Structured Shrinkage Priors

Maryclare Griffin

University of Massachusetts Amherst
Department of Mathematics and Statistics

August 5, 2025

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2024, VOL. 33, NO. 1, 1–14
https://doi.org/10.1080/10618600.2023.2233577

Taylor & Francis
Taylor & Francis Group

Structured Shrinkage Priors

Maryclare Griffin[a] and Peter D. Hoff[b]

[a]Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA; [b]Department of Statistical Science, Duke University, Durham, NC

ABSTRACT
In many regression settings the unknown coefficients may have some known structure, for instance they may be ordered in space or correspond to a vectorized matrix or tensor. At the same time, the unknown coefficients may be sparse, with many nearly or exactly equal to zero. However, many commonly used priors and corresponding penalties for coefficients do not encourage simultaneously structured and sparse estimates. In this article we develop structured shrinkage priors that generalize multivariate normal, Laplace, exponential power and normal-gamma priors. These priors allow the regression coefficients to be correlated a priori without sacrificing elementwise sparsity or shrinkage. The primary challenges in working with these structured shrinkage priors are computational, as the corresponding penalties are intractable integrals and the full conditional distributions that are needed to approximate the posterior mode or simulate from the posterior distribution may be nonstandard. We overcome these issues using a flexible elliptical slice sampling procedure, and demonstrate that these priors can be used to introduce structure while preserving sparsity. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received November 2021
Accepted June 2023

KEYWORDS
Bayesian lasso; Lasso; Multivariate Laplace; Multivariate normal-scale mixture; Shrinkage; Sparsity

Supported by NSF Grants DGE-1256082, DMS-1505136.

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2024, VOL. 33, NO. 1, 1–14
https://doi.org/10.1080/10618600.2023.2233577

Taylor & Francis
Taylor & Francis Group

Check for updates

Structured Shrinkage Priors

Maryclare Griffin[a] and Peter D. Hoff[b]

[a]Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA; [b]Department of Statistical Science, Duke University, Durham, NC

ABSTRACT
In many regression settings the unknown coefficients may have some known structure, for instance they may be ordered in space or correspond to a vectorized matrix or tensor. At the same time, the unknown coefficients may be sparse, with many nearly or exactly equal to zero. However, many commonly used priors and corresponding penalties for coefficients do not encourage simultaneously structured and sparse estimates. In this article we develop structured shrinkage priors that generalize multivariate normal, Laplace, exponential power and normal-gamma priors. These priors allow the regression coefficients to be correlated a priori without sacrificing elementwise sparsity or shrinkage. The primary challenges in working with these structured shrinkage priors are computational, as the corresponding penalties are intractable integrals and the full conditional distributions that are needed to approximate the posterior mode or simulate from the posterior distribution may be nonstandard. We overcome these issues using a flexible elliptical slice sampling procedure, and demonstrate that these priors can be used to introduce structure while preserving sparsity. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received November 2021
Accepted June 2023

KEYWORDS
Bayesian lasso; Lasso; Multivariate Laplace; Multivariate normal-scale mixture; Shrinkage; Sparsity

Supported by NSF Grants DGE-1256082, DMS-1505136.

Model-Based Penalized Regression

Model-Based Penalized Regression

Model for data (**likelihood**): $p\left(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta},\boldsymbol{\phi}\right) \propto \exp\left\{-h\left(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta},\boldsymbol{\phi}\right)\right\}$

Model-Based Penalized Regression

$$\begin{aligned} \text{Model for data } (\textbf{likelihood})\text{:} \quad & p\left(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta},\boldsymbol{\phi}\right) \propto \exp\left\{-h\left(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta},\boldsymbol{\phi}\right)\right\} \\ \text{Model for } \boldsymbol{\beta} \text{ } (\textbf{prior})\text{:} \quad & p\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right) \propto \exp\left\{-f\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right)\right\} \end{aligned}$$

Model-Based Penalized Regression

$$\begin{aligned}\text{Model for data } (\textbf{likelihood})\text{:} \quad & p\left(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta},\boldsymbol{\phi}\right) \propto \exp\left\{-h\left(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta},\boldsymbol{\phi}\right)\right\} \\ \text{Model for } \boldsymbol{\beta} \text{ } (\textbf{prior})\text{:} \quad & p\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right) \propto \exp\left\{-f\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right)\right\}\end{aligned}$$

- Choice of **likelihood** guided by type of data observed
- Choice of **prior** guided by $\underbrace{\text{scientific context}}_{\text{ideal ☺}}$ or $\underbrace{\text{convenience}}_{\text{realistic ☹}}$

Model-Based Penalized Regression

$$\begin{aligned} \text{Model for data (}\textbf{likelihood}\text{):} \quad & p\left(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta},\boldsymbol{\phi}\right) \propto \exp\left\{-h\left(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta},\boldsymbol{\phi}\right)\right\} \\ \text{Model for } \boldsymbol{\beta} \text{ (}\textbf{prior}\text{):} \quad & p\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right) \propto \exp\left\{-f\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right)\right\} \end{aligned}$$

- Choice of **likelihood** guided by type of data observed
- Choice of **prior** guided by $\underbrace{\text{scientific context}}_{\text{ideal ☺}}$ or $\underbrace{\text{convenience}}_{\text{realistic ☹}}$

Estimate $\boldsymbol{\beta}$ according to posterior $p\left(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\phi},\boldsymbol{\theta}\right)$, can use

Model-Based Penalized Regression

$$\begin{aligned} \text{Model for data } (\textbf{likelihood})\text{:} \quad & p\left(\boldsymbol{y}|\boldsymbol{X\beta},\boldsymbol{\phi}\right) \propto \exp\left\{-h\left(\boldsymbol{y}|\boldsymbol{X\beta},\boldsymbol{\phi}\right)\right\} \\ \text{Model for } \boldsymbol{\beta} \text{ } (\textbf{prior})\text{:} \quad & p\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right) \propto \exp\left\{-f\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right)\right\} \end{aligned}$$

- Choice of **likelihood** guided by type of data observed
- Choice of **prior** guided by $\underbrace{\text{scientific context}}_{\text{ideal ☺}}$ or $\underbrace{\text{convenience}}_{\text{realistic ☹}}$

Estimate $\boldsymbol{\beta}$ according to posterior $p\left(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\phi},\boldsymbol{\theta}\right)$, can use

- Posterior mode, solves $\min_{\boldsymbol{\beta}} h\left(\boldsymbol{y}|\boldsymbol{X\beta},\boldsymbol{\phi}\right) + f\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right)$
  $\rightarrow$ Penalized estimate
- Posterior mean, $\mathbb{E}\left[\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\phi},\boldsymbol{\theta}\right]$

Model-Based Penalized Regression

$$\begin{aligned} \text{Model for data (\textbf{likelihood}):} \quad & p\left(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\phi}\right) \propto \exp\left\{-h\left(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\phi}\right)\right\} \\ \text{Model for } \boldsymbol{\beta} \text{ (\textbf{prior}):} \quad & p\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right) \propto \exp\left\{-f\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right)\right\} \end{aligned}$$

- Choice of **likelihood** guided by type of data observed
- Choice of **prior** guided by $\underbrace{\text{scientific context}}_{\text{ideal ☺}}$ or $\underbrace{\text{convenience}}_{\text{realistic ☹}}$

Estimate $\boldsymbol{\beta}$ according to posterior $p\left(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\phi}, \boldsymbol{\theta}\right)$, can use

- Posterior mode, solves $\min_{\boldsymbol{\beta}} h\left(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\phi}\right) + f\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right)$
  $\rightarrow$ Penalized estimate
- Posterior mean, $\mathbb{E}\left[\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\phi}, \boldsymbol{\theta}\right]$

What do we mean by "Structured Shrinkage Priors?"

What do we mean by "Structured Shrinkage Priors?"

Most sparsity inducing priors assume **independence**

$\rightarrow$ **Separable** penalties, $f\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right) = \sum_{j=1}^{p} f_j\left(\beta_j|\boldsymbol{\theta}\right)$

What do we mean by “Structured Shrinkage Priors?”

