# Models more general than two-variance models

It would be great if this analysis could be done for all mixed linear models.

Alas, it can't.

I'll show this for 2NRCAR models.

H2013 shows it can be done for some big classes of models.

- I'll list these and spare you the details, except for one.

Two expedients may help for models with RLs that can't be simplified.

- Each involves ignoring some part of the RL.
- This makes them approximate or sloppy, as you prefer.
- They appear to provide useful information and might suggest ways to extend or supersede the approach used for two-variance models.

# You can't always diagonalize the RL for 2NRCAR models

Suppose $\mathbf{y} = \boldsymbol{\delta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma_e^2 \mathbf{I}_n)$, and $\boldsymbol{\delta}$ has a 2NRCAR density:

$$f(\boldsymbol{\delta}|\sigma_{s1}^2, \sigma_{s2}^2) \propto \exp\left[-\frac{1}{2}\boldsymbol{\delta}'\left(\mathbf{Q}_1/\sigma_{s1}^2 + \mathbf{Q}_2/\sigma_{s2}^2\right)\boldsymbol{\delta}\right];$$

- $\mathbf{Q}_k = (q_{ij,k}), k = 1, 2$ encodes class-$k$ neighbor pairs
- $\sigma_{sk}^2$ controls similarity induced by class-$k$ neighbor pairs.

From Newcomb (1961) & Graybill (1983, Theorem 12.2.12):

- $\exists$ non-singular $\mathbf{B} \ni \mathbf{Q}_k = \mathbf{B}'\mathbf{D}_k\mathbf{B}$, where
- $\mathbf{D}_k$ is diagonal with non-negative diagonal elements.
- AND $\mathbf{B}$ is orthogonal $\Leftrightarrow$ $\mathbf{Q}_1\mathbf{Q}_2$ is symmetric.

$\mathbf{y} = \boldsymbol{\delta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma_e^2 \mathbf{I}_n)$ and $\boldsymbol{\delta}$ has a 2NRCAR density:

$$f(\boldsymbol{\delta} | \sigma_{s1}^2, \sigma_{s2}^2) \propto \exp\left[-\frac{1}{2}\boldsymbol{\delta}' \left(\mathbf{Q}_1/\sigma_{s1}^2 + \mathbf{Q}_2/\sigma_{s2}^2\right)\boldsymbol{\delta}\right];$$

$\exists$ non-singular $\mathbf{B} \ni \mathbf{Q}_k = \mathbf{B}'\mathbf{D}_k\mathbf{B}$

- $\mathbf{D}_k$ has $I_k$ zero diagonal elements; $I_k = \#\mathbf{Q}_k$'s zero eigenvalues.
- $I_k \geq I = \#(\mathbf{Q}_1 + \mathbf{Q}_2)$'s zero eigenvalues $= \#$ zero diagonal entries in $\mathbf{D}_1 + \mathbf{D}_2$.

Define $\mathbf{D}_{k+}$ = upper-left $(n - I) \times (n - I)$ submatrix of $\mathbf{D}_k$

Let $d_{kj}, j = 1, \ldots, n - I$, be the diagonal elements of $\mathbf{D}_{k+}$.

Finally, define the precisions $\tau_e = 1/\sigma_e^2$, $\tau_1 = 1/\sigma_{s1}^2$, and $\tau_2 = 1/\sigma_{s2}^2$.

Then the joint posterior for $(\boldsymbol{\delta}, \tau_e, \tau_1, \tau_2) \propto$

$$\pi(\tau_e, \tau_1, \tau_2) \qquad |\tau_e \mathbf{I}_n|^{0.5} \quad \exp\left(-0.5\tau_e(\mathbf{y}-\boldsymbol{\delta})'(\mathbf{y}-\boldsymbol{\delta})\right)$$
$$\prod_{j=1}^{n-I} (d_{1j}\tau_1 + d_{2j}\tau_2)^{0.5} \exp\left(-0.5\boldsymbol{\delta}'(\tau_1\mathbf{Q}_1 + \tau_2\mathbf{Q}_2)\boldsymbol{\delta}\right).$$

$\prod_{j=1}^{n-I} (d_{1j}\tau_1 + d_{2j}\tau_2)$ is the determinant of $\tau_1\mathbf{D}_{1+} + \tau_2\mathbf{D}_{2+}$

