Mixed models, array methods and multidimensional density estimation

Maria Durbán    Iain Currie    Paul Eilers

Universidad Carlos III de Madrid, Spain
Heriot-Watt University, Scotland
Leiden University, The Netherlands

July 2006
Why mixed model approach?

What we know so far:
- How to estimate multidimensional densities.
- How to do it with P-splines
- How to do it efficiently

So, why bother with mixed models?
- They offer a unified approach
- Smoothing can be included easily in complex models: random effects, correlated data, etc
- You might feel “at home” with them
Why mixed model approach?

What we know so far:
- How to estimate multidimensional densities.
- How to do it with P-splines
- How to do it efficiently

So, why bother with mixed models?
- They offer a unified approach
- Smoothing can be included easily in complex models: random effects, correlated data, etc
- You might feel “at home” with them
P-splines as mixed models

Equivalence:
- P-splines for Gaussian data ⇒ LMM
- P-splines for non-Gaussian data ⇒ GLMM

The LMM/GLMM approach consists of two steps:
1. Reparameterizing the linear predictor
2. Estimating the model parameters

What are the benefits of reparameterizing the model?
- Solves the problems with the identifiability of the model in the case of additive models
- Allows the decomposition of a surface into additive an interaction components ⇒ hierarchical approach
- Clarifies the role of the penalty in a multidimensional setting.
- Immediate connection with mixed model approach
P-splines as mixed models

Equivalence:
- P-splines for Gaussian data $\Rightarrow$ LMM
- P-splines for non-Gaussian data $\Rightarrow$ GLMM

The LMM/GLMM approach consists of two steps:
1. Reparameterizing the linear predictor
2. Estimating the model parameters

What are the benefits of reparameterizing the model?
- Solves the problems with the identifiability of the model in the case of additive models
- Allows the decomposition of a surface into additive an interaction components $\Rightarrow$ hierarchical approach
- Clarifies the role of the penalty in a multidimensional setting
- Immediate connection with mixed model approach
P-splines as mixed models

Equivalence:

- P-splines for Gaussian data $\Rightarrow$ LMM
- P-splines for non-Gaussian data $\Rightarrow$ GLMM

The LMM/GLMM approach consists of two steps:
1. Reparameterizing the linear predictor
2. Estimating the model parameters

What are the benefits of reparameterizing the model?

- Solves the problems with the identifiability of the model in the case of additive models
- Allows the decomposition of a surface into additive an interaction components $\Rightarrow$ hierarchical approach
- Clarifies the role of the penalty in a multidimensional setting.
- Immediate connection with mixed model approach
Reparameterizing the linear predictor

1-D

We want to reparameterize \( y = B a + \epsilon, \epsilon \sim N(0, \sigma^2 I) \)

The smoothness is imposed via the penalty matrix \( P = D'D \)

\( P \) is rank deficient \( \Rightarrow \) we look for a one-to-one transformation of the coefficients:

\[
a = T \begin{bmatrix} \beta \\ \alpha \end{bmatrix}
\]

- \( \beta \) corresponds to the part of the smooth function not penalized by \( P \)
- \( \alpha \) is orthogonal to \( \beta \) and is penalized by \( P \)

\( T \) is not unique, we use the s.v.d of the penalty to construct it:
Reparameterizing the linear predictor

1-D

We want to reparameterize $y = Ba + \epsilon$, $\epsilon \sim N(0, \sigma^2 I)$

The smoothness is imposed via the penalty matrix $P = D' D$

$P$ is rank deficient $\Rightarrow$ we look for a one-to-one transformation of the coefficients:

$$a = T \begin{bmatrix} \beta \\ \alpha \end{bmatrix}$$

- $\beta$ corresponds to the part of the smooth function not penalized by $P$
- $\alpha$ is orthogonal to $\beta$ and is penalized by $P$

$T$ is not unique, we use the s.v.d of the penalty to construct it:
Reparameterizing the linear predictor