Most sparsity inducing priors assume **independence**

$\rightarrow$ **Separable** penalties, $f(\boldsymbol{\beta}|\boldsymbol{\theta}) = \sum_{j=1}^{p} f_j(\beta_j|\boldsymbol{\theta})$

In many settings, $\boldsymbol{\beta}$ may be **structured**

- Treatment effects at different spatial locations or over time
- Autoregressive parameters of different orders
- Main effects and higher order interactions

An Example Motivating Structured Shrinkage

Forney et al. (2013); Bruneel (2018)

An Example Motivating Structured Shrinkage

Subject wears cap of **EEG** sensors

- Different spatial locations
- High temporal resolution

Forney et al. (2013); Bruneel (2018)

An Example Motivating Structured Shrinkage

Subject wears cap of **EEG** sensors

- Different spatial locations
- High temporal resolution

EEG measurements collected over many **trials**

- Grid of randomly selected letters shown each trial
- Subject counts the number of times a target letter appears

Forney et al. (2013); Bruneel (2018)

An Example Motivating Structured Shrinkage

Data: For trials $i = 1, \ldots, 240$

- $y_i \in \{0, 1\}$, presence of target
- $\boldsymbol{x}_i$ EEG measurements during trial i

An Example Motivating Structured Shrinkage

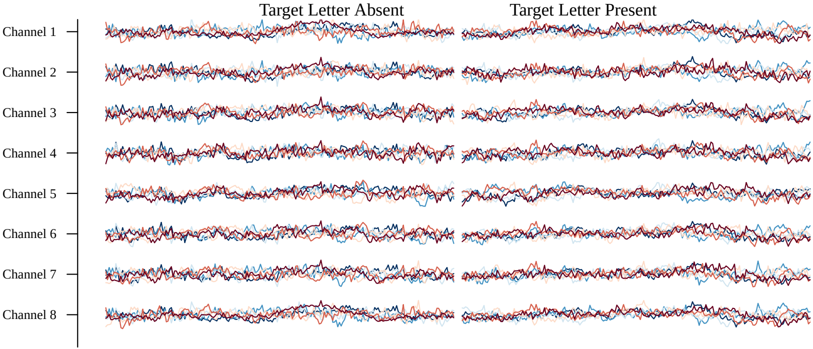
Data: For trials $i = 1, \ldots, 240$

- $y_i \in \{0, 1\}$, presence of target
- $\boldsymbol{x}_i$ EEG measurements during trial i

An Example Motivating Structured Shrinkage

Data: For trials $i = 1, \dots, 240$

- $y_i \in \{0, 1\}$, presence of target
- $\boldsymbol{x}_i$ EEG measurements during trial i



An Example Motivating Structured Shrinkage

An Example Motivating Structured Shrinkage

Given scientific context, we expect

An Example Motivating Structured Shrinkage

Given scientific context, we expect

- Only some channels and time points are relevant

  $\rightarrow$ Exactly or nearly **sparse** $\boldsymbol{\beta}$

An Example Motivating Structured Shrinkage

Given scientific context, we expect

- Only some channels and time points are relevant
  - $\rightarrow$ Exactly or nearly **sparse** $\boldsymbol{\beta}$
- Similarity across nearby channels, consecutive time points
  - $\rightarrow$ **Structured** $\boldsymbol{\beta}$

An Example Motivating Structured Shrinkage

Given scientific context, we expect

- Only some channels and time points are relevant
  - $\rightarrow$ Exactly or nearly **sparse** $\boldsymbol{\beta}$
- Similarity across nearby channels, consecutive time points
  - $\rightarrow$ **Structured** $\boldsymbol{\beta}$

Structured shrinkage priors may yield:

- Better estimation of $\boldsymbol{\beta}$
- Improved out-of-sample prediction

Our Approach to Structured Shrinkage

Many **independent** priors are normal scale mixtures

$$\boldsymbol{\beta} \overset{d}{=} \boldsymbol{s} \circ \boldsymbol{z},\ \boldsymbol{z} \sim \text{normal}\left(\boldsymbol{0}, \boldsymbol{I}_p\right),\ \boldsymbol{s} \sim ???$$

Polson and Scott (2010); Finegold and Drton (2011); Griffin and Hoff (2024)

Our Approach to Structured Shrinkage

Many **independent** priors are normal scale mixtures

$$\boldsymbol{\beta} \stackrel{d}{=} \boldsymbol{s} \circ \boldsymbol{z},\ \boldsymbol{z} \sim \text{normal}\left(\boldsymbol{0}, \boldsymbol{I}_p\right),\ \boldsymbol{s} \sim ???$$

A natural approach to defining **dependent** priors is

$$\boldsymbol{\beta} \stackrel{d}{=} \boldsymbol{s} \circ \boldsymbol{z},\ \boldsymbol{z} \sim \text{normal}\left(\boldsymbol{0}, \boldsymbol{\Omega}\right),\ \boldsymbol{s} \sim ???$$

Polson and Scott (2010); Finegold and Drton (2011); Griffin and Hoff (2024)

Our Approach to Structured Shrinkage

Many **independent** priors are normal scale mixtures

$$\boldsymbol{\beta} \stackrel{d}{=} \boldsymbol{s} \circ \boldsymbol{z},\ \boldsymbol{z} \sim \text{normal}\left(\boldsymbol{0}, \boldsymbol{I}_p\right),\ \boldsymbol{s} \sim ???$$

A natural approach to defining **dependent** priors is

$$\boldsymbol{\beta} \stackrel{d}{=} \boldsymbol{s} \circ \boldsymbol{z},\ \boldsymbol{z} \sim \text{normal}\left(\boldsymbol{0}, \boldsymbol{\Omega}\right),\ \boldsymbol{s} \sim ???$$

Dependence introduces challenges

- Marginal prior $p\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right) = \int p\left(\boldsymbol{\beta}|\boldsymbol{s}, \boldsymbol{\theta}\right) p\left(\boldsymbol{s}|\boldsymbol{\theta}\right) d\boldsymbol{s}$ intractable
- $p\left(\boldsymbol{s}|\boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\theta}\right)$ nonstandard

Polson and Scott (2010); Finegold and Drton (2011); Griffin and Hoff (2024)

Comparison to Existing Priors

$$\boldsymbol{\beta} \stackrel{d}{=} \boldsymbol{s} \circ \boldsymbol{z},\ \boldsymbol{z} \sim \text{normal}\left(\mathbf{0}, \boldsymbol{\Omega}\right),\ \boldsymbol{s} \sim ???$$

van Gerven et al. (2009); van Gerven et al. (2010); Kyung et al. (2010); Kalli and Griffin (2014); Wu et al. (2014); Zhao et al. (2016); Kowal et al. (2017)

Comparison to Existing Priors

$$\boldsymbol{\beta} \stackrel{d}{=} \boldsymbol{s} \circ \boldsymbol{z},\ \boldsymbol{z} \sim \text{normal}\left(\boldsymbol{0}, \boldsymbol{\Omega}\right),\ \boldsymbol{s} \sim ???$$

Marginally Uncorrelated Structured Priors:

- $\boldsymbol{\Omega} \propto \boldsymbol{I}_p$ and dependent $\boldsymbol{s}^2 \rightarrow$ Uncorrelated $\boldsymbol{\beta}$ a priori
- Still challenging to implement

---

van Gerven et al. (2009); van Gerven et al. (2010); Kyung et al. (2010); Kalli and Griffin (2014); Wu et al. (2014); Zhao et al. (2016); Kowal et al. (2017)

Comparison to Existing Priors

$$\boldsymbol{\beta} \stackrel{d}{=} \boldsymbol{s} \circ \boldsymbol{z},\ \boldsymbol{z} \sim \text{normal}\left(\mathbf{0}, \boldsymbol{\Omega}\right),\ \boldsymbol{s} \sim ???$$

Marginally Uncorrelated Structured Priors:

- $\boldsymbol{\Omega} \propto \boldsymbol{I}_p$ and dependent $\boldsymbol{s}^2 \rightarrow$ Uncorrelated $\boldsymbol{\beta}$ a priori
- Still challenging to implement

Elliptically Contoured/Group Lasso:

- $\boldsymbol{s} = s\mathbf{1}_p \rightarrow$ entire $\boldsymbol{\beta}$ shares single shrinkage factor
- Don't generalize their independent counterparts

---

van Gerven et al. (2009); van Gerven et al. (2010); Kyung et al. (2010); Kalli and Griffin (2014); Wu et al. (2014); Zhao et al. (2016); Kowal et al. (2017)

Comparison to Existing Priors

$$\boldsymbol{\beta} \stackrel{d}{=} \boldsymbol{s} \circ \boldsymbol{z},\ \boldsymbol{z} \sim \mathrm{normal}\left(\mathbf{0}, \boldsymbol{\Omega}\right),\ \boldsymbol{s} \sim ???$$

