

# Ordered monotone regression

Kaspar Rufibach

Biostatistics Unit  
Institute of Social and Preventive Medicine  
University of Zurich

2010 Joint Statistical Meetings, Vancouver  
August 3, 2010

Joint work with **Fadoua Balabdaoui** and **Filippo Santambrogio**  
(both Paris-Dauphiné)

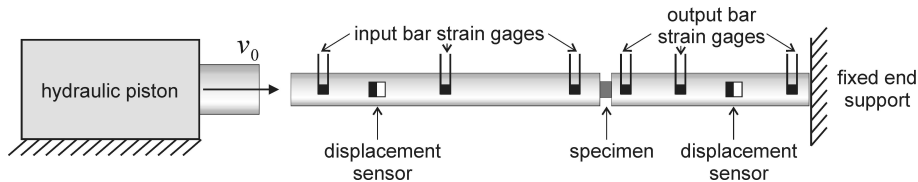
# Agenda

- 1 Motivating example
- 2 Isotonic regression
- 3 Ordered monotone regression
- 4 Algorithms
- 5 Smooth estimates
- 6 Back to stress-strain example

## Motivating example from mechanical engineering

Dynamic material tests, Shim and Mohr (2009).

Determine **deformation resistance** (strength of material) from uniaxial compression tests at **different loading velocities**.



# Motivating example from mechanical engineering

Yields **stress-strain** curves.

x-axis: **strain**,

- force applied to experimental unit via piston,
- negative logarithm of ratio of current over initial piston stroke,
- maximum shortening of piston stroke: 63%  $\Rightarrow$  x-values lie in  $[0, 1]$ .  
–  $\log(1) = 0$ ,  $-\log(1 - 0.63) = 1$ .
- x-values in fact measured with error  $\Rightarrow$  ignored.

y-axis: **stress**,

- measured compression of the sample,
- computed from cross-sectional area of specimen.

## Motivating example from mechanical engineering

Substantial noise due to

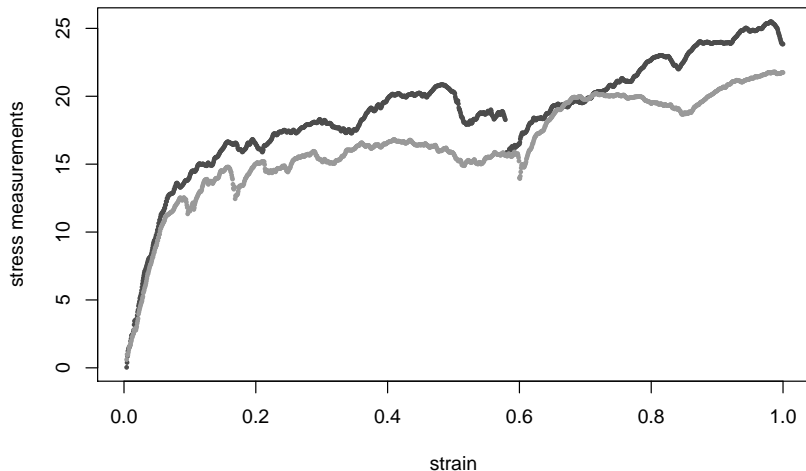
- variability in adjusting loading velocity and
- electrical noise in data acquisition.

Stress-strain curves at different loading velocities.

Regression curves are expected to be

- 1 monotone **increasing**  $\Rightarrow$  stress = increasing function of strain,
- 2 **ordered**  $\Rightarrow$  deformation resistance is higher for lower loading velocity.

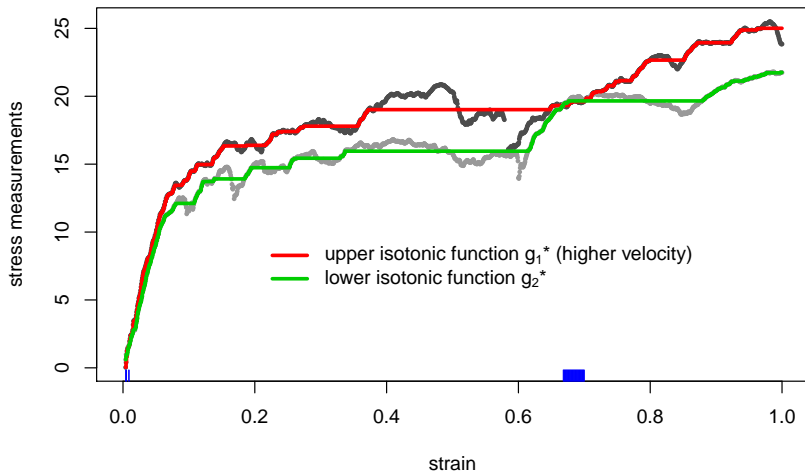
## Motivating example from mechanical engineering



1495 pairs  $(x_i, y_i)$  and  $(x_i, z_i)$  with

- $x_i$ : invoked strain,
- $y_i, z_i$ : stress measurements for **two loading velocities**.

