

# Cleaning von Zeitreihen

Von der Anomalieerkennung zur Anomalienreparatur

---

Jose Rodriguez Parra Flores  
Klaus-Johan Ziegert

16. September 2019



# Gliederung

---

- 1 Einführung
- 2 Grundlagen
- 3 Iterative Minimum Repairing
- 4 Evaluierung
- 5 Schluss

# Einführung

---

- 1 Einführung
  - Motivation
  - Zielsetzung

2 Grundlagen

3 Iterative Minimum Repairing

4 Evaluierung

5 Schluss

# Motivation

## Messgeräte liefern unzuverlässige Daten

- GPS Tracker sind nahe von Gebäuden unzuverlässig
- Sensoren sind empfindlich gegenüber äußere Einflüsse
  - Z.B. starker Fall der Temperaturen bei einem Windzug



Abbildung: GPS-Tracking auf dem Campus der Tsinghua Universität [1]

# Motivation

## Umgang von unzuverlässigen Daten mit Anomalieerkennung

### ① Unzuverlässige Datenpunkte entfernen

- Ausreißer werden entfernt 😊
- Entfernen aufeinanderfolgende Fehler machen Ergebnis unbrauchbar **oder** werden als solche ggf. nicht entfernt ☹️

### ② Unzuverlässige Datenpunkte reparieren

- Einzelne Ausreißer werden leicht korrigiert 😐
- Aufeinanderfolgende Fehler werden zu stark verändert (In der Praxis liegen die Messungen nahe bei den korrekten Werten) 😐

# Motivation

## Hinzunahme von korrekt markierten Werten

- ① Markierung durch den Benutzer
  - Z.B. markiert der Benutzer in beliebigen Zeitabständen seinen aktuellen Standort
- ② Präzise Messgeräte liefern in längeren Zeitabstände korrekte Werte



# Zielsetzung

## Ziel der Arbeit

- ① Berücksichtigung der markierten Werte in der Anomalieerkennung
  - Aufeinanderfolgende Fehler sollen besser abgeschätzt werden
- ② Anomalienreparatur mit den Minimum-Change-Prinzip vereinbaren
  - Keine drastische Veränderungen der Messwerte
- ③ Neue Anomalienreparatur hinsichtlich Berechnungslaufzeit, Ergebnisgenauigkeit usw. optimieren
- ④ Neue Anomalienreparatur mit unterschiedlichen Einstellungen mit den anderen Verfahren empirisch vergleichen

# Grundlagen

## 1 Einführung

## 2 Grundlagen

- Problemstellung
- Anomalien
- Anomalienreparaturen
- Reparatur durch Anomalieerkennung