Marginally Uncorrelated Structured Priors:

- $\boldsymbol{\Omega} \propto \boldsymbol{I}_p$ and dependent $\boldsymbol{s}^2 \rightarrow$ Uncorrelated $\boldsymbol{\beta}$ a priori
- Still challenging to implement

Elliptically Contoured/Group Lasso:

- $\boldsymbol{s} = s\mathbf{1}_p \rightarrow$ entire $\boldsymbol{\beta}$ shares single shrinkage factor
- Don't generalize their independent counterparts

Fused Lasso/Structured Penalty Priors:

- Elements of $\boldsymbol{\Omega}$ also random
- Prior parameters hard to interpret, relate to moments of $\boldsymbol{\beta}$

van Gerven et al. (2009); van Gerven et al. (2010); Kyung et al. (2010); Kalli and Griffin (2014); Wu et al. (2014); Zhao et al. (2016); Kowal et al. (2017)

Our Approach: Three Structured Shrinkage Priors

Structured Product Normal (**SPN**)

- $\boldsymbol{s} \sim \text{normal}\left(\mathbf{0}, \boldsymbol{\Psi}\right)$
- Quadratic log lik. in $\boldsymbol{\beta} \rightarrow p\left(\boldsymbol{s} | \boldsymbol{z}, \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{\Omega}, \boldsymbol{\Psi}\right)$ is normal
- Variance and fourth order moments of $\boldsymbol{\beta}$ determine $\boldsymbol{\Omega}$, $\boldsymbol{\Psi}$

Our Approach: Three Structured Shrinkage Priors

Structured Product Normal (**SPN**)

- $\boldsymbol{s} \sim \text{normal}\left(\mathbf{0}, \boldsymbol{\Psi}\right)$
- Quadratic log lik. in $\boldsymbol{\beta} \rightarrow p\left(\boldsymbol{s}|\boldsymbol{z}, \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{\Omega}, \boldsymbol{\Psi}\right)$ is normal
- Variance and fourth order moments of $\boldsymbol{\beta}$ determine $\boldsymbol{\Omega}$, $\boldsymbol{\Psi}$
- **Symmetric SPN** (**sSPN**)
  - Special case, requires $|\omega_{ij}| = |\psi_{ij}|$
  - Variance of $\boldsymbol{\beta}$ determines unknown elements of $\boldsymbol{\Omega}$, $\boldsymbol{\Psi}$

Our Approach: Three Structured Shrinkage Priors

Structured Product Normal (**SPN**)

- $\boldsymbol{s} \sim \text{normal}\left(\mathbf{0}, \boldsymbol{\Psi}\right)$
- Quadratic log lik. in $\boldsymbol{\beta} \rightarrow p\left(\boldsymbol{s}|\boldsymbol{z}, \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{\Omega}, \boldsymbol{\Psi}\right)$ is normal
- Variance and fourth order moments of $\boldsymbol{\beta}$ determine $\boldsymbol{\Omega}$, $\boldsymbol{\Psi}$
- **Symmetric SPN** (**sSPN**)
  - Special case, requires $|\omega_{ij}| = |\psi_{ij}|$
  - Variance of $\boldsymbol{\beta}$ determines unknown elements of $\boldsymbol{\Omega}$, $\boldsymbol{\Psi}$

Structured Normal-Gamma (**SNG**)

- $s_j^2 \overset{i.i.d.}{\sim} \text{gamma}\left(\text{shape} = c, \text{rate} = c\right)$
- c treated as fixed and known

Our Approach: Three Structured Shrinkage Priors

Structured Product Normal (SPN)

- $\boldsymbol{s} \sim \text{normal}\left(\boldsymbol{0}, \boldsymbol{\Psi}\right)$
- Quadratic log lik. in $\boldsymbol{\beta} \rightarrow p\left(\boldsymbol{s}|\boldsymbol{z}, \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{\Omega}, \boldsymbol{\Psi}\right)$ is normal
- Variance and fourth order moments of $\boldsymbol{\beta}$ determine $\boldsymbol{\Omega}$, $\boldsymbol{\Psi}$
- **Symmetric SPN (sSPN)**
  - Special case, requires $|\omega_{ij}| = |\psi_{ij}|$
  - Variance of $\boldsymbol{\beta}$ determines unknown elements of $\boldsymbol{\Omega}$, $\boldsymbol{\Psi}$

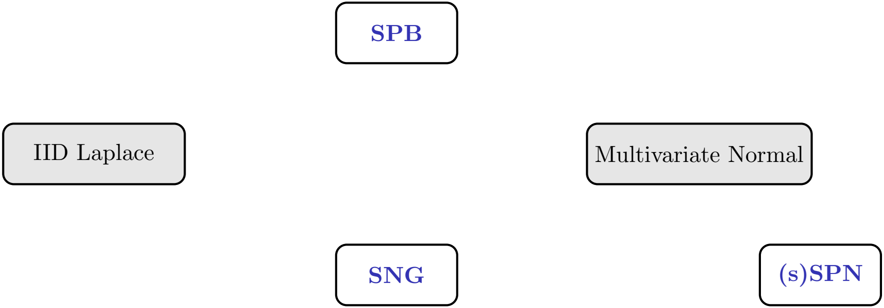
Structured Normal-Gamma (SNG)

- $s_j^2 \overset{i.i.d.}{\sim} \text{gamma}\left(\text{shape} = c, \text{rate} = c\right)$
- c treated as fixed and known

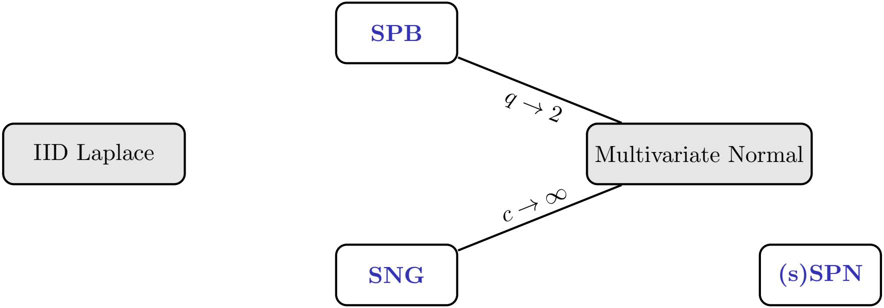
Structured Power/Bridge (SPB)

- $s_j^2 \overset{i.i.d.}{\sim}$ polynomially tilted positive $\frac{q}{2}$-stable distribution
- q treated as fixed and known

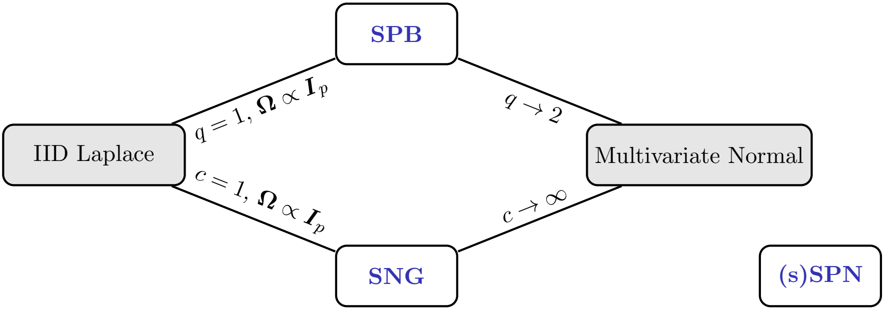
Relationships Between SPN, SNG and SPB Priors



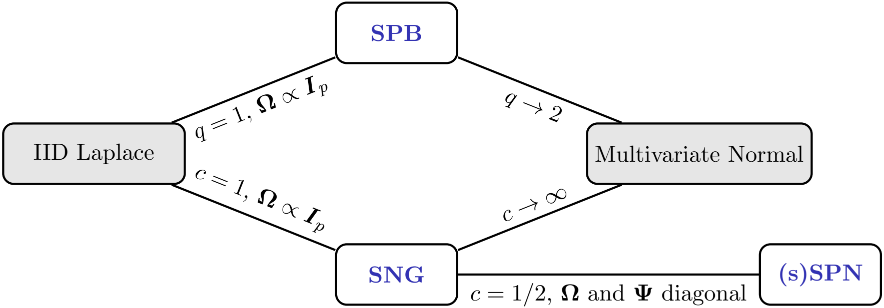
Relationships Between SPN, SNG and SPB Priors