## Naive approach



Two naive approaches to “estimation”:

- “Trial and error” with “parametric” functions  $\Rightarrow$  original approach!
- Two **independent** monotone functions  $\Rightarrow$  ordering not guaranteed.

## Problem

For data  $(x_i, y_i)$  and  $(x_i, z_i)$  find the minimizer of

$$L_2(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^n w_{1,i} (y_i - a_i)^2 + \sum_{i=1}^n w_{2,i} (z_i - b_i)^2$$

over  $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^n$  such that

- $\mathbf{a}$  and  $\mathbf{b}$  are **increasing** and
- $\mathbf{a} \leq \mathbf{b}$ , with  $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}_+^n$  given vectors of weights.

## A simpler problem

Minimize

$$L_1(\mathbf{a}) = \sum_{i=1}^n w_i (y_i - a_i)^2$$

over  $\mathbf{a} \in \mathbb{R}^n$  s.t.  $a_1 \leq a_2 \leq \dots \leq a_n$  where  $w_i > 0$ .

Two representations of the minimizer  $\hat{\mathbf{a}}^*$ :

- Slope of **greatest convex minorant (GCM)** of cumsum diagram.
- Via **min-max formula**:

$$\hat{a}_i^* = \max_{s \leq i} \min_{t \geq i} \sum_{i=s}^t y_i w_i / \sum_{i=s}^t w_i.$$

Used to derive theoretical results, not for computation.

Computation of  $\hat{\mathbf{a}}^*$  via **PAVA**.

## Two-curve problem: obvious questions

- 1 Geometrical characterizations using GCMs?
- 2 PAVA-like algorithm?
- 3 min-max representation (that only depends on data)?

## Two-curve problem: obvious questions

- 1 Geometrical characterizations using GCMs? **Not found.**
- 2 PAVA-like algorithm? **Not found.**
- 3 min-max representation (that only depends on data)? **Partially.**

## Two-curve problem: obvious questions

- 1 Geometrical characterizations using GCMs? **Not found.**
- 2 PAVA-like algorithm? **Not found.**
- 3 min-max representation (that only depends on data)? **Partially.**

Alternatives:

- 1 Analytical characterizations.
- 2 Projected subgradient algorithm (general), Dykstra's algorithm (least squares).
- 3 What we found: min-max formulas for...
  - ...first value of lower and last of upper curve.
  - ... $\hat{a}_i^*$  that depends on  $\hat{\mathbf{b}}^*$  and vice versa...

## Dykstra's algorithm

Problem of minimizing

$$L_2(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^n w_{1,i} (y_i - a_i)^2 + \sum_{i=1}^n w_{2,i} (z_i - b_i)^2$$

over  $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^n$  s.t.  $\mathbf{a}$  and  $\mathbf{b}$  are increasing and  $\mathbf{a} \leq \mathbf{b}$

is **equivalent**

to **least squares projection** of  $(\mathbf{y}, \mathbf{z})$  on intersection of **convex cones**

$\{(\mathbf{a}, \mathbf{b}) : \mathbf{a} \text{ is increasing}\}$ ,  $\{(\mathbf{a}, \mathbf{b}) : \mathbf{b} \text{ is increasing}\}$ , and  $\{(\mathbf{a}, \mathbf{b}) : \mathbf{a} \leq \mathbf{b}\}$ .

What the doctor ordered: algorithm by Dykstra (1983).

Simply **iterate projections**.

## Dykstra's algorithm

$$C_1 := \{(\mathbf{a}, \mathbf{b}) : \mathbf{a} \text{ is increasing}\},$$

$$C_2 := \{(\mathbf{a}, \mathbf{b}) : \mathbf{b} \text{ is increasing}\},$$

$$C_3 := \{(\mathbf{a}, \mathbf{b}) : \mathbf{a} \leq \mathbf{b}\}.$$

Computation of projections:

- $C_1, C_2$ : PAVA,
- $C_3$ : replace each pair  $(a_i, b_i)$  violating constraint (i.e.  $a_i > b_i$ ) by weighted average  $(w_{1,i}a_i + w_{2,i}b_i)/(w_{1,i} + w_{2,i})$ .