Integrate out $\boldsymbol{\delta}$ to give the RL $\propto$:

$$|\tau_e \mathbf{I}_n|^{0.5} \quad \prod_{j=1}^{n-I} (d_{1j}\tau_1 + d_{2j}\tau_2)^{0.5} \quad |\mathbf{H}|^{-0.5} \exp\left(-0.5\tau_e(\mathbf{y}'\mathbf{y} - \tau_e\mathbf{y}'\mathbf{H}^{-1}\mathbf{y})\right)$$

where $\mathbf{H} = \tau_e\mathbf{I}_n + \tau_1\mathbf{Q}_1 + \tau_2\mathbf{Q}_2$. We need to diagonalize $\mathbf{H}$.

We need to diagonalize

$$\begin{aligned} \mathbf{H} &= \tau_e \mathbf{I}_n + \tau_1 \mathbf{Q}_1 + \tau_2 \mathbf{Q}_2 \\ &= \tau_e \mathbf{I}_n + \mathbf{B}'(\tau_1 \mathbf{D}_1 + \tau_2 \mathbf{D}_2)\mathbf{B} \\ &= \mathbf{B}' \left[ \tau_e (\mathbf{B}\mathbf{B}')^{-1} + \tau_1 \mathbf{D}_1 + \tau_2 \mathbf{D}_2 \right] \mathbf{B} \end{aligned}$$

It's enough to diagonalize the part inside the square brackets.

This happens $\Leftrightarrow$ $\mathbf{B}$ is $\perp$ $\Leftrightarrow$ $\mathbf{Q}_1\mathbf{Q}_2$ is symmetric; not true in general.

The RL for the 2NRCAR model diagonalizes $\Leftrightarrow$ $\mathbf{Q}_1\mathbf{Q}_2$ is symmetric.

- Examples: Row-and-column grids; spatiotemporal 2NRCAR.
- Counterexamples: Periodontal 2NRCARs.

# The RL does diagonalize for some big classes of models

Balanced designs:

- For any design that satisfies the conditions of general balance and is also an orthogonal design, the RL diagonalizes.
- This includes everything you think of as a balanced ANOVA.

Separable models:

- Model $\mathbf{y} = \boldsymbol{\delta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma_e^2 \mathbf{I}_n)$
- $\boldsymbol{\delta} \sim$ normal mean $\mathbf{0}$, precision $\sum_{k=1}^{K} \tau_k \mathbf{Q}_k$ for $\mathbf{Q}_k$ $n \times n$, $\tau_k$ scalar.
- $\mathbf{Q}_k = \mathbf{A}_1 \otimes \cdots \otimes \mathbf{A}_M$ with $\mathbf{A}_l = \mathbf{I}$ for $l \neq k$, $\mathbf{A}_k$ is psd.
- This includes diagonalizable kNRCAR models, among others.
- This isn't general balance; RLs are not gamma GLM likelihoods.

Miscellaneous others:

- Clustering + heterogeneity model.
- Certain extensions of separable models.

# Approximating the GP gives a diagonalizable RL

Idea:

- Use a spectral (Fourier) approximation to a GP observed on a rectangular grid.
- This gives a model to which our tools are easily extended.
- Applying them to the approximation is of interest because . . .
- We can develop hypotheses about GPs & test them with simulations.
- Some propose using the approximation instead of the GP.

My development is built on Paciorek (2007).

We'll consider only the one-dimensional case.

Example: Global-mean surface temp series with $n = 128$ observations

Model the $n$-vector $\mathbf{y}$, for $n = 2^k$, as

$$\begin{aligned} y_t &= \beta_0 + 2 \sum_{m=1}^{\frac{n}{2}-1} \left[ u_{1m} \cos(\omega_m 2\pi t/n) - u_{2m} \sin(\omega_m 2\pi t/n) \right] \\ &\quad + u_{1,n/2} \cos(\omega_{n/2} 2\pi t/n) + \epsilon, \end{aligned}$$

where

- $\omega_m = 1, 2, \ldots, n/2$, so the $\omega_m$ are equally-spaced frequencies.
- The observations are equally spaced ($t$ is an integer).
- $\epsilon \sim$ iid $N(0, \sigma_e^2)$.