Relationships Between SPN, SNG and SPB Priors



Relationships Between SPN, SNG and SPB Priors



Understanding Properties of SPN, SNG and SPB Priors

No closed form marginal prior, $p\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right) = \int p\left(\boldsymbol{\beta}|\boldsymbol{s}, \boldsymbol{\theta}\right) p\left(\boldsymbol{s}|\boldsymbol{\theta}\right) d\boldsymbol{s}$

Understanding Properties of SPN, SNG and SPB Priors

No closed form marginal prior, $p\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right) = \int p\left(\boldsymbol{\beta}|\boldsymbol{s},\boldsymbol{\theta}\right) p\left(\boldsymbol{s}|\boldsymbol{\theta}\right) d\boldsymbol{s}$

→ Simulate $\boldsymbol{\beta} \in \mathbb{R}^2$ with unit variance, correlation $\rho = 0.5$

- Prior density contours → **Shrinkage and Structure**

Understanding Properties of SPN, SNG and SPB Priors

No closed form marginal prior, $p\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right) = \int p\left(\boldsymbol{\beta}|\boldsymbol{s},\boldsymbol{\theta}\right) p\left(\boldsymbol{s}|\boldsymbol{\theta}\right) d\boldsymbol{s}$

$\rightarrow$ Simulate $\boldsymbol{\beta} \in \mathbb{R}^2$ with unit variance, correlation $\rho = 0.5$

- Prior density contours $\rightarrow$ **Shrinkage and Structure**

$\rightarrow$ Examine behavior of along axes, at origin $\rightarrow$ **Shrinkage**

Understanding Properties of SPN, SNG and SPB Priors

No closed form marginal prior, $p(\boldsymbol{\beta}|\boldsymbol{\theta}) = \int p(\boldsymbol{\beta}|\boldsymbol{s}, \boldsymbol{\theta})\, p(\boldsymbol{s}|\boldsymbol{\theta})\, d\boldsymbol{s}$

→ Simulate $\boldsymbol{\beta} \in \mathbb{R}^2$ with unit variance, correlation $\rho = 0.5$

- Prior density contours → **Shrinkage and Structure**

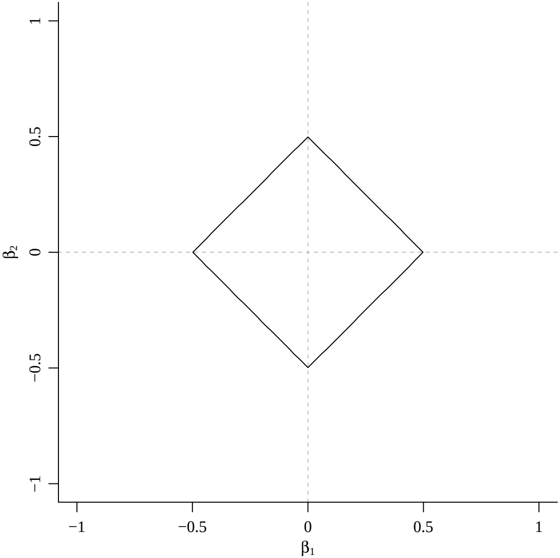
→ Examine behavior of along axes, at origin → **Shrinkage**

→ Study prior moments

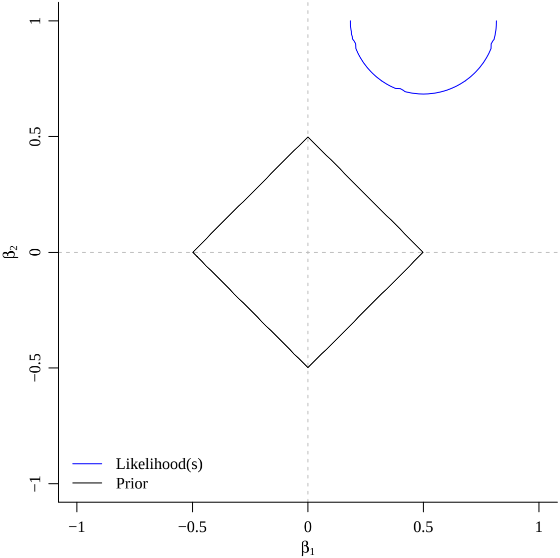
- Determine range of $\boldsymbol{\Sigma}$ → Achievable **Structure**

Property	**SNG**	**SPB**	**SPN**
Generalizes an Independent Shrinkage Prior	✓	✓	✓
Univariate Marginal is an Independent Shrinkage Prior	✓	✓	✓
Generalizes a Laplace Prior	$c = 1$	$q = 1$	
Generalizes a Normal Prior	$c \to \infty$	$q \to 2$	
Infinite Spike or Pole at Zero		$c \leq 1/2$	✓
Quadratic Scale Log Full Conditional			✓
Arbitrary Covariance Structure Achievable			✓

