Monotone polynomials in constrained mixed effects models

Joshua J. Bon with

B.A. Turlach & **K. Murray**

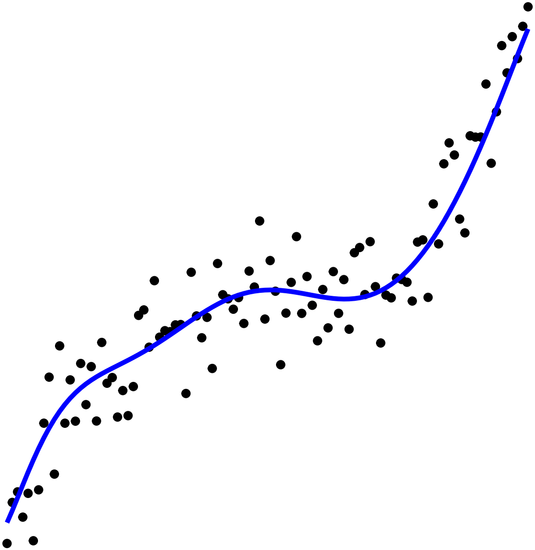
9th of May, 2017

School of Mathematics and Statistics, University of Western Australia

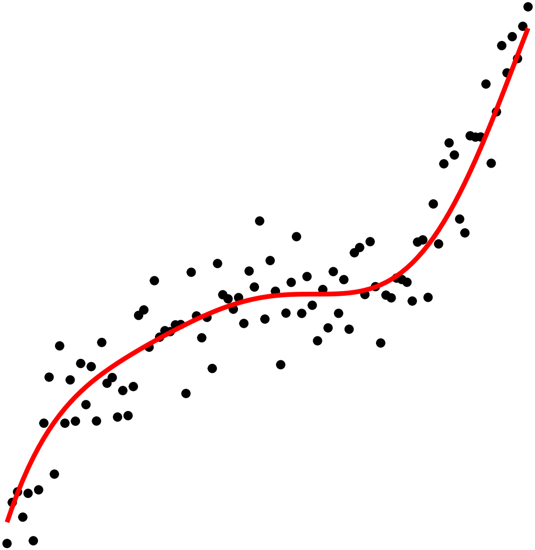
Introduction

An uphill battle

An uphill battle



An uphill battle



Some solutions

- Isotonic regression[1]
- Constrained smoothing splines[2]
- Reparameterised polynomial regression[3]

[1] J. Friedman, R. Tibshirani, *Technometrics* **26**, 243–250 (1984).
[2] I. P. Dierckx, *Computing* **24**, 349–371 (1980).
[3] K. Murray *et al.*, *Computational Statistics* **28**, 1989–2005 (2013).

Reparameterised polynomial regression

What do reparameterised polynomials offer?

Reparameterised polynomial regression

What do reparameterised polynomials offer?

+ Parametric interpretation (after transformation)
+ Likelihood based
+ Smooth curves
+ Continuous derivatives (inflection point calculation)
+ Implemented in `MonoPoly`[4] package in `R`

[4] K. Murray *et al.*, *Computational Statistics* **28**, 1989–2005 (2013).

Reparameterised polynomial regression

What do reparameterised polynomials offer?

+ Parametric interpretation (after transformation)
+ Likelihood based
+ Smooth curves
+ Continuous derivatives (inflection point calculation)
+ Implemented in `MonoPoly`[4] package in `R`

– Non-linear optimiser
– Not applicable to other (shape) constraints
– Can not accomodate mixed effects

[4] K. Murray *et al.*, *Computational Statistics* **28**, 1989–2005 (2013).

Outline

Aim: Develop a method for fitting monotone polynomials with mixed effects in a parametric frequentist framework.

Outline

Aim: Develop a method for fitting monotone polynomials with mixed effects in a parametric frequentist framework.

Results:

- COLS - Constrained **fixed** effects model estimation
- COLS & EM - Constrained *mean* **mixed** effects models
- COLS, EM, & RE truncation - Constrained *individual* curves
- Demonstration with monotonicity constraints

The least squares problem

Minimising the RSS...

$$\min_{\boldsymbol{\beta}} \left\{ (\boldsymbol{Y} - \boldsymbol{X}\beta)^T (\boldsymbol{Y} - \boldsymbol{X}\beta) \right\} \text{ s.t. } \boldsymbol{\beta} \in \Omega_{\boldsymbol{\beta}}$$

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \ \boldsymbol{X} = \begin{bmatrix} p_0(x_1) & p_1(x_1) & \cdots & p_q(x_1) \\ p_0(x_2) & p_1(x_2) & \cdots & p_q(x_2) \\ \vdots & \vdots & & \vdots \\ p_0(x_n) & p_1(x_n) & \cdots & p_q(x_n) \end{bmatrix}, \ \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \end{bmatrix}$$

using polynomial basis defined by the p_i's of degree i.

...with monotonicity

Take, for example, the set of parameters describing a monotonically increasing polynomial,

$$\Omega_{\boldsymbol{\beta}} = \{\boldsymbol{\beta} : p'(x; \boldsymbol{\beta}) \geq 0, \forall x \in S\}$$

...with monotonicity

Take, for example, the set of parameters describing a monotonically increasing polynomial,

$$\Omega_{\boldsymbol{\beta}} = \{\boldsymbol{\beta} : p'(x; \boldsymbol{\beta}) \geq 0, \forall x \in S\}$$

What can we say about $\Omega_{\boldsymbol{\beta}}$?

- $\Omega_{\boldsymbol{\beta}} \neq$ a finite set of parameter inequalities (e.g. $\beta_i \geq a_i$)
- Boundaries for each β_i are dependent
- We **can** check if $p(x; \boldsymbol{\beta}) \in \Omega_{\boldsymbol{\beta}}$

A new solution

A new solution

We use two complementary techniques to optimise the RSS.