We'll represent this as a mixed linear model.

- The intercept is the only fixed effect.
- The $u$'s have dimension $(n-1)$ and diagonal covariance matrix $\mathbf{G}$.
- With the right choice of $\mathbf{G}$, this approximates the GP's covariance.

$$
\begin{aligned}
y_t &= \beta_0 + 2\sum_{m=1}^{\frac{n}{2}-1} [u_{1m}\cos(\omega_m 2\pi t/n) - u_{2m}\sin(\omega_m 2\pi t/n)] \\
&\quad + u_{1,n/2}\cos(\omega_{n/2} 2\pi t/n) + \boldsymbol{\epsilon},
\end{aligned}
$$

The RE design matrix $\mathbf{Z}$ is $n \times (n-1)$:

- The first $n-2$ columns are cos/sin pairs with common frequency $\omega_m 2\pi$ but different $u_{1m}$ and $u_{2m}$.
- The last column is unpaired, with frequency $\omega_{n/2} 2\pi$.
- $\mathbf{Z}'\mathbf{Z} = n \text{ diag}(2, 2, \ldots, 2, 1)$ and $\mathbf{1}_n'\mathbf{Z} = \mathbf{0}$.

$\mathbf{G} = \text{cov}(\mathbf{u})$ is diagonal:

- $\text{var}(u_{1,n/2}) = \sigma_s^2 \ \phi(\omega_{n/2}; \boldsymbol{\theta})$ while
- $\text{var}(u_{1m}) = \text{var}(u_{2m}) = 0.5 \ \sigma_s^2 \ \phi(\omega_m; \boldsymbol{\theta})$.

where $\sigma_s^2 \ \phi(\omega; \boldsymbol{\theta})$ is the spectral density of the GP's covariance function for unknown parameters $\boldsymbol{\theta}$.

The approximate model is

$$\mathbf{y} = \mathbf{1}_n\beta + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon} \quad \text{for } \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma_e^2\mathbf{I}_n) \text{ and } \mathbf{u} \sim N(\mathbf{0}, \sigma_s^2\mathbf{D}(\boldsymbol{\theta}))$$

where

- $\mathbf{D}(\boldsymbol{\theta})$ is diagonal with diag elements $q\ \sigma_s^2\ \phi(\omega_m; \boldsymbol{\theta})$, $q = 0.5$ or 1
- $\mathbf{1}_n'\mathbf{Z} = \mathbf{0}$ and $\mathbf{Z}$'s columns are orthogonal.

Pre-multiply by $(\mathbf{Z}'\mathbf{Z})^{-0.5}\mathbf{Z}'$ to give

$$\begin{aligned} \hat{\mathbf{v}} \equiv (\mathbf{Z}'\mathbf{Z})^{-0.5}\mathbf{Z}'\mathbf{y} &= (\mathbf{Z}'\mathbf{Z})^{0.5}\mathbf{u} + (\mathbf{Z}'\mathbf{Z})^{-0.5}\mathbf{Z}'\ \boldsymbol{\epsilon} \\ &\equiv \mathbf{v} + \boldsymbol{\xi} \end{aligned}$$

where $\boldsymbol{\xi}$ and $\mathbf{v}$ are $(n-1) \times 1$ with $\text{cov}(\boldsymbol{\xi}) = \sigma_e^2\mathbf{I}_{n-1}$ and

$$\begin{aligned} \text{cov}(\mathbf{v}) &= \sigma_s^2\ (\mathbf{Z}'\mathbf{Z})^{0.5}\mathbf{D}(\boldsymbol{\theta})(\mathbf{Z}'\mathbf{Z})^{0.5} \\ &= \sigma_s^2\ \text{diag}(\ n\ \phi(\ \omega_{m(j)}; \boldsymbol{\theta})\ ) \\ \text{for } m(j) &= \left\{ \begin{array}{l} j/2 \text{ for even } j \\ (j+1)/2 \text{ for odd } j \end{array} \right\} j = 1, \ldots, n-1. \end{aligned}$$