Dykstra's algorithm:

- Often converges slow  $\Rightarrow$  here very fast.
- Only three projections involved (that are easy to compute).
- Implemented in R package OrdMonReg.

## Projected subgradient algorithm

Dykstra's algorithm: only applicable to functionals of type

$$L_2(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^n w_{1,i}(y_i - a_i)^2 + \sum_{i=1}^n w_{2,i}(z_i - b_i)^2.$$

Projected subgradient algorithm: can handle

$$L_3(\mathbf{a}, \mathbf{b}) = F_{\mathbf{y}, \mathbf{w}_1}(\mathbf{a}) + \sum_{i=1}^n w_{2,i}(z_i - b_i)^2$$

where  $F$  is **convex** and **differentiable** (as a function of  $\mathbf{a}$ ).

Second term can likely be generalized as well.

Derivation and implementation more involved.  
Computes solution in least squares case slower.

## A smoothed version of the estimates

In fact: estimates only determined at  $x_i$ 's.

One-curve problem: estimate step function  $\Rightarrow$  derivative of GCM of cumsum.

We **define** estimates in two-curve as step functions as well  $\Rightarrow$  non-smooth.

Define smooth estimates for some kernel  $K_h$ :

$$\tilde{a}_h^*(x) = \frac{\sum_{i=1}^n K_h(x - x_i) \hat{a}_i^*}{\sum_{i=1}^n K_h(x - x_i)} \quad \tilde{b}_h^*(x) = \frac{\sum_{i=1}^n K_h(x - x_i) \hat{b}_i^*}{\sum_{i=1}^n K_h(x - x_i)}.$$

### Theorem (Mukerjee, 1988)

*A monotone function remains monotone after kernel smoothing if the kernel is **log-concave**.*

Proof: Log-concave densities have **monotone likelihood ratio**.

Ordering: maintained if we use same kernel and bandwidth for both functions  
 $\Rightarrow$  smoothing is simply taking weighted means.

## Back to our motivating example

1495 pairs  $(x_i, y_i), (x_i, z_i)$ .

$y_i, z_i$ : stress results for two different loading velocities.

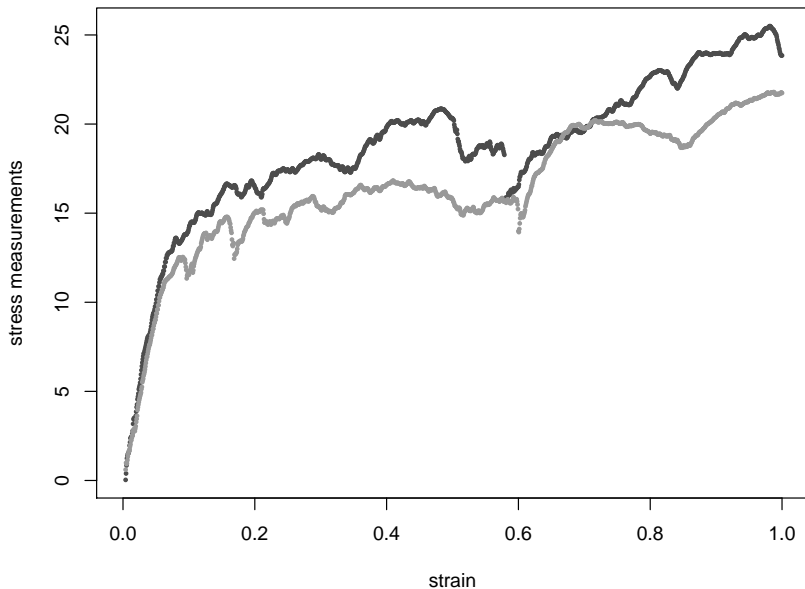
Estimated curves have to be **isotonic**, **ordered**, and ideally **smooth**.

Practitioners: fit “parametric” models using trial-and-error until above properties are met  $\Rightarrow$  ad-hoc, time-consuming.