- A coordinate descent algorithm
- An orthonormal design matrix

Coordinate descent for constrained problems

Coordinate descent:

- Minimise each coordinate of input successively
- Take "blind" step in direction that minimises objective function
- Find best permissible value with line search

Conditioning the least squares problem

Monomial polynomials are highly dependent, resulting in;

- Ill-conditioned least squares problem
- High coefficient correlation, inferential problems[5,6]
- Slower coordinate descent

[5] R. A. Bradley, S. S. Srivastava, *The American Statistician* **33**, 11–14 (1979).

[6] S. C. Narula, *International Statistical Review* **47**, 31–36 (1979).

Conditioning the least squares problem

Monomial polynomials are highly dependent, resulting in;

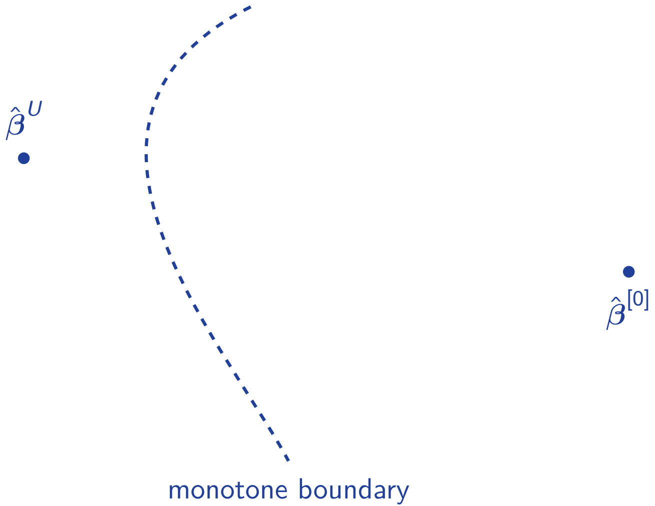
- Ill-conditioned least squares problem
- High coefficient correlation, inferential problems[5,6]
- Slower coordinate descent

Orthonormal design using discrete orthonormal polynomials removes a source of dependence;

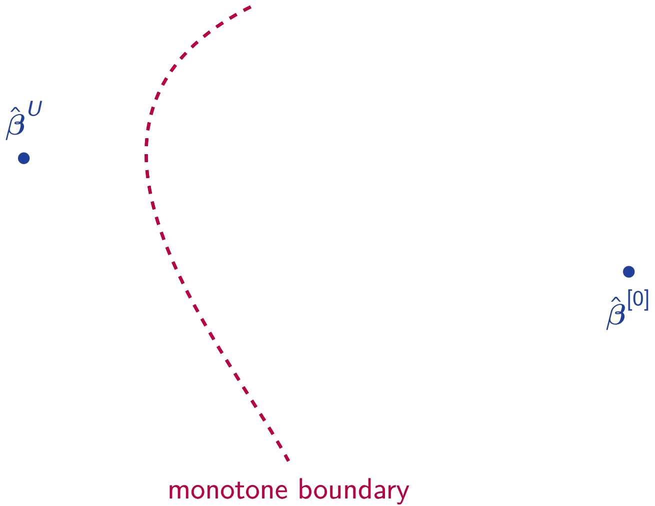
- $\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{I}$
- $\frac{\partial \text{RSS}}{\partial \beta_i} = f(\beta_i)$

[5] R. A. Bradley, S. S. Srivastava, *The American Statistician* **33**, 11–14 (1979).
[6] S. C. Narula, *International Statistical Review* **47**, 31–36 (1979).