So we have

$$\begin{aligned}
\hat{\mathbf{v}} &= (\mathbf{Z}'\mathbf{Z})^{-0.5}\mathbf{Z}'\mathbf{y} = \mathbf{v} + \boldsymbol{\xi} \\
\text{cov}(\boldsymbol{\xi}) &= \sigma_e^2 \ \mathbf{I}_{n-1} \\
\text{cov}(\mathbf{v}) &= \sigma_s^2 \ \text{diag}( \ n \ \phi(\omega_{m(j)}; \boldsymbol{\theta}) \ )
\end{aligned}$$

The restricted likelihood for this approximate model is the likelihood for $(\sigma_e^2, \sigma_s^2, \boldsymbol{\theta})$ from this model:

$$\begin{aligned}
\log RL(\sigma_e^2, \sigma_s^2, \boldsymbol{\theta}) &= K - \frac{1}{2} \sum_{j=1}^{n-1} \left[ \log(\sigma_s^2 a_j(\boldsymbol{\theta}) + \sigma_e^2) + \hat{v}_j^2 / (\sigma_s^2 a_j(\boldsymbol{\theta}) + \sigma_e^2) \right] \\
\text{for } a_j(\boldsymbol{\theta}) &= n \ \phi(\omega_{m(j)}; \boldsymbol{\theta}), \quad j = 1, \ldots, n-1
\end{aligned}$$

for $K$ an unimportant constant.

This has the desired simple form; $a_j(\boldsymbol{\theta})$ is non-increasing in $j$ and a function of the unknown $\boldsymbol{\theta}$.

# Why this is potentially cool

$$\begin{aligned} \log RL(\sigma_e^2, \sigma_s^2, \boldsymbol{\theta}) &= K - \frac{1}{2}\sum_{j=1}^{n-1}\left[\log(\sigma_s^2 a_j(\boldsymbol{\theta}) + \sigma_e^2) + \hat{v}_j^2/(\sigma_s^2 a_j(\boldsymbol{\theta}) + \sigma_e^2)\right] \\ \text{for } a_j(\boldsymbol{\theta}) &= n\,\phi(\omega_{m(j)};\boldsymbol{\theta}), \quad j = 1, \ldots, n-1 \\ \text{and } \hat{\mathbf{v}} &= (\mathbf{Z}'\mathbf{Z})^{-0.5}\mathbf{Z}'\mathbf{y}. \end{aligned}$$

For a given dataset, **Z** is the same for all GPs, so

- $\hat{\mathbf{v}}$ (the transformed data) is the same for all GPs
- alternative GPs differ only in their $a_j(\boldsymbol{\theta})$.

Given $\boldsymbol{\theta}$, this is a gamma GLM with $E(\hat{v}_j^2) = \sigma_s^2 a_j(\boldsymbol{\theta}) + \sigma_e^2$, so:

- $\hat{\sigma}_e^2$ is "the middle" of the $\hat{v}_j^2$ for large $j$;
- $\hat{\sigma}_s^2\ a_j(\boldsymbol{\theta})$ then fits the decline with $j$ of the $\hat{v}_j^2$ for small $j$.

# Two expedients for problems that don't diagonalize

Expedient #1: Ignore the error variance

- Example: 2NRCAR puzzle.
- Brian Reich did this work *before* developing the re-expressed RL; this inspired the re-expression.
- I'll show you this.

Expedient #2: Ignore the non-zero off-diagonals

- Example: Optical imaging (DLM) puzzle.
- I won't show you this – it's in H2013, Sec. 17.2.2.

# Expedient #1: Ignore the error variance – 2NRCAR models

$\mathbf{y} = \boldsymbol{\delta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma_e^2 \mathbf{I}_n)$ and $\boldsymbol{\delta}$ has a 2NRCAR density:

$$f(\boldsymbol{\delta} | \sigma_{s1}^2, \sigma_{s2}^2) \propto \exp \left[ -\frac{1}{2} \boldsymbol{\delta}' \left( \mathbf{Q}_1 / \sigma_{s1}^2 + \mathbf{Q}_2 / \sigma_{s2}^2 \right) \boldsymbol{\delta} \right];$$

$\exists$ non-singular $\mathbf{B} \ni \mathbf{Q}_k = \mathbf{B}' \mathbf{D}_k \mathbf{B}$

- $\mathbf{D}_k$ has $I_k$ zero diagonal elements; $I_k = \# \mathbf{Q}_k$'s zero eigenvalues.
- $I_k \geq I = \#(\mathbf{Q}_1 + \mathbf{Q}_2)$'s zero eigenvalues $= \#$ zero diagonal entries in $\mathbf{D}_1 + \mathbf{D}_2$.

