Template-Based Piecewise Affine Regression

Guillaume Berger and Sriram Sankaranarayanan

University of Colorado Boulder

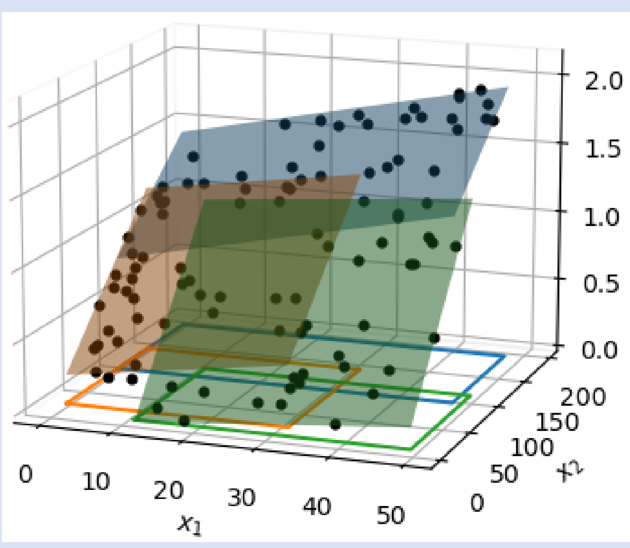
Problem of Interest

Given N data points (x_k, y_k), find q regions H_i and matrices A_i such that $x_k \in H_i \Rightarrow \|A_i x_k - y_k\| \le \epsilon$

Additional constraints on regions H_i:

- cover all points x_k
- belong to a template: $H_i = \{x : p(x) \le c_i\}$

Optimization form: minimize q (# regions)



Applications

- Model identification
- Template: adjust the complexity:
  - Tractable models
  - Avoid overfitting

Computational Complexity

NP-hard with respect to dimension of x_k

Proof: Reduction of Switched Linear Regression to Template-Based Linear Regression

Polynomial with respect to N (# data): $O(N^{qh})$

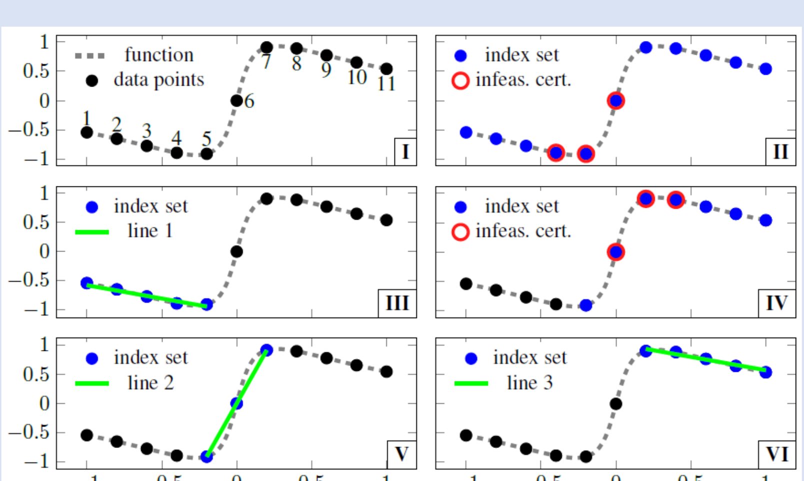
Proof: Enumerate all template-compatible sets of points and check for "linear fit"

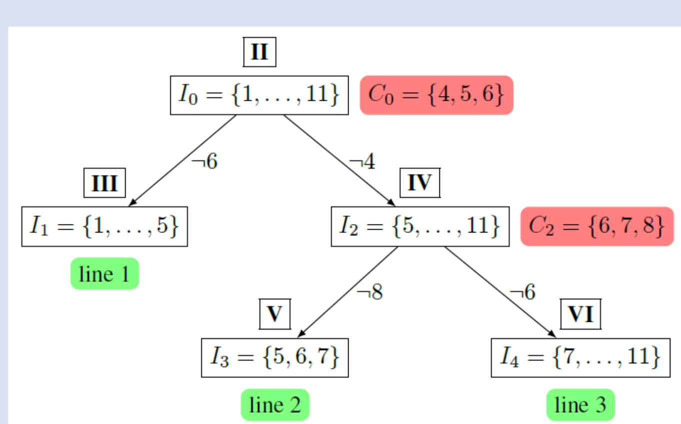
Highly data-inefficient!