1-D

We want to reparameterize $y = Ba + \epsilon$, $\epsilon \sim N(0, \sigma^2 I)$

The smoothness is imposed via the penalty matrix $P = D' D$

$P$ is rank deficient $\Rightarrow$ we look for a one-to-one transformation of the coefficients:

$$a = T \begin{bmatrix} \beta \\ \alpha \end{bmatrix}$$

- $\beta$ corresponds to the part of the smooth function not penalized by $P$
- $\alpha$ is orthogonal to $\beta$ and is penalized by $P$

$T$ is not unique, we use the s.v.d of the penalty to construct it:
Reparameterizing the linear predictor

1-D

We want to reparameterize \( y = B\alpha + \epsilon, \epsilon \sim N(0, \sigma^2 I) \)

The smoothness is imposed via the penalty matrix \( P = D' D \)

\( P \) is rank deficient \( \Rightarrow \) we look for a one-to-one transformation \( T \) of the coefficients:

\[
\mathbf{a} = T \begin{bmatrix} \beta \\ \alpha \end{bmatrix}
\]

- \( \beta \) correspond to the part of the smooth function not penalized by \( P \)
- \( \alpha \) is orthogonal to \( \beta \) and is penalized by \( P \)

\( T \) is not unique, we use the s.v.d of the penalty to construct it:

\[
D' D = U_s \bar{\Sigma} U_n^T = \begin{bmatrix} 0 & \bar{\Sigma} \\ \Sigma \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \Sigma \end{bmatrix}
\]
Reparameterizing the linear predictor

1-D

We want to reparameterize $y = B a + \epsilon, \epsilon \sim N(0, \sigma^2 I)$

The smoothness is imposed via the penalty matrix $P = D' D$

$P$ is rank deficient $\Rightarrow$ we look for a one-to-one transformation of the coefficients:

$$a = T \begin{bmatrix} \beta \\ \alpha \end{bmatrix}$$

- $\beta$ correspond to the part of the smooth function not penalized by $P$
- $\alpha$ is orthogonal to $\beta$ and is penalized by $P$

$T$ is not unique, we use the s.v.d of the penalty to construct it:

$$T = [U_n : U_s] \Rightarrow \beta = U'_n a \quad \alpha = U'_s a$$
Reparameterizing the linear predictor

1-D

\[ U_n' P U_n = 0 \Rightarrow a' Pa = \alpha' \begin{bmatrix} \tilde{\Sigma} \\ \text{diagonal} \end{bmatrix} \alpha \]

\[ Ba = BT \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = X\beta + Z\alpha \]

**Penalized Log-likelihood**

\[ y = Ba + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I) \]

\[ (y - Ba)'(y - Ba) + \lambda a' Pa \]

**Log-likelihood mixed model**

\[ (y - X\beta - Z\alpha)'(y - X\beta - Z\alpha) + \lambda \alpha' \tilde{\Sigma} \alpha \]

\[ y = X\beta + Z\alpha + \epsilon, \quad \alpha \sim N(0, \sigma^2 \tilde{\Sigma}^{-1}), \quad \epsilon \sim N(0, \sigma^2 I) \]

\[ \lambda = \frac{\sigma^2}{\sigma^2_{\alpha}} \]
Reparameterizing the linear predictor

1-D

- \( U'_n P U_n = 0 \Rightarrow a' Pa = \alpha' \begin{bmatrix} \beta \\ \alpha \end{bmatrix} \)
  
  - \( U'_n P U_n = 0 \Rightarrow a' Pa = \alpha' \begin{bmatrix} \beta \\ \alpha \end{bmatrix} \)
    
    \( \begin{bmatrix} \beta \\ \alpha \end{bmatrix} \)

- \( Ba = BT \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = X\beta + Z\alpha \)

Penalized Log-likelihood

\[ y = Ba + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I) \]

\[ (y - Ba)'(y - Ba) + \lambda a' Pa \]

