.  $\mathbf{A}^\top \mathbf{A}$  zu  $\lambda_i$  seien.

*Beweis.* Folgt (unter Benutzung der Singulärwertzerlegung von  $\mathbf{A}$ ) etwa direkt aus [BZ21, Satz 4.6].  $\square$

**Bemerkung.** Insbesondere zeigt Lemma 3.2, dass das dort definierte  $\mathbf{A}_b$  den Abstand von  $\mathbf{A}$  unter allen Rang- $b$ -Matrizen in  $\mathbb{R}^{m \times n}$  in der Spektralnorm minimiert.

*Beweis von Satz 3.1.* Seien  $\boldsymbol{\mu} \in \mathbb{R}^s$ ,  $\mathbf{A} \in \mathbb{R}^{s \times t}$ ,  $\mathbf{B} \in \mathbb{R}^{t \times r}$  und setze  $\mathbf{X}_c := \mathbf{X} - \boldsymbol{\mu}_\mathbf{X}$ ,  $\mathbf{Y}_c = \mathbf{Y} - \boldsymbol{\mu}_\mathbf{Y}$ ,  $\mathbf{C} := \mathbf{A}\mathbf{B}$ . Dann gilt:

$$\begin{aligned} W_{t,\Gamma}(\boldsymbol{\mu}, \mathbf{C}) &= \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu} - \mathbf{C}\mathbf{X})^\top \Gamma (\mathbf{Y} - \boldsymbol{\mu} - \mathbf{C}\mathbf{X})] \\ &= \mathbb{E}[(\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c + (\boldsymbol{\mu}_\mathbf{Y} - \mathbf{C}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu}))^\top \Gamma (\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c + (\boldsymbol{\mu}_\mathbf{Y} - \mathbf{C}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu}))] \\ &= \mathbb{E}[(\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c)^\top \Gamma (\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c)] \\ &\quad + \underbrace{(\boldsymbol{\mu}_\mathbf{Y} - \mathbf{C}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu})^\top \Gamma (\boldsymbol{\mu}_\mathbf{Y} - \mathbf{C}\boldsymbol{\mu}_\mathbf{X} - \boldsymbol{\mu})}_{\geq 0 \quad (\text{da } \Gamma \text{ symmetrisch positiv definit})} \end{aligned} \quad (3.3)$$

Unser Ziel ist es nun, (3.3) zu minimieren, indem wir zunächst den ersten Term nach  $\mathbf{C}$  minimieren. Für das so festgelegte  $\mathbf{C}_{\min}$  lässt sich  $\boldsymbol{\mu}_{\min} := \boldsymbol{\mu}_\mathbf{Y} - \mathbf{C}_{\min} \boldsymbol{\mu}_\mathbf{X}$  aber immer noch so wählen, dass der zweite Term zu 0 wird, was impliziert, dass die so bestimmte Kombination  $(\boldsymbol{\mu}_{\min}, \mathbf{C}_{\min})$  tatsächlich  $W_{t,\Gamma}(\boldsymbol{\mu}, \mathbf{C})$  minimiert.

Mit  $\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^* := \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}$ ,  $\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Y}}^* := \Gamma^{1/2} \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Y}} \Gamma^{1/2}$ ,  $\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}}^* := \Gamma^{1/2} \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}}$ ,  $\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}}^* := \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}} \Gamma^{1/2}$  und  $\mathbf{C}^* := \Gamma^{1/2} \mathbf{C}$  ist

$$\begin{aligned} \mathbb{E}[(\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c)^\top \Gamma (\mathbf{Y}_c - \mathbf{C}\mathbf{X}_c)] &= \mathbb{E}[\mathbf{Y}_c^\top \Gamma \mathbf{Y}_c - \mathbf{X}_c^\top \mathbf{C}^\top \Gamma \mathbf{Y}_c - \mathbf{Y}_c^\top \Gamma \mathbf{C} \mathbf{X}_c + \mathbf{X}_c^\top \mathbf{C}^\top \Gamma \mathbf{C} \mathbf{X}_c] \\ &= \text{tr}(\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Y}}^* - \mathbf{C}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}}^* - \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}}^* \mathbf{C}^{*\top} + \mathbf{C}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^* \mathbf{C}^{*\top}) \\ &= \text{tr}(\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Y}}^* - \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*-1} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}}^*) \\ &\quad + \text{tr}((\mathbf{C}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*1/2} - \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*-1/2}) \\ &\quad \cdot (\mathbf{C}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*1/2} - \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*-1/2})^\top). \end{aligned} \quad (3.4)$$