## 3 Iterative Minimum Repairing

## 4 Evaluierung

## 5 Schluss

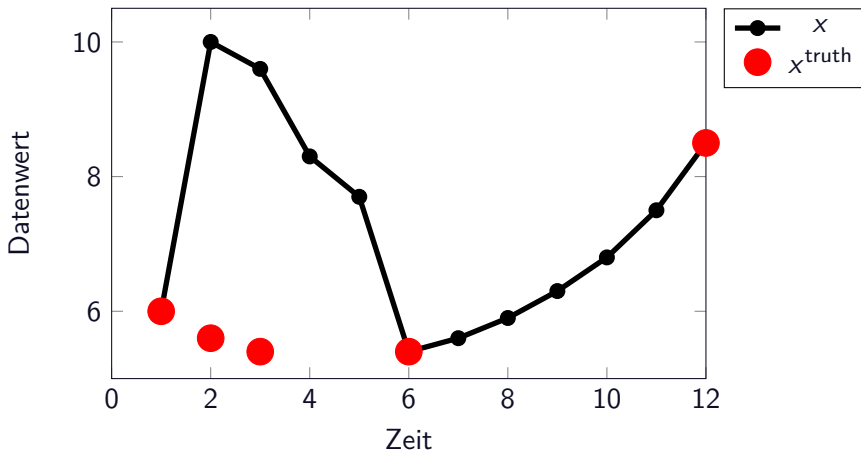


# Problemstellung

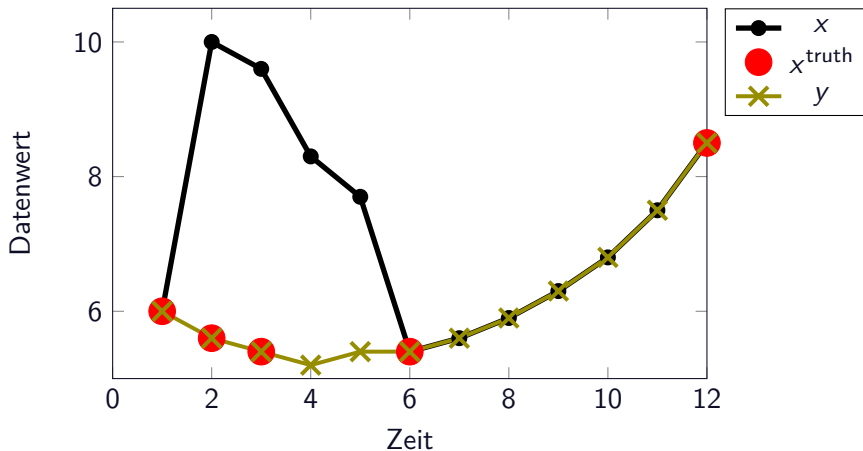
## Zeitreihenreparatur

- Gegeben:
  - Unzuverlässige Messung  $x = x[1], \dots, x[n]$
  - Unvollständige, aber dafür ausschließlich korrekte Messung  $x^{\text{truth}}$
  - (Nur bei der Evaluierung: vollständige, korrekte Messung  $x^{\text{truth}*}$ )
- Gesucht:
  - Reparatur  $y$  mit minimalen RMS-Fehler  $\Delta(x^{\text{truth}*}, y)$
  - $\Delta(x^{\text{truth}*}, y) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i^{\text{truth}*} - y_i)^2}$

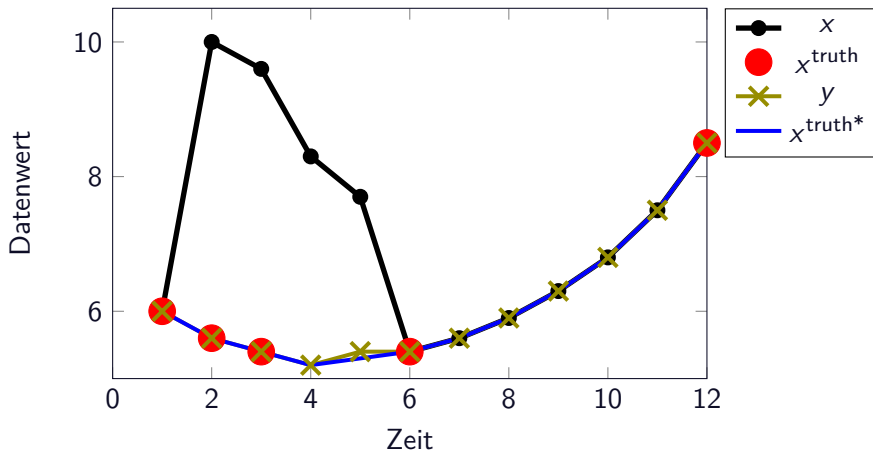
# Anomalien: Beispiel Eingabe



# Anomalien: Beispiel Eingabe & Reperatur



# Anomalien: Beispiel Eingabe & Reparatur



# Anomalien: Beispiel Zahlen & Bewertung

## Zeitreihen vom Beispiel

- $x = \{6, 10, 9.6, 8.3, 7.7, 5.4, 5.6, 5.9, 6.3, 6.8, 7.5, 8.5\}$
- $x^{\text{truth}} = \{6, 5.6, 5.4, \_, \_, 5.4, \_, \_, \_, \_, \_, 8.5\}$
- $y = \{6, 5.6, 5.4, \underline{5.2}, \underline{5.4}, \underline{5.6}, \underline{5.9}, \underline{6.3}, \underline{6.8}, \underline{7.5}, 8.5\}$
- $x^{\text{truth}*} = \{6, 5.6, 5.4, \underline{5.2}, \underline{5.3}, 5.4, \underline{5.6}, \underline{5.9}, \underline{6.3}, \underline{6.8}, \underline{7.5}, 8.5\}$

## Bewertung des Beispiels

$$\Delta(x^{\text{truth}*}, y) =$$

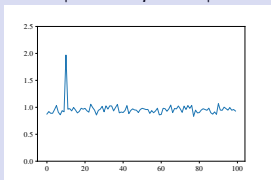
$$\sqrt{\frac{1}{12} ((6 - 6)^2 + \dots + (5.3 - 5.4)^2 + \dots + (8.5 - 8.5)^2)} \approx 0.03$$