Log-likelihood mixed model

\[ y = X\beta + Z\alpha + \epsilon, \quad \alpha \sim N(0, \sigma^2_{\alpha} \tilde{\Sigma}^{-1}), \quad \epsilon \sim N(0, \sigma^2 I) \]

\[ (y - X\beta - Z\alpha)'(y - X\beta - Z\alpha) + \lambda \alpha' \tilde{\Sigma} \alpha \]

\[ y = X\beta + Z\alpha + \epsilon, \quad \alpha \sim N(0, \sigma^2_{\alpha} \tilde{\Sigma}^{-1}), \quad \epsilon \sim N(0, \sigma^2 I) \]

\[ \lambda = \frac{\sigma^2}{\sigma^2_{\alpha}} \]
Reparameterizing the linear predictor

1-D

- \( U_n' P U_n = 0 \Rightarrow a' P a = \alpha' \begin{pmatrix} \quad \bar{\Sigma} \quad \end{pmatrix} \alpha \) diagonal
- \( B a = B T \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = X \beta + Z \alpha \)

Penalized Log-likelihood

\[
y = B a + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I)
\]

\[
(y - B a)'(y - B a) + \lambda a' P a
\]

Log-likelihood mixed model

\[
(y - X \beta - Z \alpha)'(y - X \beta - Z \alpha) + \lambda \alpha' \bar{\Sigma} \alpha
\]

\[
y = X \beta + Z \alpha + \epsilon, \quad \alpha \sim N(0, \sigma^2 \bar{\Sigma}^{-1}), \ \epsilon \sim N(0, \sigma^2 I)
\]

\[
\lambda = \frac{\sigma^2}{\sigma^2_\alpha}
\]
Reparameterizing the linear predictor

2-D

Now \( y = B \theta + \epsilon, \epsilon \sim N(0, \sigma^2 I), B = B_2 \otimes B_1 \)

The penalty matrix \( P = \lambda_1 I_{c_2} \otimes D_1' D_1 + \lambda_2 D_2' D_2 \otimes I_{c_1} \)

Currie, Durbán and Eilers (2006): Extension of the 1-D case computing the s.v.d of \( P \).

What is new?

- s.v.d of \( P \) as a function of the s.v.d. of the marginal penalties.
- Allows decomposition of the surface as the sum of marginal smooth functions plus an interaction term.
- Clarifies the role of the penalty for each component of the model
- Allows the use of the fast algorithms presented earlier
- No need to impose further constraints to achieve identifiability
Reparameterizing the linear predictor

2-D

Now $y = B\theta + \epsilon$, $\epsilon \sim N(0, \sigma^2 I)$, $B = B_2 \otimes B_1$

The penalty matrix $P = \lambda_1 I_{c_2} \otimes D_1' D_1 + \lambda_2 D_2' D_2 \otimes I_{c_1}$


What is new?

- s.v.d of $P$ as a function of the s.v.d. of the marginal penalties.

- Allows decomposition of the surface as the sum of marginal smooth functions plus an interaction term.

- Clarifies the role of the penalty for each component of the model

- Allows the use of the fast algorithms presented earlier

- No need to impose further constraints to achieve identifiability
Reparameterizing the linear predictor

2-D

Now \( y = B\theta + \epsilon, \epsilon \sim N(0, \sigma^2 I), B = B_2 \otimes B_1 \)

The penalty matrix \( P = \lambda_1 I_{c_2} \otimes D_1' D_1 + \lambda_2 D_2' D_2 \otimes I_{c_1} \)

Currie, Durbán and Eilers (2006): Extension of the 1-D case computing the s.v.d of \( P \).

What is new?

- s.v.d of \( P \) as a function of the s.v.d. of the marginal penalties.

  \[ \downarrow \]

  - Allows decomposition of the surface as the sum of marginal smooth functions plus an interaction term.
  - Clarifies the role of the penalty for each component of the model.
  - Allows the use of the fast algorithms presented earlier.
  - No need to impose further constraints to achieve identifiability.
Reparameterizing the linear predictor

2-D

Now \( y = B\theta + \epsilon, \epsilon \sim N(0, \sigma^2 I), B = B_2 \otimes B_1 \)

The penalty matrix \( P = \lambda_1 I_{c_2} \otimes D_1' D_1 + \lambda_2 D_2' D_2 \otimes I_{c_1} \)

Currie, Durbán and Eilers (2006): Extension of the 1-D case computing the s.v.d of \( P \).