Sei nun

$$k := \text{rk } \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} = \text{rk } \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*-1/2}$$

und seien  $\mathbf{u}_i$  orthonormale Eigenvektoren zum jeweils  $i$ -größten Eigenwert  $\lambda_i$  von

$$(\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*-1/2})(\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*-1/2})^\top = \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*-1} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}}^* = \Gamma^{1/2} \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}} \Gamma^{1/2}$$

für  $1 \leq i \leq s$  sowie

$$\mathbf{v}_i := \lambda_i^{-1/2} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*-1/2} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}}^* \mathbf{u}_i = \lambda_i^{-1/2} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}} \boldsymbol{\Gamma}^{1/2} \mathbf{u}_i, \quad 1 \leq i \leq k$$

zugehörige orthonormale Eigenvektoren von  $(\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*-1/2})^\top (\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*-1/2})$ , wobei wir für  $i > k$  die restlichen  $\mathbf{v}_i$  so definieren, dass wir eine Orthonormalbasis von  $\mathbb{R}^r$  erhalten. Dann lässt sich der Satz von Eckart-Young (Lemma 3.2) wie folgt anwenden: Der zweite Term von (3.4) wird minimal für

$$\mathbf{C}^* \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{*1/2} = \boldsymbol{\Gamma}^{1/2} \mathbf{C} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{1/2} = \sum_{i=1}^{\min\{k,t\}} \lambda_i^{1/2} \mathbf{u}_i \mathbf{v}_i^\top$$

(denn die Spur ist als Summe der Eigenwerte monoton steigend in jedem Eigenwert), also folgt, dass

$$\begin{aligned} \mathbf{C}_{\min}^{(t)} &:= \boldsymbol{\Gamma}^{-1/2} \left( \sum_{i=1}^{\min\{k,t\}} \lambda_i^{1/2} \mathbf{u}_i \mathbf{v}_i^\top \right) \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1/2} \\ &= \boldsymbol{\Gamma}^{-1/2} \left( \sum_{i=1}^{\min\{k,t\}} \mathbf{u}_i \mathbf{u}_i^\top \right) \boldsymbol{\Gamma}^{1/2} \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \\ &= \boldsymbol{\Gamma}^{-1/2} \left( \sum_{i=1}^t \mathbf{u}_i \mathbf{u}_i^\top \right) \boldsymbol{\Gamma}^{1/2} \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \end{aligned}$$

den ersten Summand von (3.3) über alle Matrizen  $\mathbf{C} \in \mathbb{R}^{s \times r}$  von Rang  $\leq t$  minimiert. Daher wird  $W_{t,\boldsymbol{\Gamma}}(\boldsymbol{\mu}, \mathbf{A}, \mathbf{B})$  nun minimal für

$$\begin{aligned} \mathbf{A}_{\min}^{(t)} &= \boldsymbol{\Gamma}^{-1/2} \mathbf{U}_t && \in \mathbb{R}^{s \times t} \\ \mathbf{B}_{\min}^{(t)} &= \mathbf{U}_t^\top \boldsymbol{\Gamma}^{1/2} \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} && \in \mathbb{R}^{t \times r} \\ \boldsymbol{\mu}_{\min}^{(t)} &= \boldsymbol{\mu}_{\mathbf{Y}} - \mathbf{A}_{\min}^{(t)} \mathbf{B}_{\min}^{(t)} \boldsymbol{\mu}_{\mathbf{X}} && \in \mathbb{R}^s, \end{aligned}$$

wobei  $\mathbf{U}_t := (\mathbf{u}_1, \dots, \mathbf{u}_t) \in \mathbb{R}^{s \times t}$ . Mit (3.3), (3.4) ist das dadurch angenommene Minimum dann