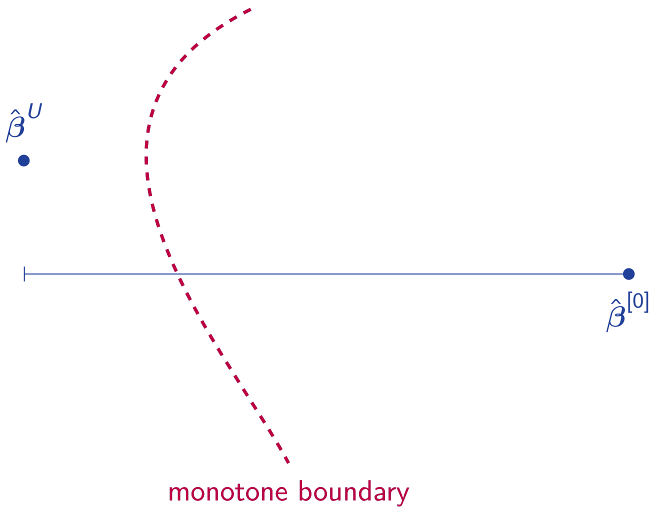
Coordinate descent - line search



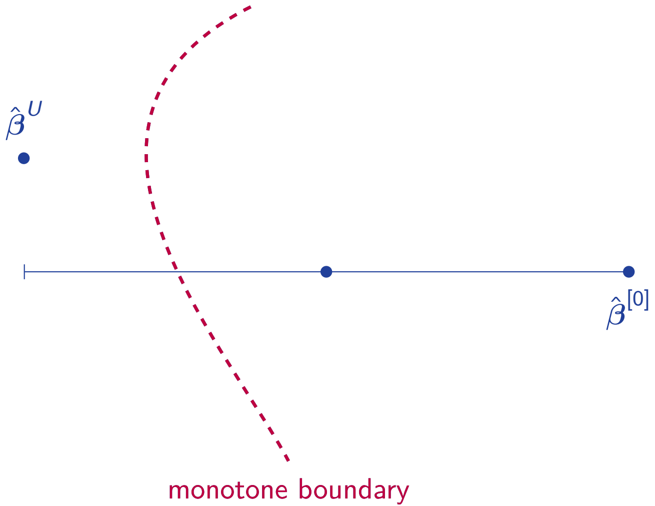
Coordinate descent - line search



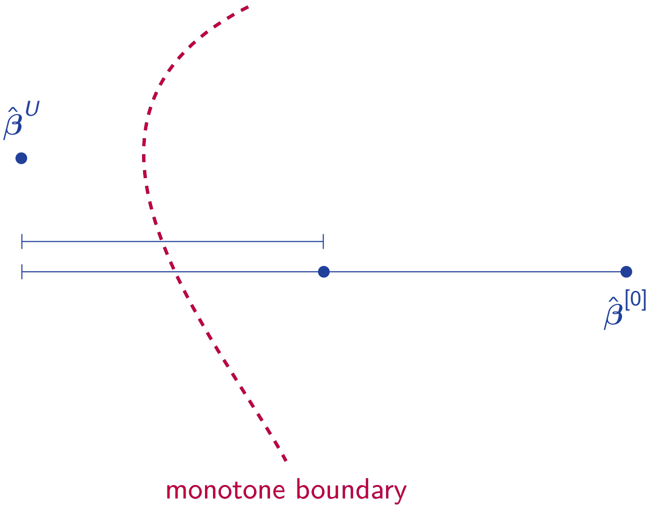
Coordinate descent - line search



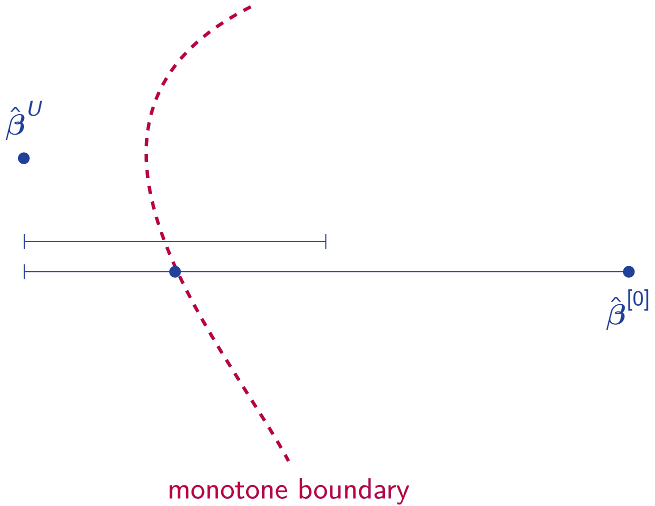
Coordinate descent - line search



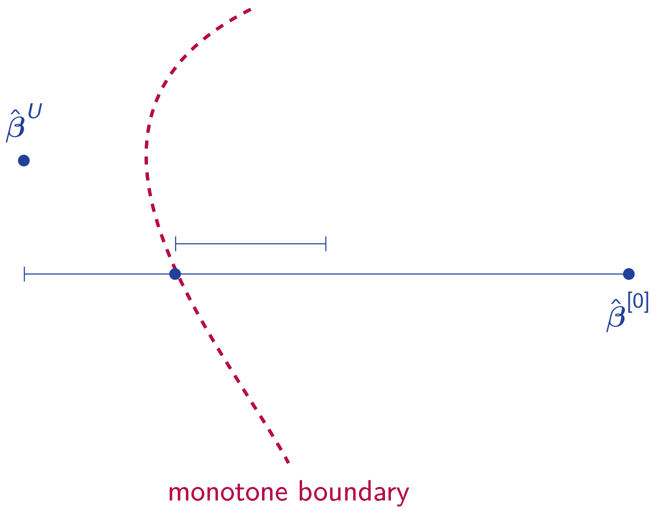
Coordinate descent - line search



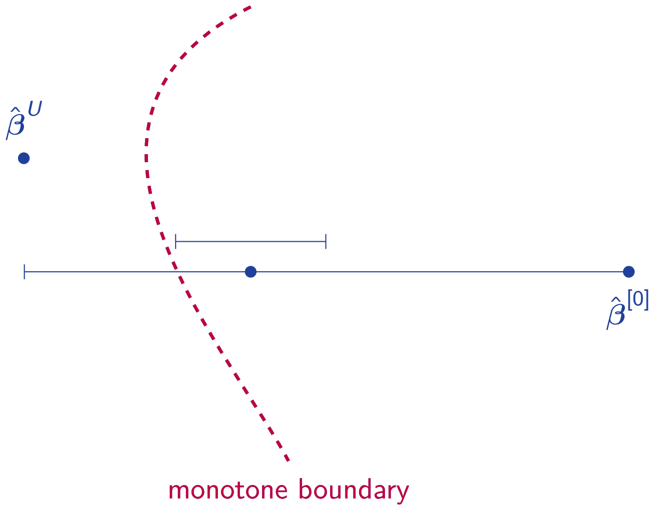
Coordinate descent - line search