What is new?

- s.v.d of \( P \) as a function of the s.v.d. of the marginal penalties.

- Allows decomposition of the surface as the sum of marginal smooth functions plus an interaction term.

- Clarifies the role of the penalty for each component of the model

- Allows the use of the fast algorithms presented earlier

- No need to impose further constraints to achieve identifiability
Reparameterizing the linear predictor

2-D

\[ D_1' D_1 = U_1 \Sigma_1 U_1' \quad D_2' D_1 = U_2 \Sigma_2 U_2' \]

Define:

\[ T = [U_{2n} \otimes U_{1n} : U_{2s} \otimes U_{1n} : U_{2n} \otimes U_{1s} : U_{2s} \otimes U_{1s}] \]

\[ X = B[U_{2n} \otimes U_{1n}] = X_2 \otimes X_1 \]

\[ Z = B[U_{2s} \otimes U_{1n} : U_{2n} \otimes U_{1s} : U_{2s} \otimes U_{1s}] = [Z_2 \otimes X_1 : X_2 \otimes Z_1 : Z_2 \otimes Z_1] \]
Reparameterizing the linear predictor

2-D

\[ D_1' D_1 = U_1 \Sigma_1 U_1' \quad D_2' D_1 = U_2 \Sigma_2 U_2' \]

Define:

\[ T = [U_{2n} \otimes U_{1n} : U_{2s} \otimes U_{1n} : U_{2n} \otimes U_{1s} : U_{2s} \otimes U_{1s}] \]

\[ X = B[U_{2n} \otimes U_{1n}] = X_2 \otimes X_1 \]

\[ Z = B[U_{2s} \otimes U_{1n} : U_{2n} \otimes U_{1s} : U_{2s} \otimes U_{1s}] \]

\[ = [Z_2 \otimes X_1 : X_2 \otimes Z_1 : Z_2 \otimes Z_1] \]
Reparameterizing the linear predictor

2-D

\[
D_1' D_1 = U_1 \Sigma_1 U_1' \quad D_2' D_1 = U_2 \Sigma_2 U_2'
\]

Define:

\[
T = [U_{2n} \otimes U_{1n} : U_{2s} \otimes U_{1n} : U_{2n} \otimes U_{1s} : U_{2s} \otimes U_{1s}]
\]

\[
X = B[U_{2n} \otimes U_{1n}] = X_2 \otimes X_1
\]

\[
Z = B[U_{2s} \otimes U_{1n} : U_{2n} \otimes U_{1s} : U_{2s} \otimes U_{1s}]
= [Z_2 \otimes X_1 : X_2 \otimes Z_1 : Z_2 \otimes Z_1]
\]
Reparameterizing the linear predictor

2-D

\[ D'_1 D_1 = U_1 \Sigma_1 U'_1 \quad D'_2 D_1 = U_2 \Sigma_2 U'_2 \]

Define:

\[ T = [U_{2n} \otimes U_{1n} : U_{2s} \otimes U_{1n} : U_{2n} \otimes U_{1s} : U_{2s} \otimes U_{1s}] \]

\[ X = B[U_{2n} \otimes U_{1n}] = X_2 \otimes X_1 \]

\[ Z = B [U_{2s} \otimes U_{1n} : U_{2n} \otimes U_{1s} : U_{2s} \otimes U_{1s}] \]

\[ = [Z_2 \otimes X_1 : X_2 \otimes Z_1 : Z_2 \otimes Z_1] \]

- Full interaction
- Linear by Non-linear interaction
Reparameterizing the linear predictor

2-D

The penalty with the new parametrization is:

\[ F = T' PT = \left( \begin{array}{ccc} \lambda_2 \Sigma_{2s} \otimes I_{q_1} \\ \lambda_1 I_{q_2} \otimes \Sigma_{1s} \\ \lambda_1 I_{c_2 - q_2} \otimes \Sigma_{1s} + \lambda_2 \Sigma_{2s} \otimes I_{c_1 - q_1} \end{array} \right) \]