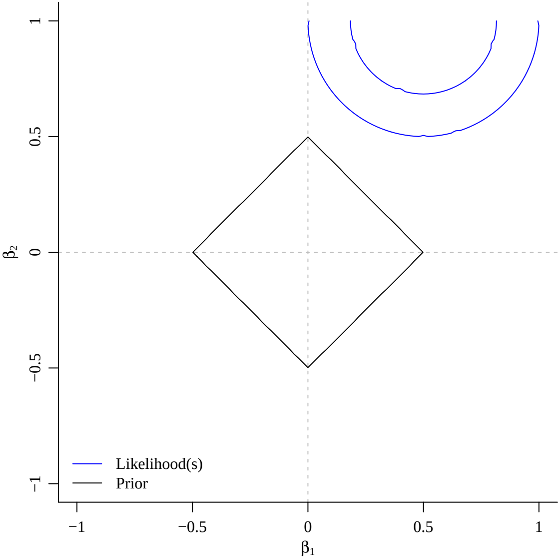
Prior Density Contours



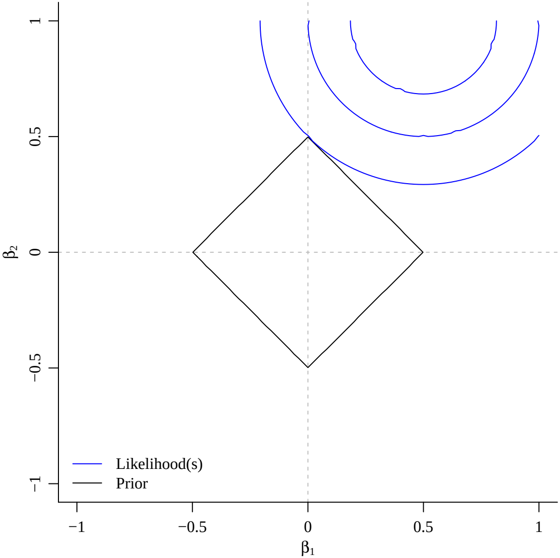
Prior Density Contours



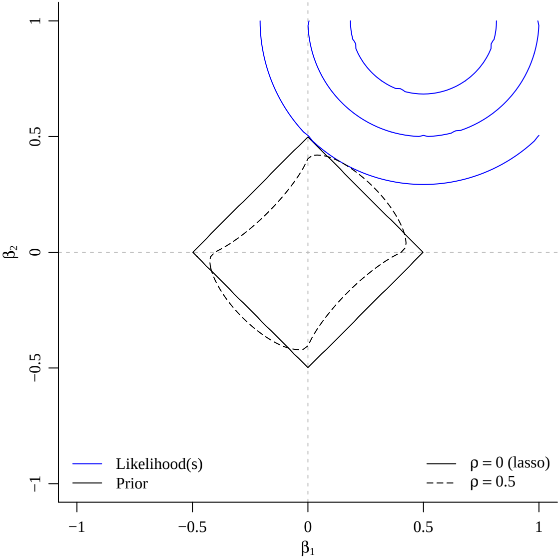
Prior Density Contours



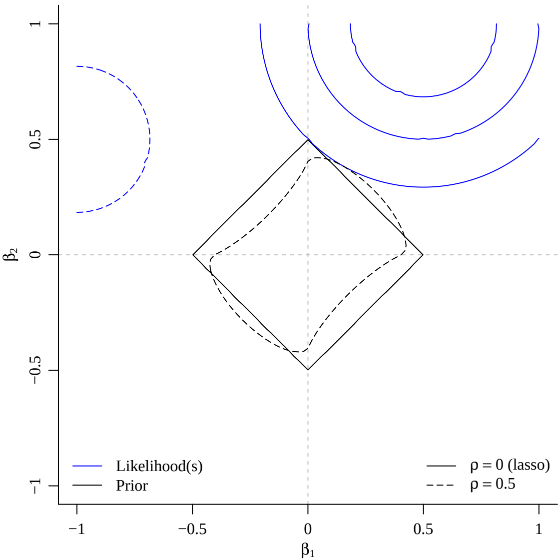
Prior Density Contours



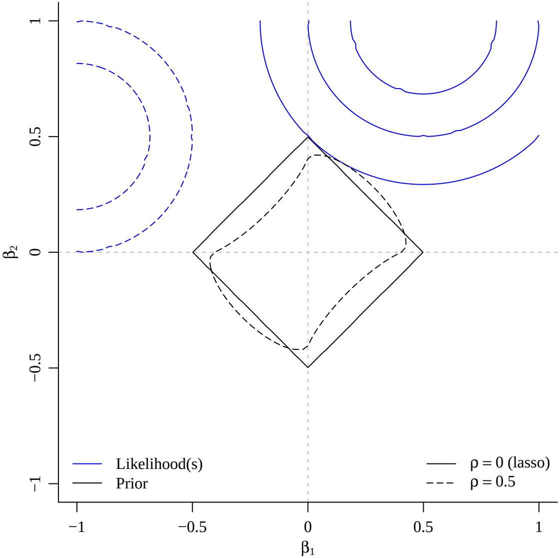
Prior Density Contours



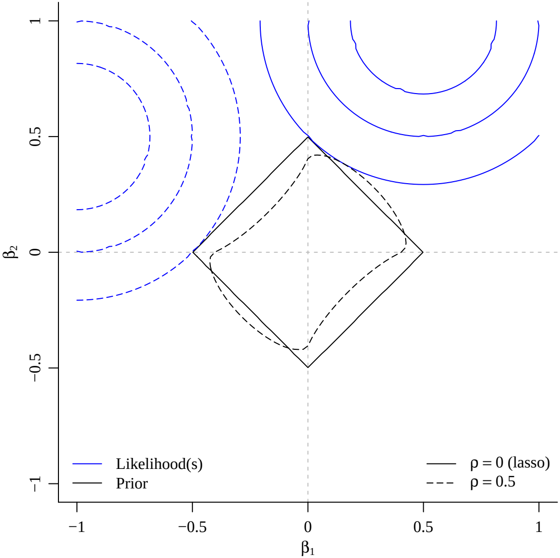
Prior Density Contours



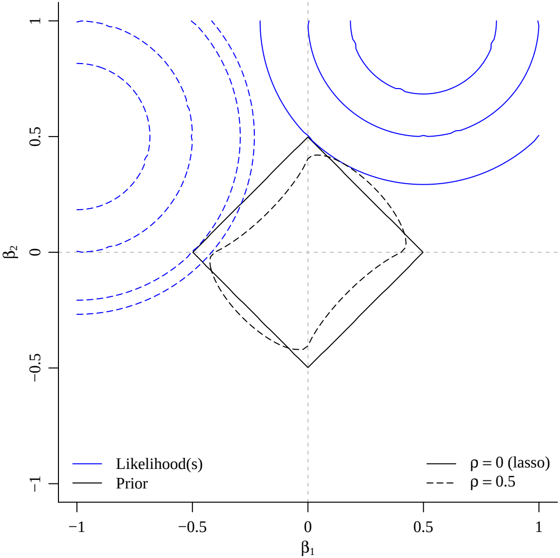
Prior Density Contours



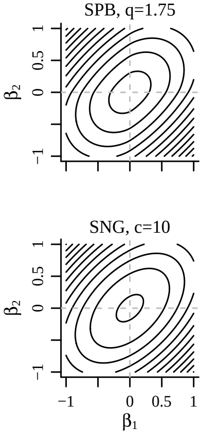
Prior Density Contours



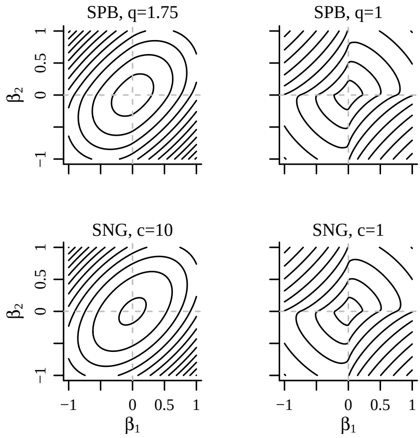
Prior Density Contours



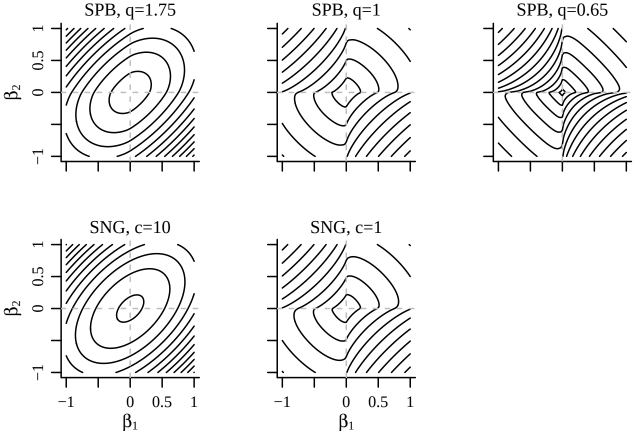
Prior Density Contours



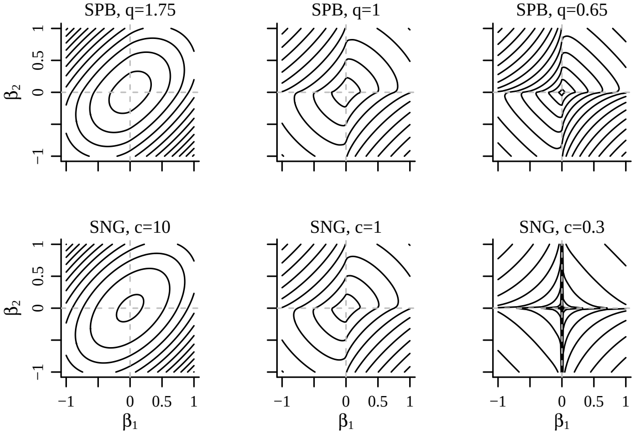
Prior Density Contours



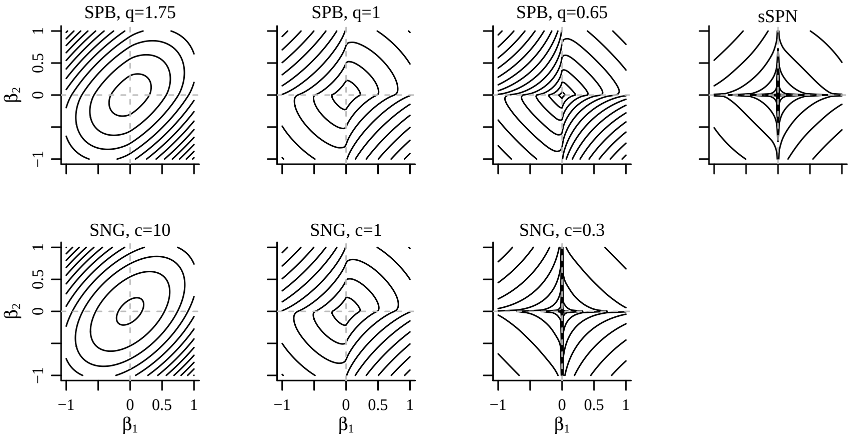
Prior Density Contours



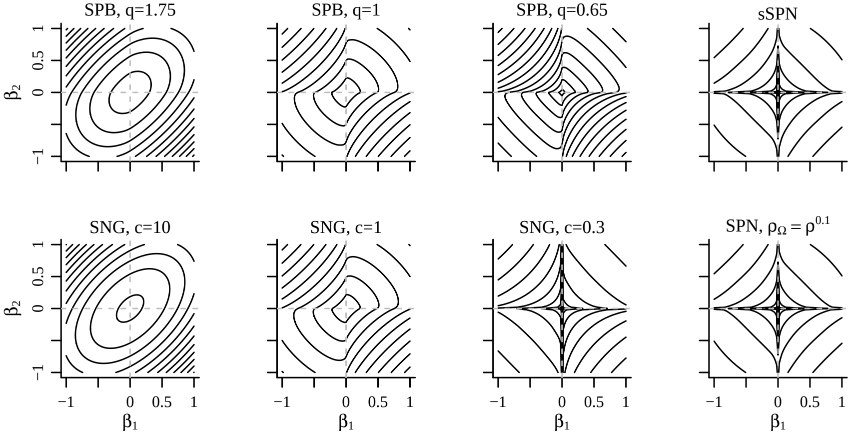
Prior Density Contours



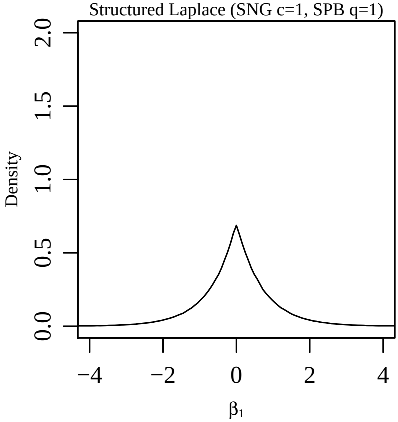
Prior Density Contours



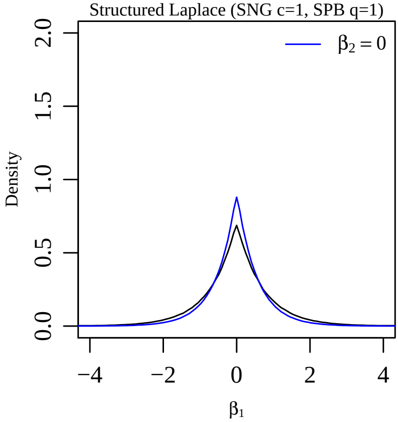