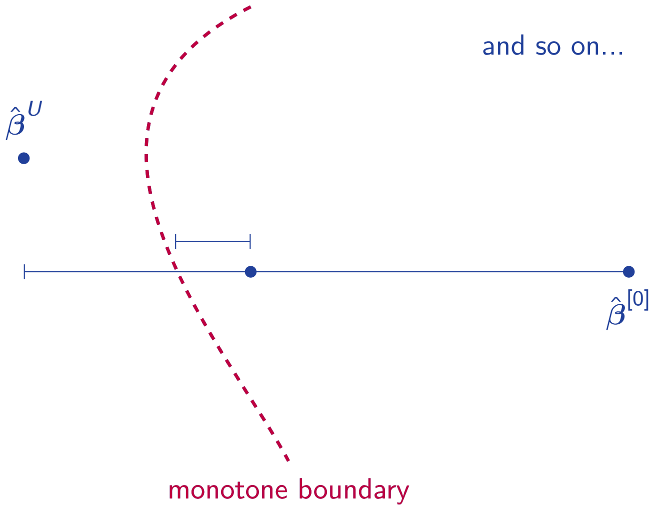
Coordinate descent - line search



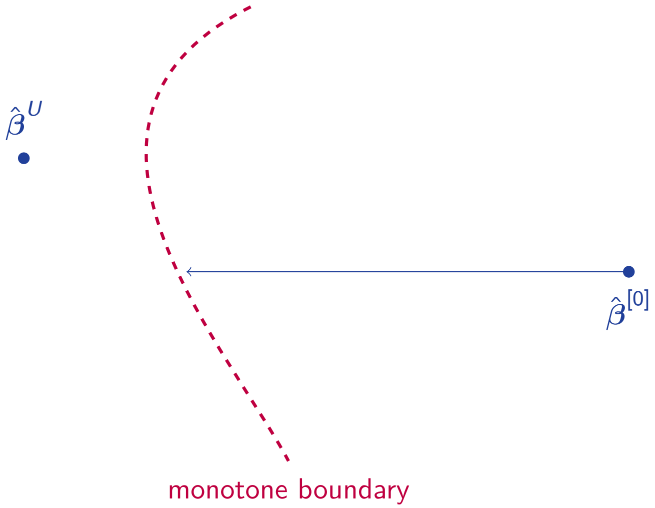
Coordinate descent - line search



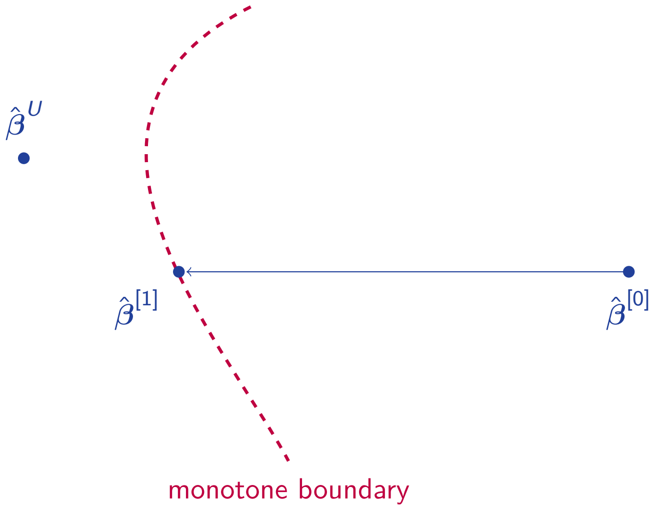
Coordinate descent - line search



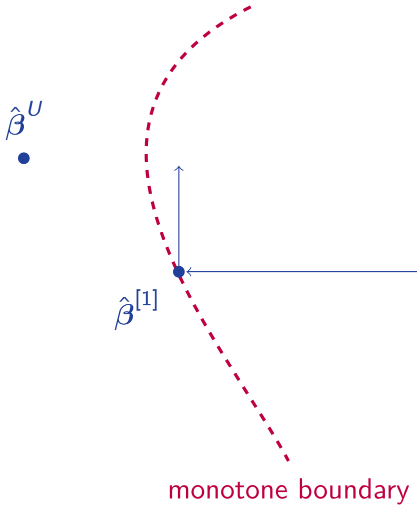
Coordinate descent - coordinate iterations



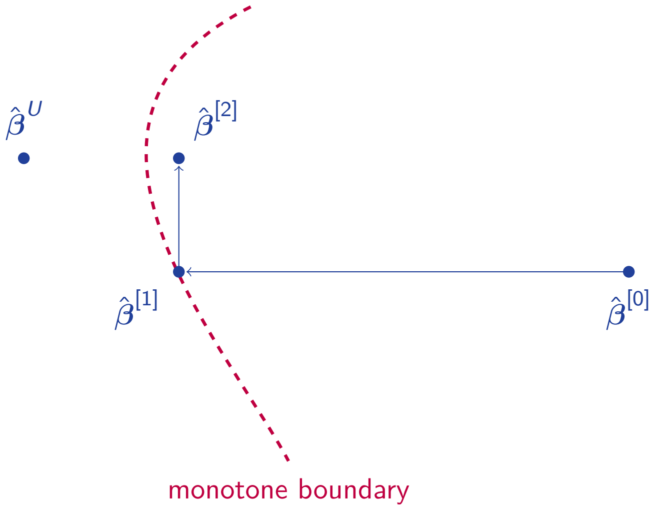
Coordinate descent - coordinate iterations



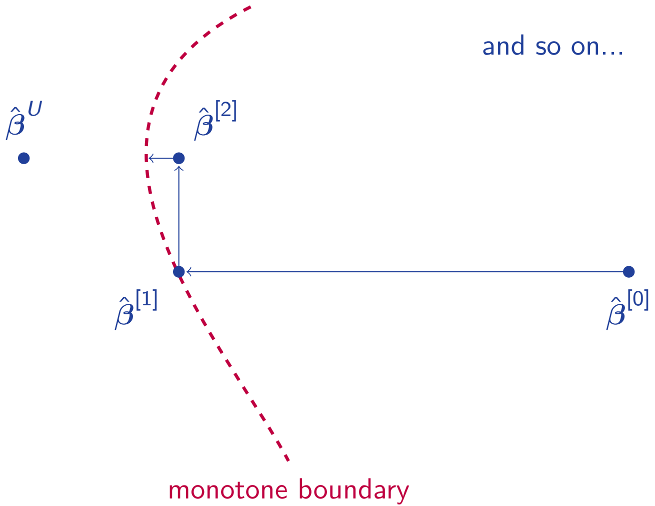
Coordinate descent - coordinate iterations