⇓

Penalized-Spline ANOVA decomposition: \( f(x) + f(y) + f(x, y) \)

Now, the mixed model representation is:

\[ y = X\beta + Z\alpha + \epsilon, \quad \alpha \sim N(0, \sigma^2 F^{-1}), \quad \epsilon \sim N(0, \sigma^2 I) \]
Reparameterizing the linear predictor

2-D

The penalty with the new parametrization is:

\[
F = T'PT = \begin{pmatrix}
\lambda_2 \Sigma_2 \otimes I_{q_1} \\
\lambda_1 I_{p_2} \otimes \Sigma_1 \\
\lambda_1 I_{q_2 - q_2} \otimes \Sigma_1 + \lambda_2 \Sigma_2 \otimes I_{c_1 - q_1}
\end{pmatrix}
\]

Penalized-Spline ANOVA decomposition: \( f(x) + f(y) + f(x, y) \)

Now, the mixed model representation is:

\[
y = X\beta + Z\alpha + \epsilon, \quad \alpha \sim N(0, \sigma^2 F^{-1}), \quad \epsilon \sim N(0, \sigma^2 I)
\]
Reparameterizing the linear predictor

2-D

The penalty with the new parametrization is:

\[ F = T' PT = \begin{pmatrix} \lambda_2 \Sigma_2 s \otimes I_{q_1} \\ \lambda_1 I_{q_2} \otimes \Sigma_1 s \\ \lambda_1 I_{c_{2-q_2}} \otimes \Sigma_1 s + \lambda_2 \Sigma_2 s \otimes I_{c_{1-q_1}} \end{pmatrix} \]

\[ \Downarrow \]

Penalized-Spline ANOVA decomposition: \( f(x) + f(y) + f(x, y) \)

Now, the mixed model representation is:

\[ y = X\beta + Z\alpha + \epsilon, \, \alpha \sim N(0, \sigma^2 F^{-1}), \, \epsilon \sim N(0, \sigma^2 I) \]
Reparameterizing the linear predictor

2-D

The penalty with the new parametrization is:

\[ F = T' PT = \begin{pmatrix} \lambda_2 \Sigma_2 s \otimes I_{q_1} \\ \lambda_1 I_{q_2} \otimes \Sigma_1 s \\ \lambda_1 I_{c_2 - q_2} \otimes \Sigma_1 s + \lambda_2 \Sigma_2 s \otimes I_{c_1 - q_1} \end{pmatrix} \]

Allowing for different smoothing parameters in the interaction:

Makes possible to compare nested models
Reparameterizing the linear predictor

Example: Data on a 30 grid

- Data simulated on a $30 \times 20$ grid.
- $B_1$, $30 \times 13$, and $B_2$, $20 \times 10$

\[
X = (1 : x_2 \otimes 1_{n_1} : 1_{n_2} \otimes x_1 : x_2 \otimes x_1)
\]
\[
Z = (Z_2 \otimes 1_{n_1} : Z_2 \otimes x_1 : 1_{n_2} \otimes Z_1 : x_2 \otimes Z_1 : Z_2 \otimes Z_1).
\]

This partition facilitates two things:

1. The fitted surface is the sum of an overall mean plus the sum of three terms: a term for $x_2$, a term for $x_1$ and an interaction term.
2. We can fit submodels $\Rightarrow$ hierarchical approach can lead to model selection and model simplification.
Reparameterizing the linear predictor

Example: Data on a 30 grid

- Data simulated on a $30 \times 20$ grid.
- $B_1$, $30 \times 13$, and $B_2$, $20 \times 10$

\[
X = (1 : x_2 \otimes 1_{n_1} : 1_{n_2} \otimes x_1 : x_2 \otimes x_1)
\]

\[
Z = (Z_2 \otimes 1_{n_1} : Z_2 \otimes x_1 : 1_{n_2} \otimes Z_1 : x_2 \otimes Z_1 : Z_2 \otimes Z_1).
\]

This partition facilitates two things:

1. The fitted surface is the sum of an overall mean plus the sum of three terms: a term for $x_2$, a term for $x_1$ and an interaction term.
2. We can fit submodels $\Rightarrow$ hierarchical approach can lead to model selection and model simplification.
Reparameterizing the linear predictor

Example: Data on a 30 grid

- Data simulated on a $30 \times 20$ grid.
- $B_1$, $30 \times 13$, and $B_2$, $20 \times 10$

\[
X = (1 : x_2 \otimes 1_{n_1} : 1_{n_2} \otimes x_1 : x_2 \otimes x_1)
\]
\[
Z = (Z_2 \otimes 1_{n_1} : Z_2 \otimes x_1 : 1_{n_2} \otimes Z_1 : x_2 \otimes Z_1 : Z_2 \otimes Z_1).
\]

This partition facilitates two things:

1. The fitted surface is the sum of an overall mean plus the sum of three terms: a term for $x_2$, a term for $x_1$ and an interaction term.

2. We can fit submodels $\Rightarrow$ hierarchical approach can lead to model selection and model simplification.
Reparameterizing the linear predictor

Example: Data on a 30 grid

- Data simulated on a $30 \times 20$ grid.
- $B_1$, $30 \times 13$, and $B_2$, $20 \times 10$

$$X = (1 : x_2 \otimes 1_{n_1} : 1_{n_2} \otimes x_1 : x_2 \otimes x_1)$$

$$Z = (Z_2 \otimes 1_{n_1} : Z_2 \otimes x_1 : 1_{n_2} \otimes Z_1 : x_2 \otimes Z_1 : Z_2 \otimes Z_1).$$

This partition facilitates two things:

1. The fitted surface is the sum of an overall mean plus the sum of three terms: a term for $x_2$, a term for $x_1$ and an interaction term.
2. We can fit submodels $\Rightarrow$ hierarchical approach can lead to model selection and model simplification.
Reparameterizing the linear predictor

Data

Term for $X_1$

Term for $X_2$

Interaction term

Fitted surface
Estimating the model parameters

- Estimates of $\beta$ and $\alpha$ follow from standard mixed model theory:

\[
\hat{\beta} = (X' V^{-1} X)^{-1} X' V^{-1} y \\
\hat{\alpha} = GZ' V^{-1} (y - X\hat{\beta})
\]

- $V = \sigma^2 I + ZGZ'$

- Smoothing parameters may be selected by maximizing the residual log likelihood (REML) $\ell(\lambda_1, \lambda_2)$

\[-\frac{1}{2} \log |V| - \frac{1}{2} \log |X' V^{-1} X| - \frac{1}{2} y' (V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1}) y.\]

- In the case of GLMM $\implies$ penalized quasi-likelihood (PQL) approach of Breslow & Clayton (1993).

Now

\[V = W_\delta^{-1} + ZGZ'\]

the weights $W_\delta$ are given by $\text{diag}(\exp(X\beta + Z\alpha))$ in the Poisson case.

Evaluation

Key point: Kronecker product structure of $X$ and $Z$
Estimating the model parameters

- Estimates of $\beta$ and $\alpha$ follow from standard mixed model theory:
  \[
  \hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y \\
  \hat{\alpha} = GZ'V^{-1}(y - X\hat{\beta})
  \]

- $V = \sigma^2I + ZGZ'$

- Smoothing parameters may be selected by maximizing the residual log likelihood (REML) $\ell(\lambda_1, \lambda_2)$
  \[
  -\frac{1}{2} \log |V| - \frac{1}{2} \log |X'V^{-1}X| - \frac{1}{2} y'(V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1})y.
  \]

- In the case of GLMM $\Longrightarrow$ penalized quasi-likelihood (PQL) approach of Breslow & Clayton (1993).

  Now
  \[
  V = W_{\delta}^{-1} + ZGZ'
  \]

  the weights $W_{\delta}$ are given by $\text{diag}(\exp(X\beta + Z\alpha))$ in the Poisson case.