$$\begin{aligned} W_{t,\boldsymbol{\Gamma}}(\boldsymbol{\mu}_{\min}^{(t)}, \mathbf{A}_{\min}^{(t)}, \mathbf{B}_{\min}^{(t)}) &= \text{tr}(\boldsymbol{\Gamma}^{1/2} \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Y}} \boldsymbol{\Gamma}^{1/2} - \boldsymbol{\Gamma}^{1/2} \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}} \boldsymbol{\Gamma}^{1/2}) \\ &\quad + \text{tr} \left( \left( \sum_{i=t+1}^k \lambda_i^{1/2} \mathbf{u}_i \mathbf{v}_i^\top \right) \left( \sum_{i=t+1}^k \lambda_i^{1/2} \mathbf{u}_i \mathbf{v}_i^\top \right)^\top \right) \\ &= \text{tr}(\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Y}}) \text{tr}(\boldsymbol{\Gamma}) - \text{tr}(\boldsymbol{\Gamma}^{1/2} \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}} \boldsymbol{\Gamma}^{1/2}) + \sum_{i=t+1}^k \lambda_i \\ &= \text{tr}(\boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Y}}) \text{tr}(\boldsymbol{\Gamma}) - \sum_{i=1}^t \lambda_i. \quad \square \end{aligned}$$

**Bemerkung.** Für  $t = s$  ist  $\sum_{i=1}^s \mathbf{u}_i \mathbf{u}_i^\top = \mathbf{I}_s \in \mathbb{R}^{s \times s}$ , also  $\mathbf{C}_{\min}^{(t)} = \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} = \boldsymbol{\Theta}$  und damit liefern in diesem Fall die Vorgehensweisen aus Satz 2.2 und Satz 3.1 dasselbe Ergebnis - unabhängig von  $\boldsymbol{\Gamma}$ .

## Spezialfälle

Für spezielle Wahlen von  $\mathbf{\Gamma}$  und  $\mathbf{X}, \mathbf{Y}$  können wir nun wie angekündigt mit der Reduced-Rank-Regression einige klassische Methoden der multivariaten Statistik zusammenfassen, die wir teilweise auch noch in diesem Seminar kennenlernen werden:

- Im Fall  $\mathbf{X} = \mathbf{Y}$  und  $\mathbf{\Gamma} = \mathbf{I}_s$  erhalten wir die *Hauptkomponentenanalyse (principal component analysis, PCA)* (Vortrag 6).
- Für  $\mathbf{\Gamma} = \mathbf{\Sigma}_{\mathbf{Y}\mathbf{Y}}^{-1}$  erhalten wir die *kanonische Korrelationsanalyse* (Vortrag 7).
- Nutzen wir das Setup der kanonischen Korrelationsanalyse für einen  $\{0, 1\}$ -wertigen Vektor  $\mathbf{Y}$ , der Gruppenzugehörigkeiten modelliert, erhalten wir die *lineare Diskriminanzanalyse* (Vortrag 8).

## Stichprobenschätzung bei der Reduced-Rank-Regression

Um Satz 3.1 zu verwenden, werden  $\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\mu}_{\mathbf{Y}}$  und  $\mathbf{\Sigma}_{\mathbf{X}\mathbf{X}}, \mathbf{\Sigma}_{\mathbf{X}\mathbf{Y}}, \mathbf{\Sigma}_{\mathbf{Y}\mathbf{X}}, \mathbf{\Sigma}_{\mathbf{Y}\mathbf{Y}}$  benötigt, die in der Regel natürlich unbekannt sind. Stattdessen liegt häufig eine Stichprobe von  $n \in \mathbb{N}$  Beobachtungen

$$\mathcal{D} := \{(\mathbf{x}_j, \mathbf{y}_j) \mid 1 \leq j \leq n\} \quad (3.5)$$

vor, die wir als unabhängige Realisierungen von  $r$ - bzw.  $s$ -dimensionalen Zufallsvektoren  $\mathbf{X}$  bzw.  $\mathbf{Y}$  (wie in der Einleitung) auffassen.