Top-Down Approach

Starts from "large subsets" of data points (x_k, y_k)

- Pick a set S
- Check S for "linear fit"
- If S is not "fitting", split S into smaller template-compatible sets $S_1, \dots, S_h$
- Repeat until all sets fit





Split using Infeasibility Certificates

If S is not fitting, then there is a **certificate** $C \subseteq S$, $|C| \le d$, that is not fitting

Computation: using Linear Programming

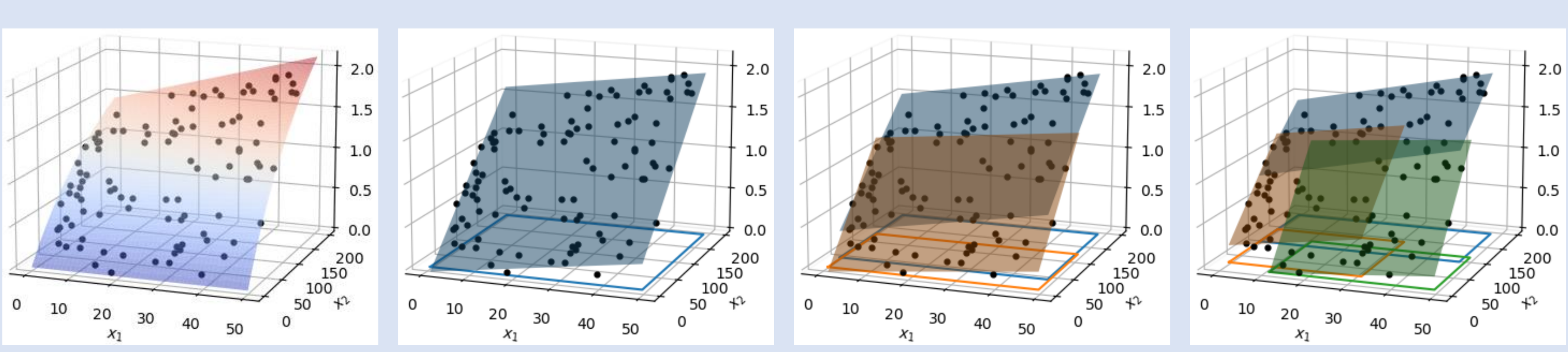
Split S into all **maximal template-compatible subsets** that do not include C

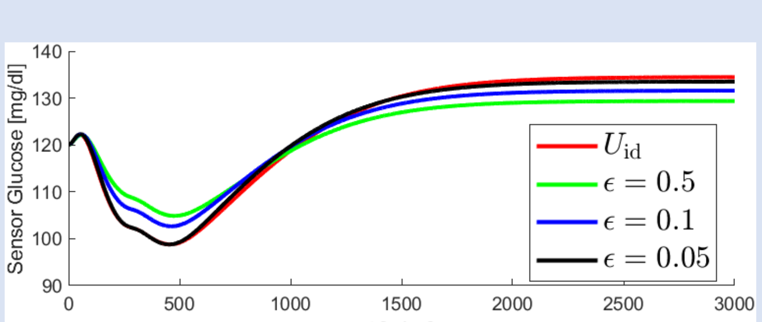
Computation: refine each component of the template $p(x) \le c$ to exclude at least one point of C

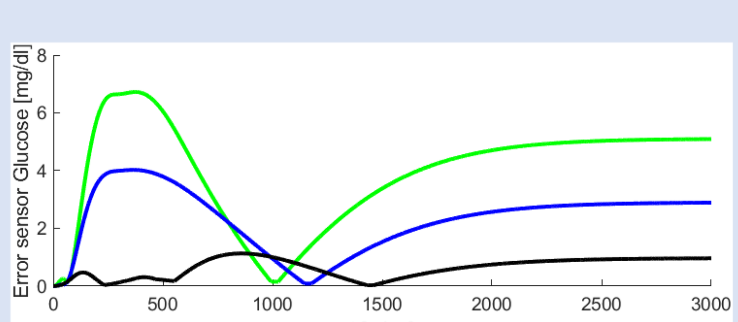
Theorem: This computes all maximal template-compatible linear-fitting subsets of data points

Insulin–Glucose Regulation Model

PWA approximation using rectangle regions

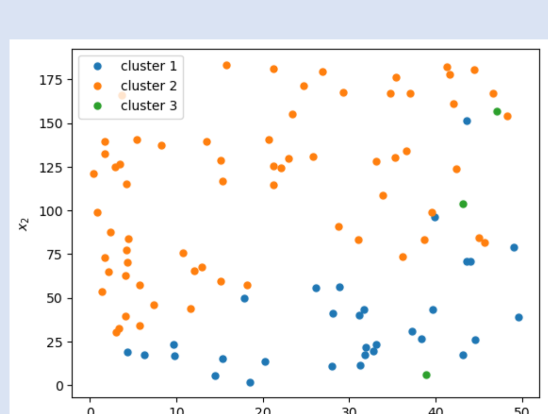






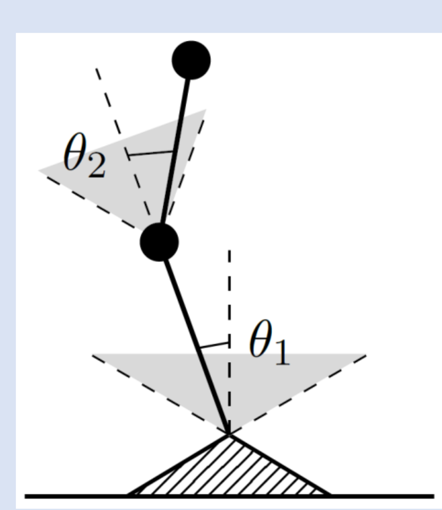
$N = 100$
Computation times: 1, 22, 112 secs