Let $d_{kj}$ be $\mathbf{D}_k$'s diagonals, for $j \ni d_{1j} > 0$ or $d_{2j} > 0$, $j = 1, \ldots, n - I$.

Finally, define the precisions $\tau_e = 1/\sigma_e^2$, $\tau_1 = 1/\sigma_{s1}^2$, and $\tau_2 = 1/\sigma_{s2}^2$.

$\mathbf{y} = \boldsymbol{\delta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma_e^2 \mathbf{I}_n)$ and $\boldsymbol{\delta}$ has a 2NRCAR density

Then the joint posterior for $(\boldsymbol{\delta}, \tau_e, \tau_1, \tau_2) \propto$

$$\begin{aligned} \pi(\tau_e, \tau_1, \tau_2) \qquad & |\tau_e \mathbf{I}_n|^{0.5} \quad \exp\left(-0.5\tau_e(\mathbf{y} - \boldsymbol{\delta})'(\mathbf{y} - \boldsymbol{\delta})\right) \\ & \prod_{j=1}^{n-I} (d_{1j}\tau_1 + d_{2j}\tau_2)^{0.5} \exp\left(-0.5\boldsymbol{\delta}'(\tau_1 \mathbf{Q}_1 + \tau_2 \mathbf{Q}_2)\boldsymbol{\delta}\right). \end{aligned}$$

Expedient #1: Condition on $(\tau_e, \boldsymbol{\delta})$ and ignore $\pi(\tau_e, \tau_1, \tau_2)$, so

$$\begin{aligned} \pi(\tau_1, \tau_2 | \boldsymbol{\delta}) & \propto \prod_{j=1}^{n-I} (d_{1j}\tau_1 + d_{2j}\tau_2)^{0.5} \exp\left(-0.5\boldsymbol{\delta}'\mathbf{B}'(\tau_1 \mathbf{D}_1 + \tau_2 \mathbf{D}_2)\mathbf{B}\boldsymbol{\delta}\right) \\ & = \prod_{j=1}^{n-I} (d_{1j}\tau_1 + d_{2j}\tau_2)^{0.5} \exp\left(-0.5 \sum_{j=1}^{n-I} u_j^2 \left(d_{1j}\tau_1 + d_{2j}\tau_2\right)\right), \end{aligned}$$

for $\mathbf{u} = \mathbf{B}\boldsymbol{\delta}$.

$$\pi(\tau_1, \tau_2|\boldsymbol{\delta}) \quad \propto \quad \prod_{j=1}^{n-I} (d_{1j}\tau_1 + d_{2j}\tau_2)^{0.5} \exp\left(-0.5 \sum_{j=1}^{n-I} u_j^2 \,(d_{1j}\tau_1 + d_{2j}\tau_2)\right)$$

for $\mathbf{u} = \mathbf{B}\boldsymbol{\delta}$; $\mathbf{B}$ is known, we're conditioning on the unknown $\boldsymbol{\delta}$.

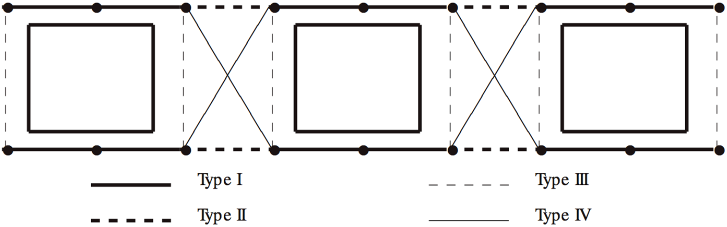
This looks like the re-expressed RL for a 2-variance model, but:

- In a 2-variance model, $\sigma_e^2$ can have free terms but $\sigma_s^2$ can't.
- Here, however, $\tau_1$ and $\tau_2$ are symmetric.
- So consider free and mixed terms for both classes of neighbor pairs.