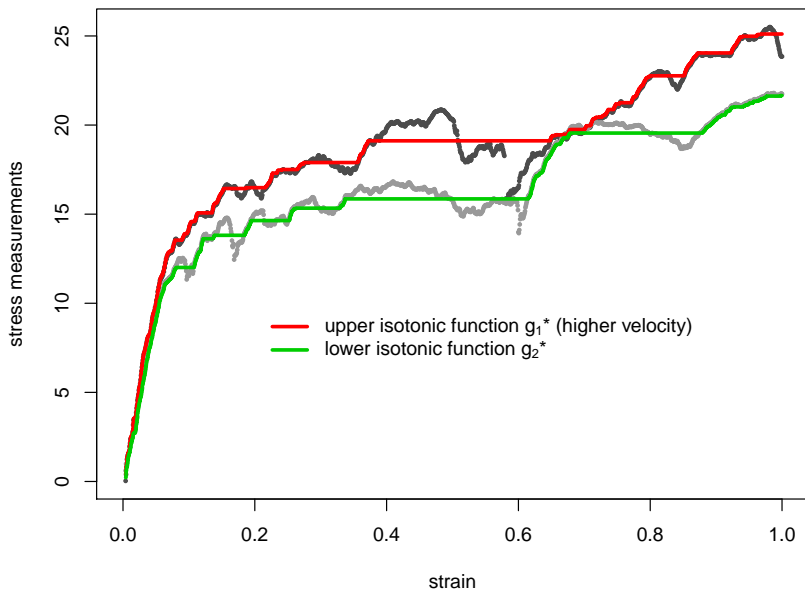
We use **normal** kernel  $K_h(x) = \phi(x/h) \Rightarrow$  log-concave.

Bandwidth:  $h = 0.1n^{-1/5} = 0.023$ .

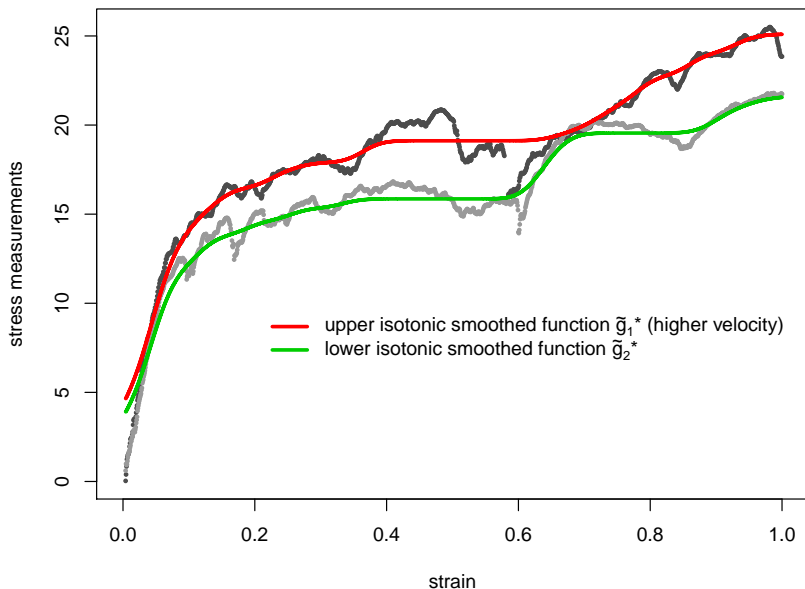
## Example: data



## Example: ordered monotone estimates



## Example: ordered monotone smoothed estimates



## Summary and outlook

### Summary:

- Existence, uniqueness, and characterization of ordered monotone regression.
- Dykstra's and projected subgradient algorithm.
- Smoothed estimates that maintain ordering and monotonicity.
- Algorithms for some generalized one- and two-curve problem implemented in R package `OrdMonReg`.

# Outlook

## Outlook:

- Initial problem of practitioners was **solved**.
- Further applications?
- More than two curves algorithm by Beran & Dümbgen (2010)?

## Setup a probabilistic model.

- Asymptotics (consistency, limiting distributions): no min-max formula available.
- Different approach: Instead of trying to find min-max formula for minimizer  $\hat{a}^*$   $\Rightarrow$  **define** estimates via min-max formula, proof consistency of such an estimate  $\Rightarrow$  PAVA at each  $x$ , similar to El Barmi & Mukerjee (2005) for survival functions.

Thank you for your attention.

Balabdaoui, F., Rufibach, K., Santambrogio, F. (2010).

*Least Squares estimation of two ordered monotone regression curves.*

J. Nonparametr. Stat., to appear.