Damit schätzen wir  $\boldsymbol{\mu}_{\mathbf{X}}$  und  $\boldsymbol{\mu}_{\mathbf{Y}}$  auf naheliegende Weise:

$$\hat{\boldsymbol{\mu}}_{\mathbf{X}} := \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j = \bar{\mathbf{x}}_n, \quad \hat{\boldsymbol{\mu}}_{\mathbf{Y}} := \frac{1}{n} \sum_{j=1}^n \mathbf{y}_j = \bar{\mathbf{y}}_n \quad (3.6)$$

Für die Kovarianzmatrizen definieren wir zunächst

$$\mathbf{x}_{cj} := \mathbf{x}_j - \bar{\mathbf{x}}_n, \quad \mathbf{y}_{cj} := \mathbf{y}_j - \bar{\mathbf{y}}_n, \quad \mathcal{X}_c := (\mathbf{x}_{c1}, \dots, \mathbf{x}_{cn}) \in \mathbb{R}^{r \times n}, \quad \mathcal{Y}_c := (\mathbf{y}_{c1}, \dots, \mathbf{y}_{cn}) \in \mathbb{R}^{s \times n}$$

und schätzen diese dann wie folgt:

$$\begin{aligned} \hat{\mathbf{\Sigma}}_{\mathbf{X}\mathbf{X}} &:= \frac{1}{n} \mathcal{X}_c \mathcal{X}_c^\top \in \mathbb{R}^{r \times r}, \quad \hat{\mathbf{\Sigma}}_{\mathbf{Y}\mathbf{Y}} := \frac{1}{n} \mathcal{Y}_c \mathcal{Y}_c^\top \in \mathbb{R}^{s \times s}, \\ \hat{\mathbf{\Sigma}}_{\mathbf{Y}\mathbf{X}} &:= \frac{1}{n} \mathcal{Y}_c \mathcal{X}_c^\top = \hat{\mathbf{\Sigma}}_{\mathbf{X}\mathbf{Y}}^\top \in \mathbb{R}^{s \times r}. \end{aligned} \quad (3.7)$$

Die entsprechenden Größen aus Satz 3.1 können wir nun einfach schätzen, indem wir die Schätzungen für die Erwartungswerte und Kovarianzmatrizen aus (3.6), (3.7) verwenden:

$$\hat{\mathbf{C}}_{\min}^{(t)} := \mathbf{\Gamma}^{-1/2} \left( \sum_{i=1}^t \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^\top \right) \mathbf{\Gamma}^{1/2} \hat{\mathbf{\Sigma}}_{\mathbf{Y}\mathbf{X}} \hat{\mathbf{\Sigma}}_{\mathbf{X}\mathbf{X}}^{-1},$$

wobei  $\hat{\mathbf{u}}_i$ ,  $1 \leq i \leq t$  orthonormale Eigenvektoren zu den je  $i$ -größten Eigenwerten  $\lambda_i$ ,  $1 \leq i \leq t$  von  $\mathbf{\Gamma}^{1/2} \hat{\mathbf{\Sigma}}_{\mathbf{Y}\mathbf{X}} \hat{\mathbf{\Sigma}}_{\mathbf{X}\mathbf{X}}^{-1} \hat{\mathbf{\Sigma}}_{\mathbf{X}\mathbf{Y}} \mathbf{\Gamma}^{1/2}$  seien.  $\boldsymbol{\Theta}_{\min}$  aus Satz 2.2 wird analog durch

$$\hat{\boldsymbol{\Theta}} := \hat{\mathbf{\Sigma}}_{\mathbf{Y}\mathbf{X}} \hat{\mathbf{\Sigma}}_{\mathbf{X}\mathbf{X}}^{-1}$$

geschätzt.

$\widehat{\Sigma}_{\mathbf{X}\mathbf{X}}$  und  $\widehat{\Sigma}_{\mathbf{Y}\mathbf{Y}}$  sind nach [EP73] für  $n \geq r + 1$  mit Wahrscheinlichkeit 1 invertierbar (hier wird natürlich auch benötigt, dass  $\mathbf{X}$  und  $\mathbf{Y}$  stetig sind), allerdings kann  $\kappa(\widehat{\Sigma}_{\mathbf{X}\mathbf{X}})$  bzw.  $\kappa(\widehat{\Sigma}_{\mathbf{Y}\mathbf{Y}})$  in der Praxis oft so groß werden, dass Invertieren schwierig ist. In diesem Fall könnte man stattdessen die Pseudoinversen von  $\widehat{\Sigma}_{\mathbf{X}\mathbf{X}}$ ,  $\widehat{\Sigma}_{\mathbf{Y}\mathbf{Y}}$  verwenden oder, wie [Iz08, Seite 182] nahelegt, durch eine kleine Veränderung der Diagonaleinträge erreichen, dass man eine besser invertierbare Matrix erhält:

$$\widehat{\Sigma}_{\mathbf{X}\mathbf{X}}^\delta := \frac{1}{n}(\mathcal{X}_c \mathcal{X}_c^\top + \delta \mathbf{I}_r), \quad \widehat{\Sigma}_{\mathbf{Y}\mathbf{Y}}^\delta := \frac{1}{n}(\mathcal{Y}_c \mathcal{Y}_c^\top + \delta \mathbf{I}_s)$$

für ein  $\delta > 0$ . In [Iz08, Kapitel 6.3.4] wird etwa erklärt, wie man dieses  $\delta$  in der Praxis sinnvoll wählen könnte.

## Ein Simulationsbeispiel

Wir wenden die Reduced-Rank-Regression nun an einem simulierten Beispiel an. Dafür erzeugen wir je  $n = 1000$  unabhängige Realisierungen

$$\mathbf{x}_j, \quad \varepsilon_j, \quad 1 \leq j \leq n,$$

wobei die  $\mathbf{x}_j$  einer  $\mathcal{N}_3(\mathbf{0}, 0.3 \cdot \mathbf{I}_3)$ -Verteilung und die  $\varepsilon_j$  einer  $\mathcal{N}_3(\mathbf{0}, 0.15 \cdot \mathbf{I}_3)$ -Verteilung entstammen, und setzen  $r = s := 3$  sowie

$$\mathbf{C} := \frac{1}{4} \cdot \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \in \mathbb{R}^{s \times r}, \quad \boldsymbol{\mu} := \begin{pmatrix} 0.5 \\ 0.5 \\ 0.5 \end{pmatrix}.$$

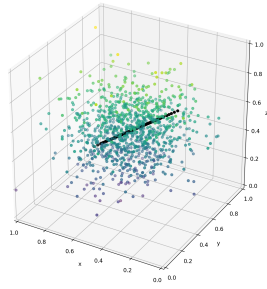
Für  $\mathbf{y}_j := \boldsymbol{\mu} + \mathbf{C}\mathbf{x}_j + \varepsilon_j$  ist unsere Stichprobe für die Reduced-Rank-Regression dann wie in (3.5)

$$\mathcal{D} := \{(\mathbf{x}_j, \mathbf{y}_j) \mid 1 \leq j \leq n\}.$$

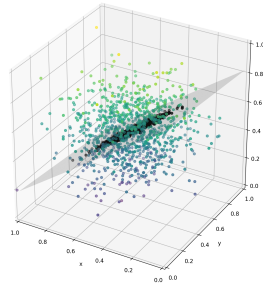
und wir können, wie im vorigen Abschnitt beschrieben, die Reduced-Rank-Regression durchführen, wobei wir der Einfachheit halber  $\boldsymbol{\Gamma} := \mathbf{I}_s$  verwenden.

In Abbildung 1, (a)-(c) sieht man Visualisierungen dieser Situation: Die bunten Punkte sind die  $\mathbf{y}_j$ , die schwarzen Punkte sind die Prädiktoren der Reduced-Rank-Regression für die jeweiligen Parameter  $t = 1, 2, 3$ . Im Vergleich dazu sieht man in Abbildung 1, (d)-(f) das analoge Ergebnis, wenn man stattdessen  $\mathcal{D} := \{(\mathbf{y}_j, \mathbf{y}_j) \mid 1 \leq j \leq n\}$  als Stichprobe verwendet - d.h. hier wird im Wesentlichen eine Hauptkomponentenanalyse von  $\{\mathbf{y}_j \mid 1 \leq j \leq n\}$  durchgeführt.