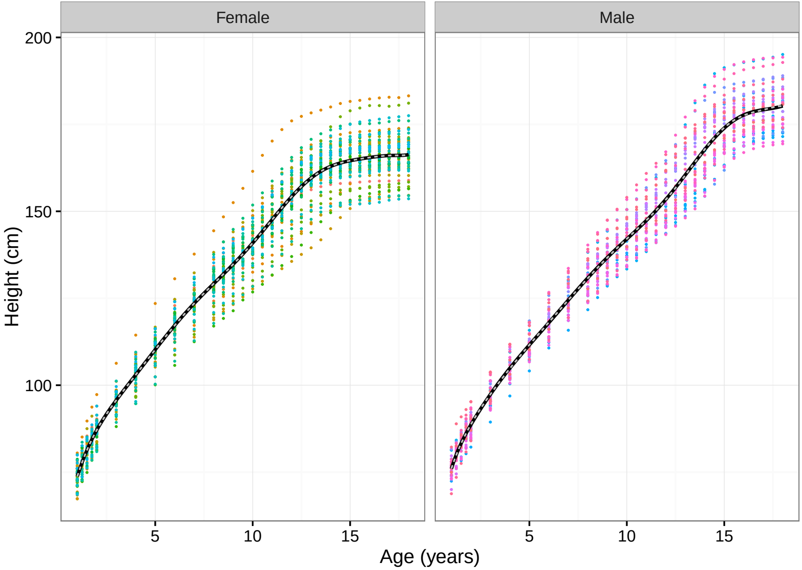
Coordinate descent - coordinate iterations



Coordinate descent - coordinate iterations



Demonstration on the Berkeley Growth Dataset



Demonstration on the Berkeley Growth Dataset

Male fit ($n = 1,209$)				
	OLS (1)	`MonoPoly` (2)	**COLS (3)**	Diff. (%) (3) - (2)
Monotonic fit?	No	Yes	**Yes**	
$\hat{\beta}_0$	141.33	141.27	**141.27**	0.00
$\hat{\beta}_1$	46.77	45.99	**45.98**	-0.03
$\hat{\beta}_2$	-8.70	-4.84	**-4.80**	-0.75
$\hat{\beta}_3$	69.40	88.83	**89.11**	0.32
$\hat{\beta}_4$	128.97	86.85	**86.58**	-0.31
$\hat{\beta}_5$	-159.89	-291.42	**-292.94**	0.52
$\hat{\beta}_6$	-449.39	-295.44	**-294.81**	-0.21
$\hat{\beta}_7$	55.33	415.40	**418.63**	0.78
$\hat{\beta}_8$	544.75	321.54	**321.07**	-0.15
$\hat{\beta}_9$	131.54	-297.78	**-300.64**	0.96
$\hat{\beta}_{10}$	-231.37	-120.38	**-120.33**	-0.05
$\hat{\beta}_{11}$	-93.63	92.01	**92.86**	0.93
RSS	42051.86	42060.93	**42060.93**	0.00
Runtime (secs)	< 0.01	17.01	**4.39**	-74.19

Demonstration on the Berkeley Growth Dataset

Female fit ($n = 1,674$)				
	OLS	MonoPoly	**COLS**	Diff. (%)
	(1)	(2)	**(3)**	(3) - (2)
Monotonic fit?	Yes	Yes	**Yes**	
$\hat{\beta}_0$	139.96	139.96	**139.96**	0.00
$\hat{\beta}_1$	56.41	56.41	**56.41**	0.00
$\hat{\beta}_2$	28.34	28.34	**28.34**	0.00
$\hat{\beta}_3$	-22.92	-22.92	**-22.92**	0.00
$\hat{\beta}_4$	-235.65	-235.65	**-235.65**	0.00
$\hat{\beta}_5$	-37.00	-37.00	**-37.00**	-0.01
$\hat{\beta}_6$	498.14	498.13	**498.14**	0.00
$\hat{\beta}_7$	169.16	169.16	**169.16**	0.00
$\hat{\beta}_8$	-480.85	-480.85	**-480.85**	0.00
$\hat{\beta}_9$	-219.32	-219.32	**-219.32**	0.00
$\hat{\beta}_{10}$	172.03	172.03	**172.03**	0.00
$\hat{\beta}_{11}$	101.86	101.87	**101.86**	0.00
RSS	55297.89	55297.89	**55297.89**	0.00
Runtime (secs)	< 0.01	17.03	**3.96**	-76.75

Constrained Orthogonal Least Squares (COLS) estimation

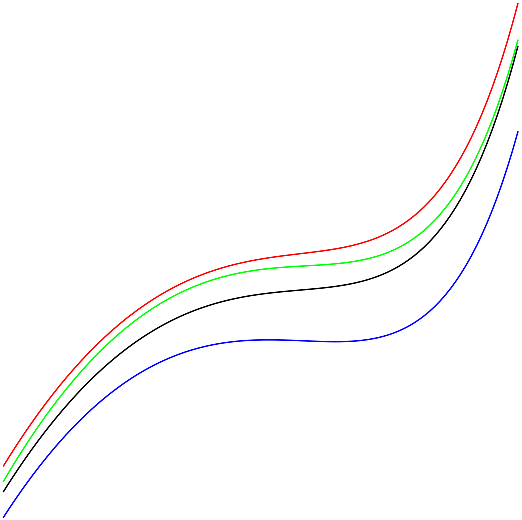
COLS in summary;

+ Testing suggests it may be faster than existing methods
+ Requires only linear reparametrisation
+ Applies to any closed convex parameter space
  + Difficult constraints such as monotonicity
  + Multiple constraints
+ Can be used in mixed effects models

– Iterative, not a closed form solution (like OLS)

Polynomial mixed effects models

Polynomial Mixed effects models



Estimation methodology

Two questions:

1. How do we constrain the **mean** polynomial curve to be monotonic?
2. How do we constrain **individuals'** polynomial curves to be monotonic, in addition to the mean curve?