Prior Density Contours



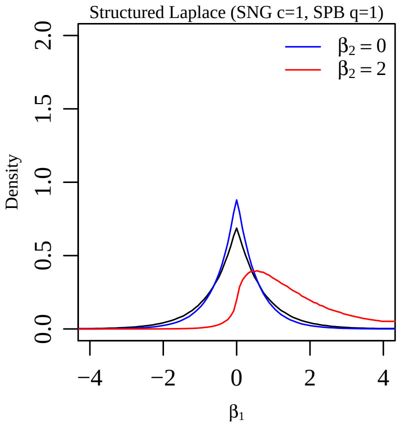
Prior Conditional Distributions



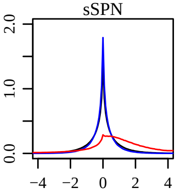
Prior Conditional Distributions



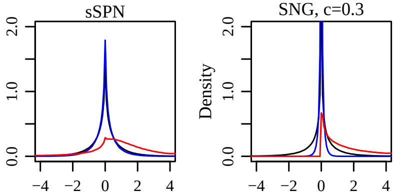
Prior Conditional Distributions



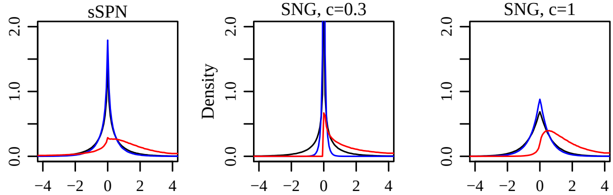
Prior Conditional Distributions



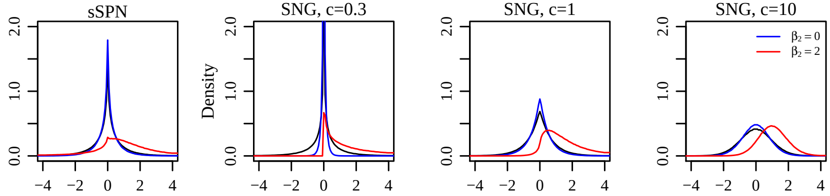
Prior Conditional Distributions



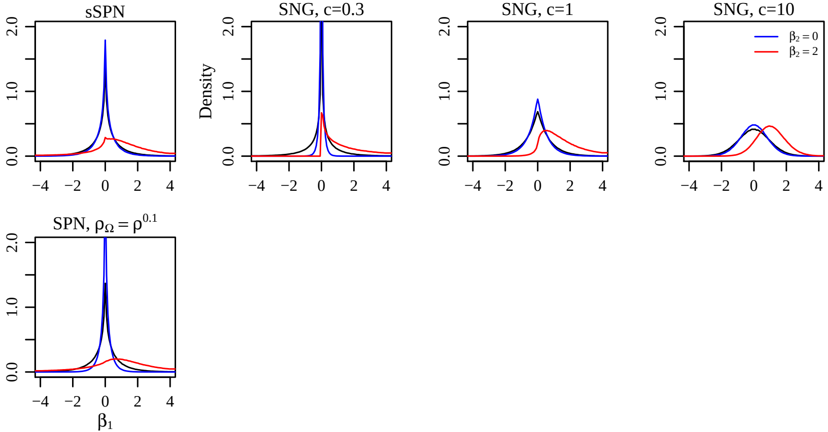
Prior Conditional Distributions



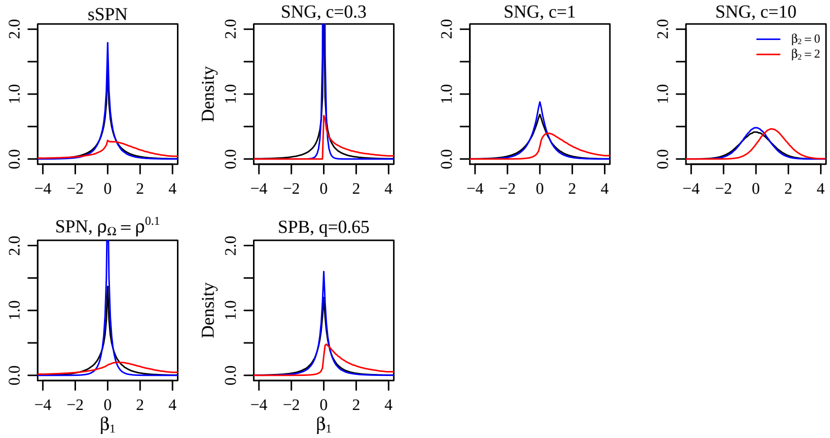
Prior Conditional Distributions



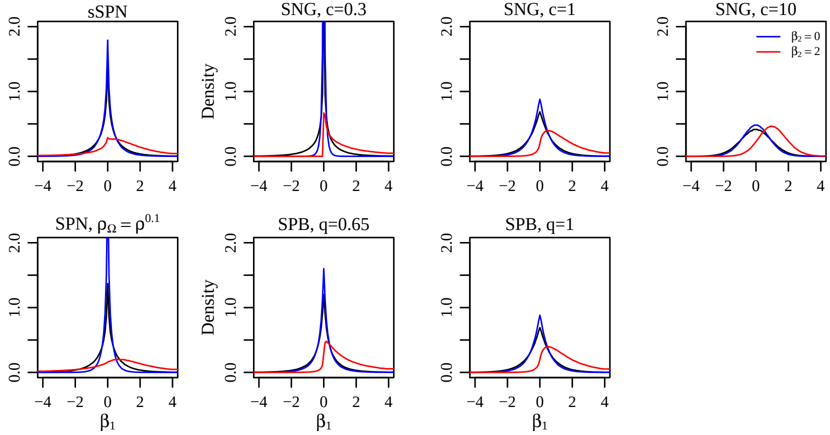
Prior Conditional Distributions



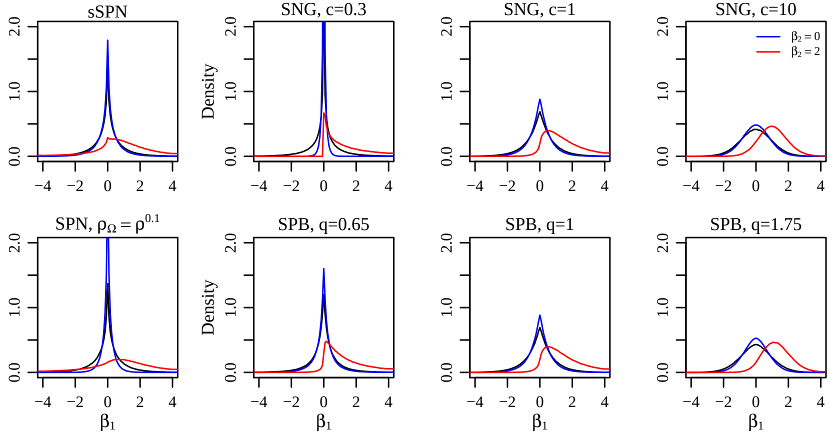
Prior Conditional Distributions



Prior Conditional Distributions



Prior Conditional Distributions



Range of $\boldsymbol{\Sigma} \to$ Achievable Structure

$$\boldsymbol{\Sigma} = \mathbb{E}\left[\boldsymbol{s}\boldsymbol{s}'\right] \circ \boldsymbol{\Omega}$$