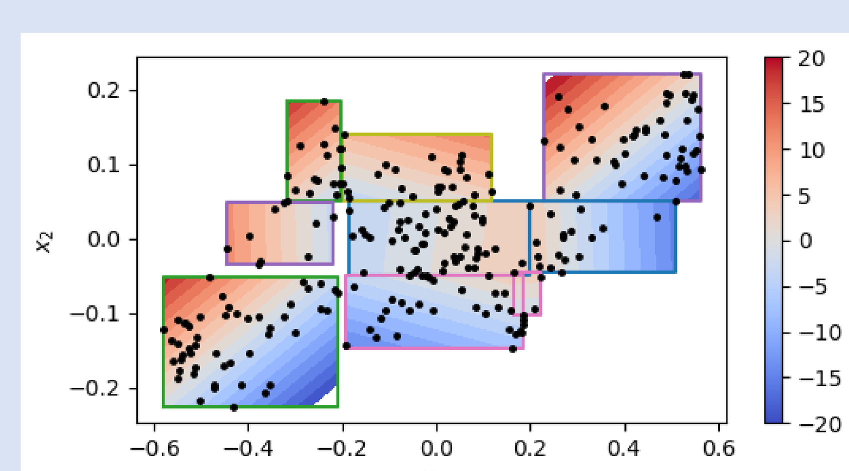
Switched Affine Regression: <10 secs, but not able to simple regions

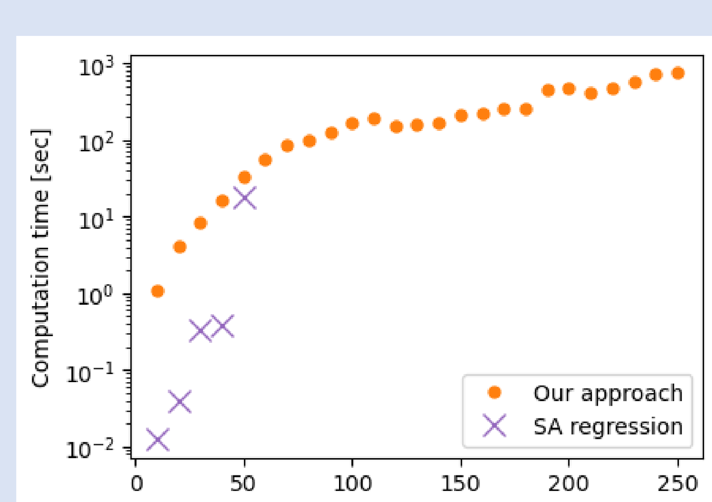


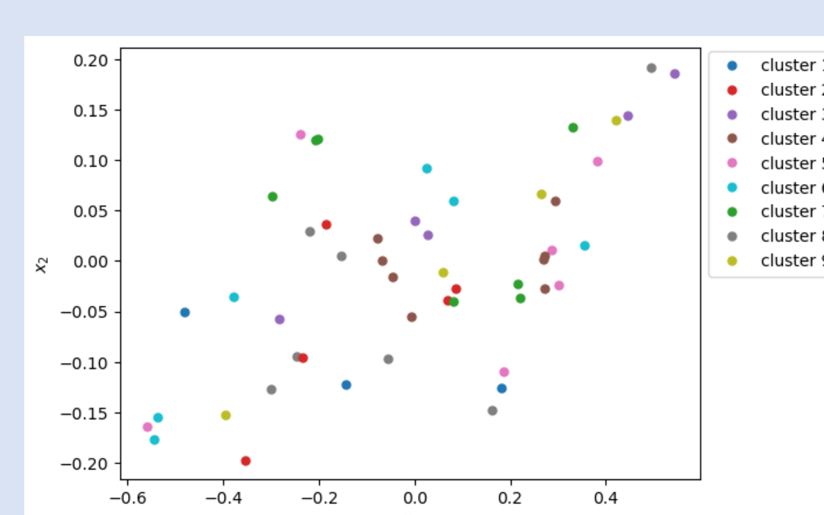
Comparison with MILP for SAR

Double Pendulum with Soft Contacts









Comparison with MILP for SAR