# Anomalienreparaturen

## Anomalien

- Wikipedia: Abweichung von der Regel
- Werte  $x_i$  mit Abweichung  $\tau$  (Bsp.  $\tau = 2\sigma$ ):

$$|x_i - x_i^{\text{truth}^*}| > \tau$$



# Reparatur durch Anomalieerkennung

## Autoregressive Modell $AR(p)$

- Lineare Regression der letzte  $p$  Werte:

$$x'_t = \sum_{i=1}^p \phi_i x_{t-i} + \epsilon_t$$

- Reparatur:

$$y_t = \begin{cases} x'_t & \text{falls kein Label und } |x'_t - x_t| > \tau \\ x_t & \text{sonst} \end{cases}$$

# Reparatur durch Anomalieerkennung

## Autoregressives exogenes Modell $ARX(p)$

- Exogenes Variabel  $y$

$$y'_t = x_t + \sum_{i=1}^p \phi_i (y_{t-i} - x_{t-i}) + \epsilon_t$$

- $y'_t$  Mögliche Reparatur:

$$y_t = \begin{cases} y'_t & \text{falls kein Label und } |y'_t - x_t| > \tau \\ y_t & \text{sonst} \end{cases}$$

- Fakten



# Iterative Minimum Repairing

- 1 Einführung
- 2 Grundlagen
- 3 Iterative Minimum Repairing**
  - IMR
  - Optimierung 1: Matrix-Pruning IMR
- 4 Evaluierung
- 5 Schluss

## Intuitiver Ansatz von IMR

- ARX nutzt markierte Werte effizient, **aber** verändert die Werte zu drastisch.
- IMR Ansatz:
  - ① Wende ARX an
  - ② Wähle Reparaturwert aus mit minimalen Abstand zur Messung
  - ③ Wiederhole Prozedur
- Motivation: Reparierte Werte verbessern zukünftige Reparaturen

# IMR = ARX + Minimum-Change-Prinzip

- 
- 1: **Eingabe:** Messung  $x$ , markierte Werte  $x^{\text{truth}}$ , Ordnung  $p$ , Schwellenwert  $\tau$  und max-num-iterations
  - 2: **Ausgabe:** Reparatur  $y$
  - 3:  $y^{(0)} \leftarrow \text{Initialize}(x, x^{\text{truth}})$
  - 4: **for**  $k \leftarrow 0$  **to** max-num-iterations **do**
  - 5:    $\phi^{(k)} \leftarrow \text{Estimate}(x, y^{(k)})$
  - 6:    $\hat{y} \leftarrow \text{Candidate}(x, y^{(k)}, \phi^{(k)})$
  - 7:    $y^{(k+1)} \leftarrow \text{Evaluate}(x, y^{(k)}, \hat{y})$
  - 8:   **if**  $\text{Converge}(y^{(k)}, y^{(k+1)})$  **then**
  - 9:     **break**
  - 10:   **end if**
  - 11: **end for**
  - 12: **return**  $y^{(k)}$
-