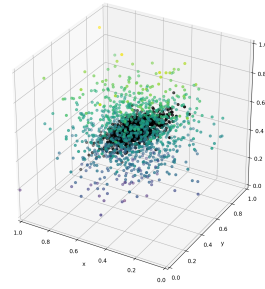




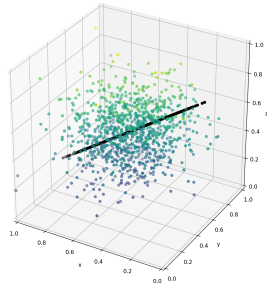
(a)  $\mathbf{x}, \mathbf{y}; t = 1$



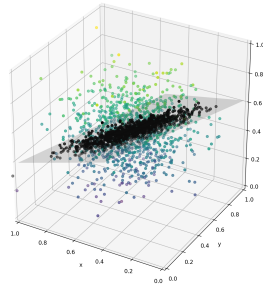
(b)  $\mathbf{x}, \mathbf{y}; t = 2$



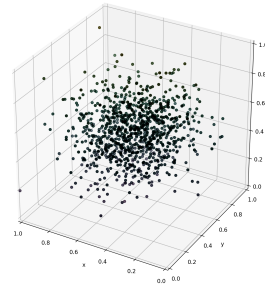
(c)  $\mathbf{x}, \mathbf{y}; t = 3$



(d)  $\mathbf{y}, \mathbf{y}; t = 1$



(e)  $\mathbf{y}, \mathbf{y}; t = 2$



(f)  $\mathbf{y}, \mathbf{y}; t = 3$

Abbildung 1: Plots der Simulationen

## Effektive Dimension der Regression

Letztlich bleibt noch die Frage zu beantworten, wie man die sogenannte *effektive Dimension*  $t$  bei der Reduced-Rank-Regression geeignet wählen kann. Dafür beachte man, dass man in der Situation von Satz 3.1, wenn der Rang der Koeffizientenmatrix  $t = t_0$  zu  $t = t_1$ ,  $t_0 < t_1$  vergrößert wird, eine Verringerung von  $W_{t,\Gamma}(\boldsymbol{\mu}_{\min}^{(t)}, \mathbf{A}_{\min}^{(t)}, \mathbf{B}_{\min}^{(t)})$  der Höhe

$$W_{t_0,\Gamma}(\boldsymbol{\mu}_{\min}^{(t_0)}, \mathbf{A}_{\min}^{(t_0)}, \mathbf{B}_{\min}^{(t_0)}) - W_{t_1,\Gamma}(\boldsymbol{\mu}_{\min}^{(t_1)}, \mathbf{A}_{\min}^{(t_1)}, \mathbf{B}_{\min}^{(t_1)}) = \sum_{i=t_0+1}^{t_1} \lambda_i$$

bewirkt. Hiermit könnte man in der Praxis (wo man dann natürlich mit  $\hat{\lambda}_i$  arbeitet) untersuchen, bis wann es sich „lohnt“,  $t$  zu vergrößern, was sich, wie in [Iz08, Seite 185] beschrieben wird, mit Tests, ob  $\text{rk } \Gamma^{1/2} \boldsymbol{\Sigma}_{\mathbf{YX}} \boldsymbol{\Sigma}_{\mathbf{XX}}^{-1} \boldsymbol{\Sigma}_{\mathbf{XY}} \Gamma^{1/2} = \text{rk } \boldsymbol{\Sigma}_{\mathbf{YX}}$  kleiner als eine festgelegte Zahl ist, überprüfen ließe. Ferner wird in [Iz08, Kapitel 6.3.4] ein anderer Algorithmus vorgestellt, mit dem die effektive Dimension  $t$  bestimmt werden kann.

## Literatur

- [Iz08] Izenman, Alan Julian. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. Springer Texts in Statistics, Springer-Verlag New York, 2008.
- [BZ21] Bandeira, Afonso S. und Zhivotovskiy, Nikita. *Lecture Notes for Mathematics of Machine Learning*. 2021. [https://metaphor.ethz.ch/x/2021/fs/401-2684-00L/sc/Math\\_of\\_ML\\_Lecture\\_Notes.pdf](https://metaphor.ethz.ch/x/2021/fs/401-2684-00L/sc/Math_of_ML_Lecture_Notes.pdf)
- [EP73] Eaton, Morris L. und Perlman, Michael D. *The Non-Singularity of Generalized Sample Covariance Matrices*. The Annals of Statistics, 1973.