Griffin and Hoff (2024)

Range of $\mathbf{\Sigma} \to$ Achievable Structure

$$\mathbf{\Sigma} = \mathbb{E}\left[\boldsymbol{s}\boldsymbol{s}'\right] \circ \mathbf{\Omega}$$

- Unrestricted under SPN prior

Griffin and Hoff (2024)

Range of $\mathbf{\Sigma} \to$ Achievable Structure

$$\mathbf{\Sigma} = \mathbb{E}\left[\boldsymbol{s}\boldsymbol{s}'\right] \circ \mathbf{\Omega}$$

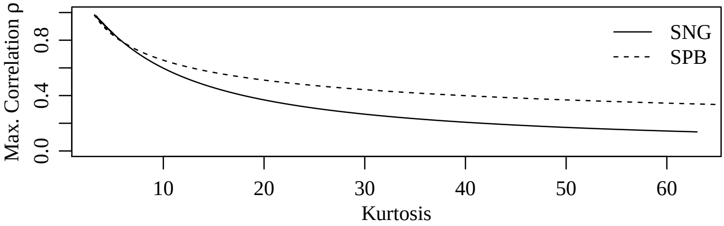
- Unrestricted under SPN prior
- Opaquely restricted under sSPN prior

Griffin and Hoff (2024)

Range of $\boldsymbol{\Sigma} \rightarrow$ Achievable Structure

$$\boldsymbol{\Sigma} = \mathbb{E}\left[\boldsymbol{s}\boldsymbol{s}'\right] \circ \boldsymbol{\Omega}$$

- Unrestricted under SPN prior
- Opaquely restricted under sSPN prior
- Clearly restricted under SNG, SPB priors



Griffin and Hoff (2024)

Computation under Structured Shrinkage Priors

Posterior Mode:
Approximated by solving

$$\hat{\boldsymbol{\beta}} = \text{maximize}_{\boldsymbol{\beta}} h\left(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\phi}\right) + \log\left(\int p\left(\boldsymbol{\beta}|\boldsymbol{s}, \boldsymbol{\theta}\right) p\left(\boldsymbol{s}|\boldsymbol{\theta}\right) d\boldsymbol{s}\right)$$

$\rightarrow$Requires simulation from $p\left(\boldsymbol{s}|\boldsymbol{\beta}, \boldsymbol{\theta}\right)$

Computation under Structured Shrinkage Priors

Posterior Mode:
Approximated by solving

$$\hat{\boldsymbol{\beta}} = \text{maximize}_{\boldsymbol{\beta}} h\left(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\phi}\right) + \log\left(\int p\left(\boldsymbol{\beta}|\boldsymbol{s}, \boldsymbol{\theta}\right) p\left(\boldsymbol{s}|\boldsymbol{\theta}\right) d\boldsymbol{s}\right)$$

$\rightarrow$Requires simulation from $p\left(\boldsymbol{s}|\boldsymbol{\beta}, \boldsymbol{\theta}\right)$

Posterior Mean, Median, Quantiles:

Approximated using samples from posterior

$$p\left(\boldsymbol{\beta}, \boldsymbol{s}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\phi}, \boldsymbol{\theta}\right)$$

$\rightarrow$Requires simulation from $p\left(\boldsymbol{\beta}|\boldsymbol{s}, \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\phi}\right)$ and $p\left(\boldsymbol{s}|\boldsymbol{\beta}, \boldsymbol{\theta}\right)$

Simulation from $p\left(\boldsymbol{s}|\boldsymbol{\beta},\boldsymbol{\theta}\right)$

Goal: Simulate $\boldsymbol{s}$ according to

$$\left(\prod_{j=1}^{p}|s|^{-1}\right)\exp\left\{-\frac{1}{2}\left(\mathbf{1}/\boldsymbol{s}\right)'\left(\boldsymbol{\Omega}^{-1}\circ\left(\boldsymbol{\beta}\boldsymbol{\beta}'\right)\right)\left(\mathbf{1}/\boldsymbol{s}\right)\right\}p\left(\boldsymbol{s}|\boldsymbol{\theta}\right)$$

Murray et al. (2010); Nishihara et al. (2014)

Simulation from $p\left(\boldsymbol{s}|\boldsymbol{\beta},\boldsymbol{\theta}\right)$

Goal: Simulate $\boldsymbol{s}$ according to

$$\left(\prod_{j=1}^{p}|s|^{-1}\right)\exp\left\{-\frac{1}{2}\left(\mathbf{1}/\boldsymbol{s}\right)'\left(\boldsymbol{\Omega}^{-1}\circ\left(\boldsymbol{\beta}\boldsymbol{\beta}'\right)\right)\left(\mathbf{1}/\boldsymbol{s}\right)\right\}p\left(\boldsymbol{s}|\boldsymbol{\theta}\right)$$

Trick: Creatively implement **elliptical slice sampling**

- Developed for densities with a normal component
- Extended to arbitrary densities later

Murray et al. (2010); Nishihara et al. (2014)

Simulation from $p\left(\boldsymbol{s}|\boldsymbol{\beta},\boldsymbol{\theta}\right)$

Goal: Simulate $\boldsymbol{s}$ according to

$$\left(\prod_{j=1}^{p}|s|^{-1}\right)\exp\left\{-\frac{1}{2}\left(\mathbf{1}/\boldsymbol{s}\right)'\left(\boldsymbol{\Omega}^{-1}\circ\left(\boldsymbol{\beta}\boldsymbol{\beta}'\right)\right)\left(\mathbf{1}/\boldsymbol{s}\right)\right\}p\left(\boldsymbol{s}|\boldsymbol{\theta}\right)$$

Trick: Creatively implement **elliptical slice sampling**

- Developed for densities with a normal component
- Extended to arbitrary densities later

Murray et al. (2010); Nishihara et al. (2014)

Specification of Hyperparameters

Maximum Marginal Likelihood: (**MMLE**) Solves

$$\text{maximize}_{\boldsymbol{\theta}} \int p\left(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\phi}\right) p\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right) d\boldsymbol{\beta}$$

- Approximate using Gibbs-within-EM

Perry (2017)

Specification of Hyperparameters

Maximum Marginal Likelihood: (**MMLE**) Solves

$$\text{maximize}_{\boldsymbol{\theta}} \int p\left(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\phi}\right) p\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right) d\boldsymbol{\beta}$$

- Approximate using Gibbs-within-EM

Moment Estimation:

1. Obtain estimate of $\hat{\boldsymbol{\Sigma}}$
   - Exact for **linear** model
   - Approximate for **nonlinear** model, e.g. normal $\boldsymbol{\beta}$ MMLE
2. Set $\hat{\boldsymbol{\Omega}}$ according to $\boldsymbol{\Sigma} = \mathbb{E}\left[\boldsymbol{s}\boldsymbol{s}'\right] \circ \boldsymbol{\Omega}$

Perry (2017)

Specification of Hyperparameters

Maximum Marginal Likelihood: (**MMLE**) Solves

$$\text{maximize}_{\boldsymbol{\theta}} \int p\left(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\phi}\right) p\left(\boldsymbol{\beta}|\boldsymbol{\theta}\right) d\boldsymbol{\beta}$$

- Approximate using Gibbs-within-EM

Moment Estimation:

1. Obtain estimate of $\hat{\boldsymbol{\Sigma}}$
    - Exact for **linear** model
    - Approximate for **nonlinear** model, e.g. normal $\boldsymbol{\beta}$ MMLE
2. Set $\hat{\boldsymbol{\Omega}}$ according to $\boldsymbol{\Sigma} = \mathbb{E}\left[\boldsymbol{s}\boldsymbol{s}'\right] \circ \boldsymbol{\Omega}$

Fully Bayes: Assume prior distributions for $\boldsymbol{\Omega}$, $\boldsymbol{\Psi}$

Perry (2017)