Estimation methodology

Two questions:

1. How do we constrain the **mean** polynomial curve to be monotonic?
2. How do we constrain **individuals'** polynomial curves to be monotonic, in addition to the mean curve?

Suggested methods:

A1. The **Expectation-Maximisation**[7] algorithm and COLS

A2. Truncated multivariate normal distribution

---

[7] A. P. Dempster *et al.*, *Journal of the Royal Statistical Society. Series B (Methodological)* **39**, 1–38 (1977).

1. Constraining the mean curve

Advantages of the **Expectation-maximisation algorithm**;

- Separates mean estimation from random effects estimation
  - COLS on RSS-like problem
- Flexible for random effects
  - Constrained
  - MCEM for non-standard random effects[8]
- Already tested on mixed effects models[9]
- Convergence properties on constrained parameter space hold[10]

---

[8] J. G. Booth, J. P. Hobert, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **61**, 265–285 (1999).
[9] N. Laird *et al.*, *Journal of the American Statistical Association* **82**, 97–105 (1987).
[10] D. Nettleton, *Canadian Journal of Statistics* **27**, 639–648 (1999).

1. Constraining the mean curve

1. Initialise parameters
2. E-step: $\boldsymbol{U}^{[t]} = \mathbb{E}\left(\boldsymbol{\mathcal{U}} \mid \boldsymbol{Y}, \beta^{[t-1]}\right)$, with $\boldsymbol{\mathcal{U}} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{G}\right)$
3. M-step: Minimise RSS with COLS and $\boldsymbol{Y}^{*} = \boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{U}^{[t]}$

$$\beta^{[t]} = \arg\min_{\boldsymbol{\beta}} \left\{ \left(\boldsymbol{Y}^{*} - \boldsymbol{X}\beta\right)^{T} \left(\boldsymbol{Y}^{*} - \boldsymbol{X}\beta\right) \right\} \text{ s.t. } \beta \in \Omega_{\boldsymbol{\beta}}$$

4. M-step: Update variance parameters
5. Iterate through E-steps and M-steps until convergence

2. Constraining the mean and individuals' curves

1. Initialise parameters
2. E-step: $\boldsymbol{U}^{[t]} = \mathbb{E}\left(\boldsymbol{\mathcal{U}}_T \mid \boldsymbol{Y}, \beta^{[t-1]}\right)$, with $\boldsymbol{\mathcal{U}}_T \sim \mathcal{N}_{T(\beta)}\left(\mathbf{0}, \boldsymbol{G}\right)$
3. M-step: Minimise RSS with COLS and $\boldsymbol{Y}^* = \boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{U}^{[t]}$

$$\beta^{[t]} = \arg\min_{\boldsymbol{\beta}} \left\{ \left(\boldsymbol{Y}^* - \boldsymbol{X}\beta\right)^T \left(\boldsymbol{Y}^* - \boldsymbol{X}\beta\right) - \eta\left(\beta\right) \right\} \text{ s.t. } \boldsymbol{\beta} \in \Omega_{\boldsymbol{\beta}}$$

4. M-step: Update variance parameters
5. Iterate through E-steps and M-steps until convergence

2. Constraining the mean and individuals' curves

Complications from constraining individuals' curves;

- $\eta(\boldsymbol{\beta})$, the "penalty" term from truncation
- $\mathbb{E}\left(\boldsymbol{\mathcal{U}}_T \mid \boldsymbol{Y}, \boldsymbol{\beta}^{[t-1]}\right)$

Complications from constraining individuals' curves;

- $\eta(\boldsymbol{\beta})$, the "penalty" term from truncation
- $\mathbb{E}\left(\boldsymbol{\mathcal{U}}_T \mid \boldsymbol{Y}, \boldsymbol{\beta}^{[t-1]}\right)$

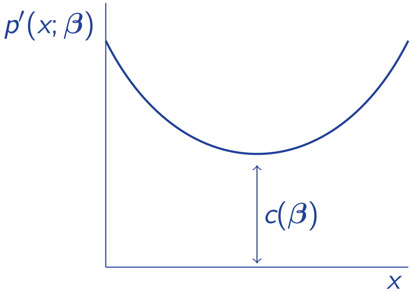
$$\eta(\boldsymbol{\beta}) = \log\left(\int_{T(\boldsymbol{\beta})} \left((2\pi)^{rg}|\boldsymbol{G}|\right)^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{W}^T\boldsymbol{G}^{-1}\boldsymbol{W}\right\} \mathrm{d}\boldsymbol{W}\right)$$

When $r = 2$ the truncation is point-wise:

$$T(\boldsymbol{\beta}) = \left\{\boldsymbol{\mathcal{U}}_T = \left[u_{0,1}\; u_{1,1} \cdots u_{0,g}\; u_{1,g}\right]^T \in \mathbb{R}^{2g} \text{ s.t. } u_{i,1} \geq -c(\boldsymbol{\beta})\right\}$$