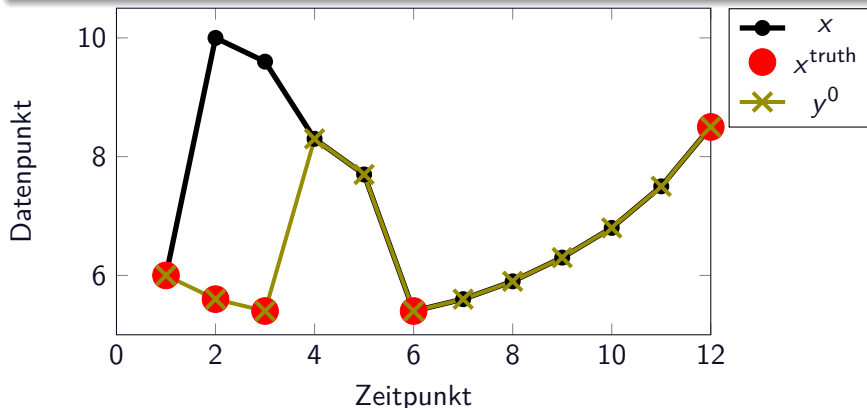
# IMR: Initialisierung

- 
- 
- 1: **Eingabe:** Messung  $x$ , markierte Werte  $x^{\text{truth}}$ , Ordnung  $p$ , Schwellenwert  $\tau$  und max-num-iterations
  - 2: **Ausgabe:** Reparatur  $y$
  - 3:  $y^{(0)} \leftarrow \text{Initialize}(x, x^{\text{truth}})$
  - 4: **for**  $k \leftarrow 0$  **to** max-num-iterations **do**
  - 5:    $\phi^{(k)} \leftarrow \text{Estimate}(x, y^{(k)})$
  - 6:    $\hat{y} \leftarrow \text{Candidate}(x, y^{(k)}, \phi^{(k)})$
  - 7:    $y^{(k+1)} \leftarrow \text{Evaluate}(x, y^{(k)}, \hat{y})$
  - 8:   **if**  $\text{Converge}(y^{(k)}, y^{(k+1)})$  **then**
  - 9:     **break**
  - 10:   **end if**
  - 11: **end for**
  - 12: **return**  $y^{(k)}$
-

# IMR: Initialisierung

## Initiale Reparatur

Initiale Reparatur  $y^{(0)}$  ist Messung  $x$  und übernimmt die markierten Werte aus  $x^{\text{truth}}$



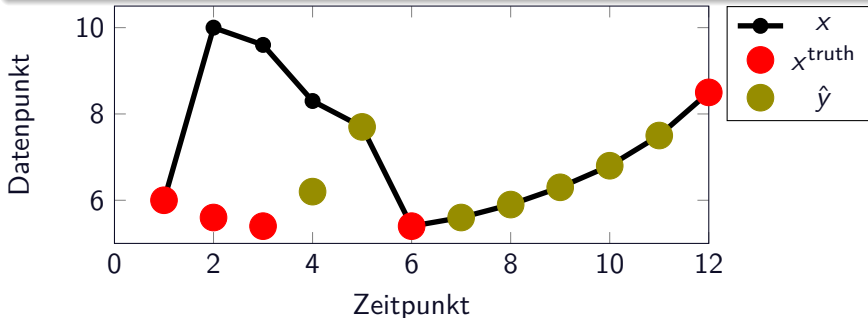
# IMR: ARX auf aktuelle Reparatur anwenden

- 1: **Eingabe:** Messung  $x$ , markierte Werte  $x^{\text{truth}}$ , Ordnung  $p$ , Schwellenwert  $\tau$  und max-num-iterations
- 2: **Ausgabe:** Reparatur  $y$
- 3:  $y^{(0)} \leftarrow \text{Initialize}(x, x^{\text{truth}})$
- 4: **for**  $k \leftarrow 0$  **to** max-num-iterations **do**
- 5:      $\phi^{(k)} \leftarrow \text{Estimate}(x, y^{(k)})$
- 6:      $\hat{y} \leftarrow \text{Candidate}(x, y^{(k)}, \phi^{(k)})$
- 7:      $y^{(k+1)} \leftarrow \text{Evaluate}(x, y^{(k)}, \hat{y})$
- 8:     **if**  $\text{Converge}(y^{(k)}, y^{(k+1)})$  **then**
- 9:         **break**
- 10:    **end if**
- 11: **end for**
- 12: **return**  $y^{(k)}$

# IMR: ARX auf aktuelle Reparatur anwenden

## Kandidaten

- Parameterschätzung  $\phi$ : aktuelle Reparatur  $y^{(k)}$  wird als  $x^{\text{truth}}$  interpretiert.
- Reparatur sind Kandidaten  $\hat{y}$  (angewendetes ARX kennt vorangegangene Aktualisierung nicht)



# IMR: Minimum-Change

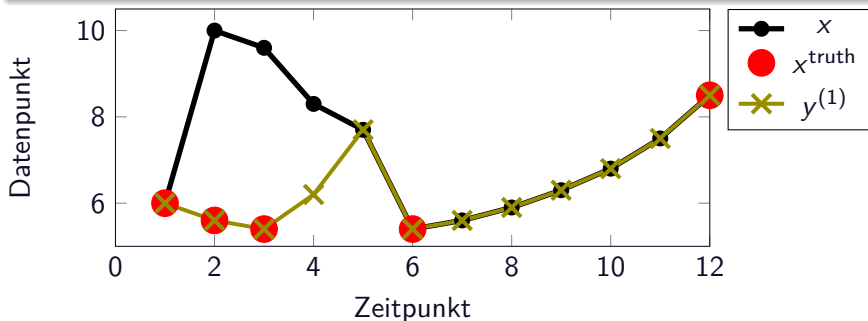
- 1: **Eingabe:** Messung  $x$ , markierte Werte  $x^{\text{truth}}$ , Ordnung  $p$ , Schwellenwert  $\tau$  und max-num-iterations
- 2: **Ausgabe:** Reparatur  $y$
- 3:  $y^{(0)} \leftarrow \text{Initialize}(x, x^{\text{truth}})$
- 4: **for**  $k \leftarrow 0$  **to** max-num-iterations **do**
- 5:    $\phi^{(k)} \leftarrow \text{Estimate}(x, y^{(k)})$
- 6:    $\hat{y} \leftarrow \text{Candidate}(x, y^{(k)}, \phi^{(k)})$
- 7:    $y^{(k+1)} \leftarrow \text{Evaluate}(x, y^{(k)}, \hat{y})$
- 8:   **if**  $\text{Converge}(y^{(k)}, y^{(k+1)})$  **then**
- 9:     **break**
- 10:   **end if**
- 11: **end for**
- 12: **return**  $y^{(k)}$