Evaluation

Key point: Kronecker product structure of $X$ and $Z$
Estimating the model parameters

- Estimates of $\beta$ and $\alpha$ follow from standard mixed model theory:
  \[
  \hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y \\
  \hat{\alpha} = GZ'V^{-1}(y - X\hat{\beta})
  \]
  
  \[V = \sigma^2I + ZGZ'\]

- Smoothing parameters may be selected by maximizing the residual log likelihood (REML) $\ell(\lambda_1, \lambda_2)$
  \[-\frac{1}{2} \log |V| - \frac{1}{2} \log |X'V^{-1}X| - \frac{1}{2} y'(V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1})y.\]

- In the case of GLMM $\Rightarrow$ penalized quasi-likelihood (PQL) approach of Breslow & Clayton (1993).
  Now
  \[V = W_\delta^{-1} + ZGZ'\]
  
  The weights $W_\delta$ are given by $\text{diag}(\exp(X\beta + Z\alpha))$ in the Poisson case.

Evaluation

Key point: Kronecker product structure of $X$ and $Z$
Estimating the model parameters

- Estimates of $\beta$ and $\alpha$ follow from standard mixed model theory:
  \[
  \hat{\beta} = (X' V^{-1} X)^{-1} X' V^{-1} y
  \]
  \[
  \hat{\alpha} = GZ' V^{-1} (y - X\hat{\beta})
  \]

- $V = \sigma^2 I + ZGZ'$

- Smoothing parameters may be selected by maximizing the residual log likelihood (REML) $\ell(\lambda_1, \lambda_2)$
  \[\frac{1}{2} \log |V| - \frac{1}{2} \log |X' V^{-1} X| - \frac{1}{2} y' (V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1}) y.\]

- In the case of GLMM $\implies$ penalized quasi-likelihood (PQL) approach of Breslow & Clayton (1993).

  Now
  \[V = W_\delta^{-1} + ZGZ',\]
  the weights $W_\delta$ are given by $\text{diag}(\exp(X\beta + Z\alpha))$ in the Poisson case.

**Evaluation**

Key point: Kronecker product structure of $X$ and $Z$
Example

Old Faithful

- Construct a two-dimensional histogram over a fine grid; we used $100 \times 100$ in this example
- $B$-spline bases $B_1$ and $B_2$ each of rank 13 $\Rightarrow B$ of size $10^4 \times 169$.
- Methods not using our array approach are quite challenging (25 minutes)
- With array methods: 40 seconds
Example

Isotropic: $\lambda = 0.0311$

Anisotropic: $\lambda_1 = 0.00663$, $\lambda_2 = 0.3088$

Isotropic: $ED = 26.27$, $AIC = 1342.68$

Anisotropic: $ED = 22.73$, $AIC = 1334.69$

Maria Durbán, Iain Currie, Paul Eilers ( )
Example

Isotropic: $\lambda=0.0311$

Anisotropic: $\lambda_1=0.00663$, $\lambda_2=0.3088$

Isotropic: ED=26.27, AIC=1342.68
Anisotropic: ED=22.73, AIC=1334.69
Conclusions

- New basis
  - Allows mixed model methods to be used to fit a $d$-dimensional surface
  - Enables the fitted surface to be decomposed as a sum of additive and interaction terms $\Rightarrow$ Hierarchical approach
  - No problem with identifiability

- Mixed models give a fast and compact method of 2-dimensional density smoothing.

But...

If you don’t like mixed models: we still have the bayesian approach
Conclusions

- New basis
  - Allows mixed model methods to be used to fit a $d$-dimensional surface
  - Enables the fitted surface to be decomposed as a sum of additive and interaction terms $\Rightarrow$ Hierarchical approach
  - No problem with identifiability
- Mixed models give a fast and compact method of 2-dimensional density smoothing.

But...

If you don’t like mixed models: we still have the bayesian approach
Conclusions

- New basis
  - Allows mixed model methods to be used to fit a $d$-dimensional surface
  - Enables the fitted surface to be decomposed as a sum of additive and interaction terms $\Rightarrow$ Hierarchical approach
  - No problem with identifiability
- Mixed models give a fast and compact method of 2-dimensional density smoothing.

But...

If you don’t like mixed models: we still have the bayesian approach