2. Constraining the mean and individuals' curves

$$u_{i,1} \geq -c(\boldsymbol{\beta})$$



2. Constraining individuals' curves

For $r = 2$;

- Expectation from point-truncated normal theory
- Analytical differentiation of $\eta(\boldsymbol{\beta})$ from chain rules. Envelope theorem for $c(\boldsymbol{\beta})$

For $r \geq 3$;

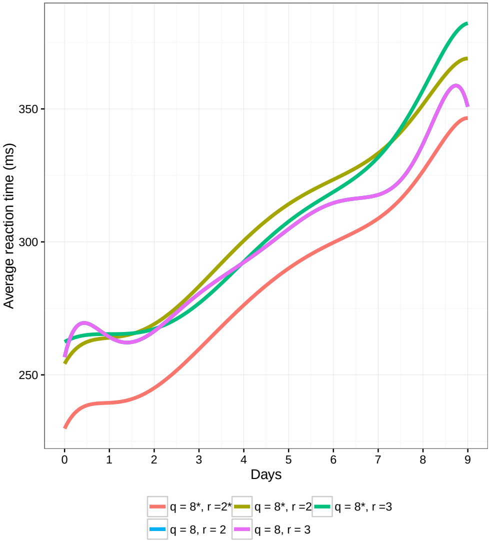
- Monte Carlo EM to deal with expectation
- Numerical differentiation of $\eta(\boldsymbol{\beta})$

Sleep Study Data - Degree 8 mean curves

$S = [0, 9]$

$r =$ random effects

$* =$ constrained

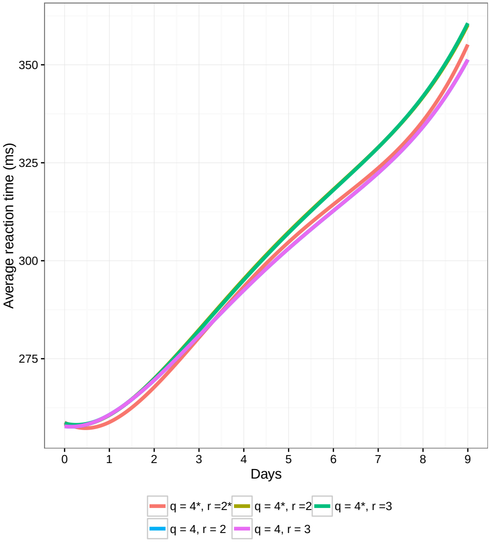


Sleep Study Data - Degree 4 mean curves

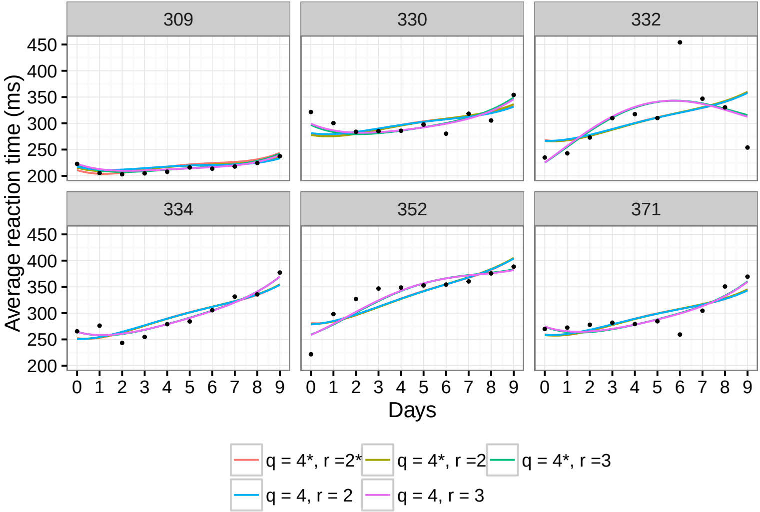
$S = [0, 9]$

$r =$ random effects

$* =$ constrained



Sleep Study Data - Degree 4 individual curves



Conclusion

Conclusion - Fixed effects models

For **fixed** effects models, this work has delivered;

- COLS - a new method constrained regression (on closed, convex sets)
- Opens up possibilities for shape constraints, joint constraints, and more...
- Can extend beyond polynomials of a single variable

Conclusion - Mixed effects models

For **mixed** effects models;

- Demonstrated COLS can estimate these with an EM-algorithm
- Derived full method for $r = 2$ with and without constrained individuals' curves
- Suggested MCEM to extend for $r \geq 3$
- Widely useful because of the flexibility of COLS and the EM-algorithm

Questions?

Appendix

Reparameterised polynomial regression

For example a monotonic polynomial can be written as[11]

$$p(x) = \delta + \alpha \int_0^x \prod_{j=1}^{K} \left\{ 1 + 2b_j t + \left(b_j^2 + c_j^2 \right) t^2 \right\} \mathrm{d}t \qquad (1)$$

with unconstrained parameters δ, b_j's, and c_j's.

[11] C. D. Elphinstone, *Communications in Statistics - Theory and Methods* **12**, 161–198 (1983), D. M. Hawkins, *Computational Statistics* **9**, 233–247 (1994), D. Heinzmann, *Computational Statistics* **23**, 343–360 (2008).