# IMR: Minimum-Change

## Minimum-Change

- Zu geringe Änderungen werden herausgefiltert  $|y_i^{(k)} - \hat{y}_i| > \tau$
- Geringste Änderung zu Messung  $x$  wird als Kandidaten herangezogen



# IMR: Terminierung

- 1: **Eingabe:** Messung  $x$ , markierte Werte  $x^{\text{truth}}$ , Ordnung  $p$ , Schwellenwert  $\tau$  und max-num-iterations
- 2: **Ausgabe:** Reparatur  $y$
- 3:  $y^{(0)} \leftarrow \text{Initialize}(x, x^{\text{truth}})$
- 4: **for**  $k \leftarrow 0$  **to** max-num-iterations **do**
- 5:    $\phi^{(k)} \leftarrow \text{Estimate}(x, y^{(k)})$
- 6:    $\hat{y} \leftarrow \text{Candidate}(x, y^{(k)}, \phi^{(k)})$
- 7:    $y^{(k+1)} \leftarrow \text{Evaluate}(x, y^{(k)}, \hat{y})$
- 8:   **if**  $\text{Converge}(y^{(k)}, y^{(k+1)})$  **then**
- 9:     **break**
- 10:   **end if**
- 11: **end for**
- 12: **return**  $y^{(k)}$

# IMR: Terminierung

## Terminierung

- Zwei Möglichkeiten der Terminierung:
  - Maximale Anzahl der Iterationen wird erreicht
  - Konvergenz: Neue Reparatur  $y^{(k+1)}$  ist gleich aktuelle Reparatur  $y^{(k)}$
- Allgemeine Konvergenzfrage ist noch offen

# Motivation von Matrix-Pruning IMR

## Laufzeitproblem

- Parameterschätzung beansprucht viel Zeit
- Matrizen  $V$  und  $Z$  bestehen aus  $y_i^{(k)} - x_i$ :
  - wenige markierte Werte vorhanden
  - markierte Werte häufig identisch zur Messung
  - Reparaturwerte ändern sich nicht signifikant
  - → dünnbesetzte Matrizen
- Matrix-Pruning: Löschen von Zeilen mit 0en

# Matrix Pruning IMR Beispiel

Beispiel

Beispiel

# Evaluierung

- 1 Einführung
- 2 Grundlagen
- 3 Iterative Minimum Repairing
- 4 Evaluierung**
  - Ordnung
  - Schwellenwert
  - maximale Anzahl von Iterationen
  - Markierungsrate
- 5 Schluss

# Ordnung

## Blank

- Faktenöasdkfjöasdöasdkfjöasd
- Fakten
- Fakten

# Schwellenwert

## Blank

- Fakten
- Fakten
- Fakten



# maximale Anzahl von Iterationen

## Blank

- Fakten
- Fakten
- Fakten

# Markierungsrate

## Blank

- Fakten
- Fakten
- Fakten

# Schluss

---

- 1 Einführung
- 2 Grundlagen
- 3 Iterative Minimum Repairing
- 4 Evaluierung
- 5 Schluss**
  - Zusammenfassung und Ausblick
  - Literatur

# Zusammenfassung und Ausblick

## Zusammenfassung

- Was wurde getan?

# Zusammenfassung und Ausblick

## Zusammenfassung

- Was wurde getan?

## Ausblick

- Wie könnten zukünftige Arbeiten aussehen?

# Literatur I

---



Shaoxu song - tsinghua university.



Aoqian Zhang, Shaoxu Song, Jianmin Wang, and Philip S Yu.

Time series data cleaning: From anomaly detection to anomaly repairing.

*Proceedings of the VLDB Endowment*, 10(10):1046–1057, 2017.