Definitions:

- **u**-free terms for $\tau_1$ have $d_{2j} = 0$; $\exists$ $I_2 - I$ of them.
- **u**-free terms for $\tau_2$ have $d_{1j} = 0$, $\exists$ $I_1 - I$ of them.
- **u**-mixed terms have $d_{1j} > 0$ and $d_{2j} > 0$; $\exists$ $n - I_1 - I_2 + I$ of them.

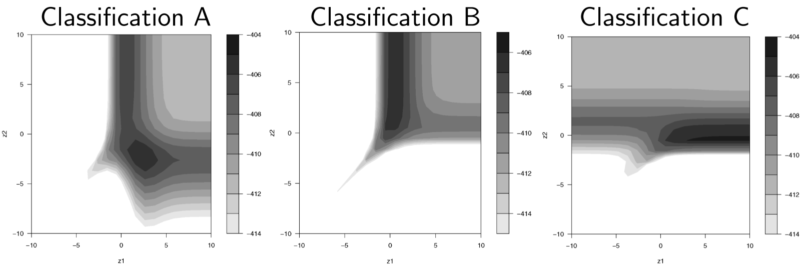
Recall: Periodontal neighbor pairs are of four distinct types.




Consider three ways of defining 2 classes of neighbor pairs:

| Classification | Type of nbr pair I | II | III | IV | Description |
|---|---|---|---|---|---|
| A | 1 | 1 | 2 | 2 | Sides vs. interproximal |
| B | 1 | 2 | 2 | 2 | Direct vs. interproximal |
| C | 2 | 1 | 2 | 2 | Type II vs. others |

Contours of one person's marginal posterior of $(z_1, z_2)$, $z_k = \log(\tau_k/\tau_1)$.

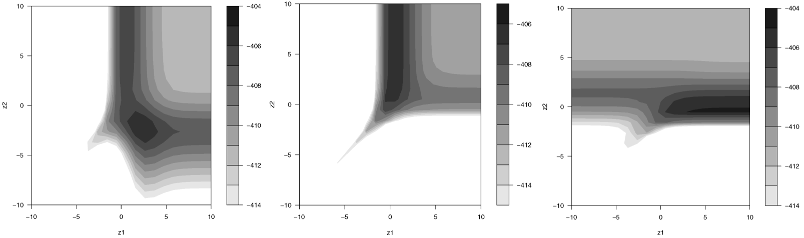



The "legs" arise from **u**-free terms (intuition on the next slide).

| THIS IS WRONG | | # of islands | | | Counts of **u**-free terms | | **u**-mixed terms |
|---|---|---|---|---|---|---|---|
| Classification | $n$ | $I$ | $I_1$ | $I_2$ | for $z_1$ | for $z_2$ | |
| A | 162 | 3 | 6 | 84 | 81 | 3 | 75 |
| B | 162 | 3 | 54 | 84 | 81 | 51 | 27 |
| C | 162 | 3 | 114 | 3 | 0 | 111 | 48 |

Contours of one person's marginal posterior of $(z_1, z_2)$, $z_k = \log(\tau_k/\tau_1)$.

Classification A　　Classification B　　Classification C




The "legs" arise from **u**-free terms (intuition on the next slide).

| CORRECTED | | | | | Counts of | | |
|---|---|---|---|---|---|---|---|
| | | | # of islands | | **u**-free terms | | **u**-mixed |
| Classif'n | $n$ | $I$ | $I_1$ | $I_2$ | for $z_1$ | for $z_2$ | terms |
| A | 162 | 3 | 6 | ~~84~~ 30 | ~~81~~ 27 | 3 | ~~75~~ 129 |
| B | 162 | 3 | 54 | ~~84~~ 30 | ~~81~~ 27 | 51 | ~~27~~ 81 |
| C | 162 | 3 | ~~114~~ 48 | 3 | 0 | ~~111~~ 45 | ~~48~~ 114 |

Reich et al (2007, Section 4) presents the argument.

Intuition:

- The **u**-free terms for $z_1$ considered alone have contours parallel to the $z_2$ axis. This gives the vertical "leg".
- The **u**-free terms for $z_2$ considered alone have contours parallel to the $z_1$ axis. This gives the horizontal "leg".
- Classification C's contour plot has only one "leg" parallel to the $z_1$ axis because it has no **u**-free terms for $z_1$.

So why don't the legs go down & to the left? Explained in H2013, p. 385.