Conditioning the least squares problem

For better properties we use discrete orthonormal polynomials

$$\boldsymbol{X}_o = \begin{bmatrix} p_0(x_1) & p_1(x_1) & p_2(x_1) & \cdots & p_q(x_1) \\ p_0(x_2) & p_1(x_2) & p_2(x_2) & \cdots & p_q(x_2) \\ \vdots & \vdots & \vdots & & \vdots \\ p_0(x_n) & p_1(x_n) & p_2(x_n) & \cdots & p_q(x_n) \end{bmatrix}$$

$$\langle p_i, p_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \qquad \text{where } \langle f, g \rangle = \sum_{x \in D} f(x) g(x)$$

Conditioning the least squares problem

For better properties we use discrete orthonormal polynomials;

- Discrete orthonormal polynomials results in an orthonormal design matrix

$$\boldsymbol{X}_o^T \boldsymbol{X}_o = \boldsymbol{I}_q$$

- Calculate $\boldsymbol{X}_o$ with a QR decomposition or as in Emerson[12]

[12] P. L. Emerson, *Biometrics* **24**, 695–701 (1968).

Conditioning the least squares problem

$$\min_{\boldsymbol{\beta}} \left\{ \text{RSS}(\boldsymbol{\beta}) \right\} \text{ s.t. } \boldsymbol{\beta} \in \Omega_{\boldsymbol{\beta}}$$

	monomial ($\boldsymbol{X}$)	orthonormal ($\boldsymbol{X}_o$)
$\frac{\partial \text{RSS}}{\partial \boldsymbol{\beta}}$	$2\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}^T\boldsymbol{Y}\right)$	$2\left(\boldsymbol{\beta} - \boldsymbol{X}_o^T\boldsymbol{Y}\right)$
$\hat{\boldsymbol{\beta}}^U$	$\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}$	$\boldsymbol{X}_o^T\boldsymbol{Y}$

Coordinate descent theory

Good global convergence properties when[13]

- Parameter space closed and convex
- Object function continuously differentiable

Both satisfied by monotone polynomials over RSS.

- Monotone increasing/decreasing
- Over $\mathbb{R}$ or a compact subset of $\mathbb{R}$
- Over a broad range of difficult constraints

[13] A Cassioli *et al.*, *European Journal of Operational Research* **231**, 274–281 (2013).

Mixed effects models

One way to define the underlying probability model is with;

- a conditional normal distribution for $\boldsymbol{\mathcal{Y}}$

$$(\boldsymbol{\mathcal{Y}} \,|\, \boldsymbol{\mathcal{U}} = \boldsymbol{U}) \sim \mathcal{N}\left(\boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{U}, \boldsymbol{R}\right)$$

- and a normal distribution for $\boldsymbol{\mathcal{U}}$

$$\boldsymbol{\mathcal{U}} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{G}\right)$$

Mixed effects models

This allows the joint pseudo-log-likelihood function to be written as

$$
\begin{aligned}
& l_{\mathcal{Y},\mathcal{U}}(\beta, \phi_{\boldsymbol{R}}, \phi_{\boldsymbol{G}} \mid \boldsymbol{Y}, \mathcal{U}) \\
&= l_{\mathcal{Y}|\mathcal{U}}(\beta, \phi_{\boldsymbol{R}}, \phi_{\boldsymbol{G}} \mid \boldsymbol{Y}, \mathcal{U}) + l_{\mathcal{U}}(\beta, \phi_{\boldsymbol{R}}, \phi_{\boldsymbol{G}} \mid \mathcal{U}) \\
&= -\frac{1}{2}\left[c + \log|\boldsymbol{R}| + \log|\boldsymbol{G}| + \boldsymbol{\mathcal{E}}^T \boldsymbol{R}^{-1} \mathcal{E} + \boldsymbol{\mathcal{U}}^T \boldsymbol{G}^{-1} \mathcal{U}\right]
\end{aligned}
$$

where $\boldsymbol{\mathcal{E}} = \boldsymbol{Y} - \boldsymbol{X}\beta - \boldsymbol{Z}\boldsymbol{\mathcal{U}}$

2. Constraining individuals' curves

Constrained random effects, such that individual curves are monotone, may be specified by the probability model;

- a conditional normal distribution for $\boldsymbol{\mathcal{Y}}$ (as before)

$$(\boldsymbol{\mathcal{Y}} \,|\, \boldsymbol{\mathcal{U}}_T = \boldsymbol{U}) \sim \mathcal{N}\left(\boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{U}, \boldsymbol{R}\right)$$

- and a truncated multivariate normal distribution for $\boldsymbol{\mathcal{U}}$

$$\boldsymbol{\mathcal{U}}_T \sim \mathcal{N}_{T(\beta)}\left(\mathbf{0}, \boldsymbol{G}\right)$$

where $T(\beta) \subseteq \mathbb{R}^{rg}$

2. Constraining individuals' curves

The general pseudo-log-likelihood becomes:

$$l_{\mathcal{Y},\mathcal{U}_T}(\beta, \phi_{\boldsymbol{R}}, \phi_{\boldsymbol{G}} \mid \boldsymbol{Y}, \mathcal{U}) = l_{\mathcal{Y},\mathcal{U}}(\beta, \phi_{\boldsymbol{R}}, \phi_{\boldsymbol{G}} \mid \boldsymbol{Y}, \mathcal{U}) - \eta(\beta)$$

Where $\eta(\beta)$ is the normalising term;

$$\eta(\beta) = \log\left(\int_{T(\beta)} \left((2\pi)^{rg}|\boldsymbol{G}|\right)^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{W}^T\boldsymbol{G}^{-1}\boldsymbol{W}\right\} \mathrm{d}\boldsymbol{W}\right)$$

2. Constraining individuals' curves

When $r = 2$ we have,

$$T(\beta) = \Big\{\mathcal{U}_T = \left[u_{0,1}\, u_{1,1} \cdots\, u_{0,g}\, u_{1,g}\right]^T \in \mathbb{R}^{2g}$$
$$\text{s.t. } u_{i,1} \geq -c(\beta), i = 1, 2, \ldots, g\Big\}$$

which we incorporate into the expectation step.