Return to EEG Example

Review of Data:

- $n \times 1$ response $\boldsymbol{y}$, indicators of target letter presence
  - $n = 20$ (220 held out for constructing "ground truth")
- $n \times p$ covariate matrix $\boldsymbol{X}$, contemporaneous EEG
  - $p = p_1 \times p_2 = 26 \times 8 = 208$

Polson et al. (2013)

Return to EEG Example

Review of Data:

- $n \times 1$ response $\boldsymbol{y}$, indicators of target letter presence
  - $n = 20$ (220 held out for constructing "ground truth")
- $n \times p$ covariate matrix $\boldsymbol{X}$, contemporaneous EEG
  - $p = p_1 \times p_2 = 26 \times 8 = 208$

$\rightarrow$ Logistic Regression

Structure:

- **Separable** across time and channel, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1$
- **AR-1** across time, $\sigma_{1,ij} = \rho_{\boldsymbol{\Sigma}}^{|i-j|}$
- **Unrestricted** across channels $\boldsymbol{\Sigma}_2$

Polson et al. (2013)

Return to EEG Example

Review of Data:

- $n \times 1$ response $\boldsymbol{y}$, indicators of target letter presence
  - $*$ $n = 20$ (220 held out for constructing "ground truth")
- $n \times p$ covariate matrix $\boldsymbol{X}$, contemporaneous EEG
  - $*$ $p = p_1 \times p_2 = 26 \times 8 = 208$

$\rightarrow$ Logistic Regression

Structure:

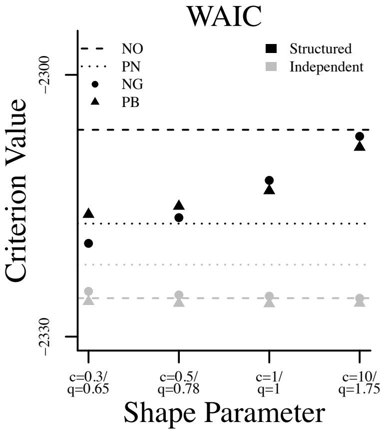
- **Separable** across time and channel, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1$
- **AR-1** across time, $\sigma_{1,ij} = \rho_{\boldsymbol{\Sigma}}^{|i-j|}$
- **Unrestricted** across channels $\boldsymbol{\Sigma}_2$

Specification of Hyperparameters:

- $\rho_{\boldsymbol{\Omega}}, \rho_{\boldsymbol{\Psi}} \sim \text{beta}\left(\left(p_1+1\right)/2, \left(p_1+1\right)/2\right)$,
  $\boldsymbol{\Omega}_2^{-1} \sim \text{Wishart}(p_2+2, \kappa^{-1}\boldsymbol{I}_{p_2})$, $\boldsymbol{\Psi}_2^{-1} \sim \text{Wishart}(p_2+2, \boldsymbol{I}_{p_2})$

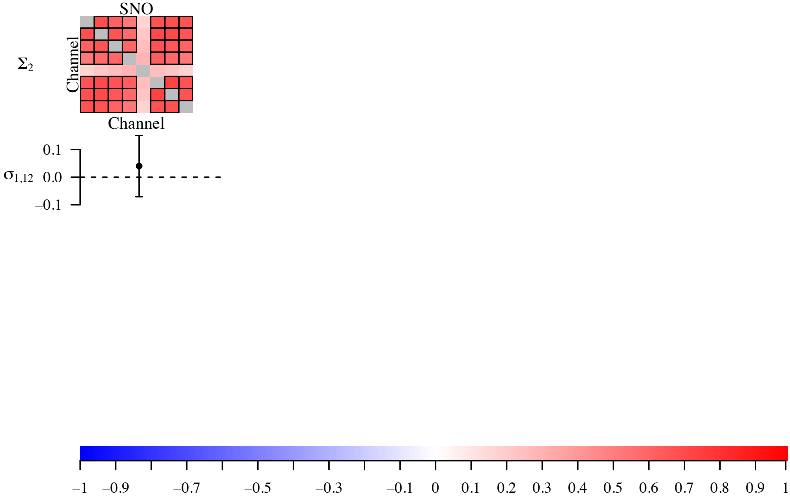
Polson et al. (2013)

Measures of Fit

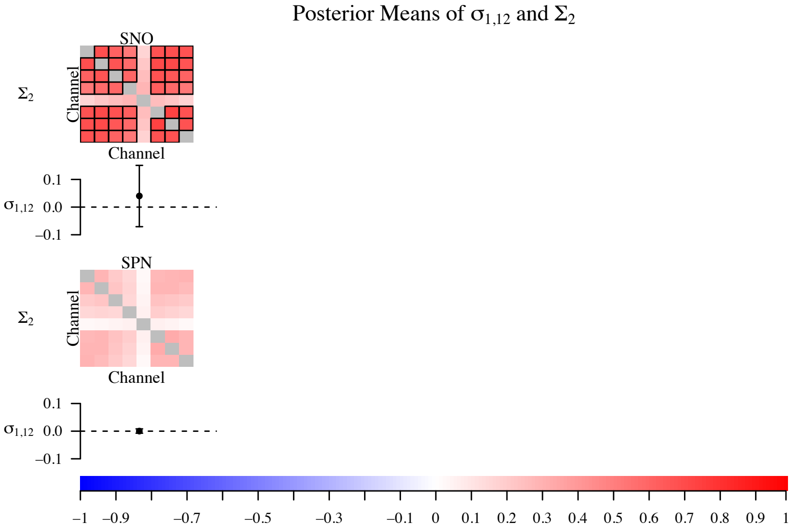


Approximate Estimates of Covariances

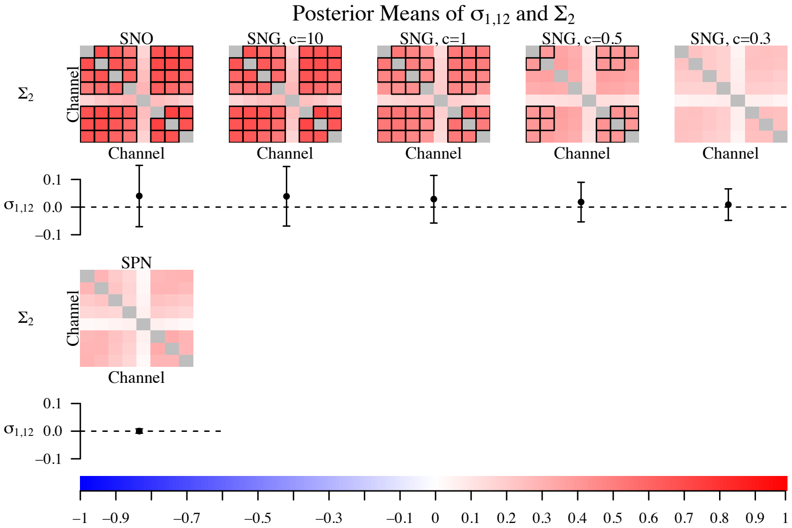
Posterior Means of $\sigma_{1,12}$ and Σ_2



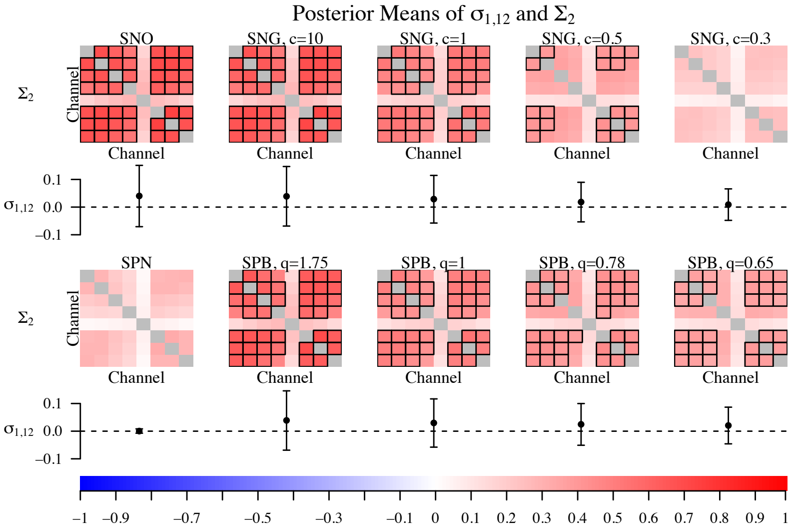
Approximate Estimates of Covariances



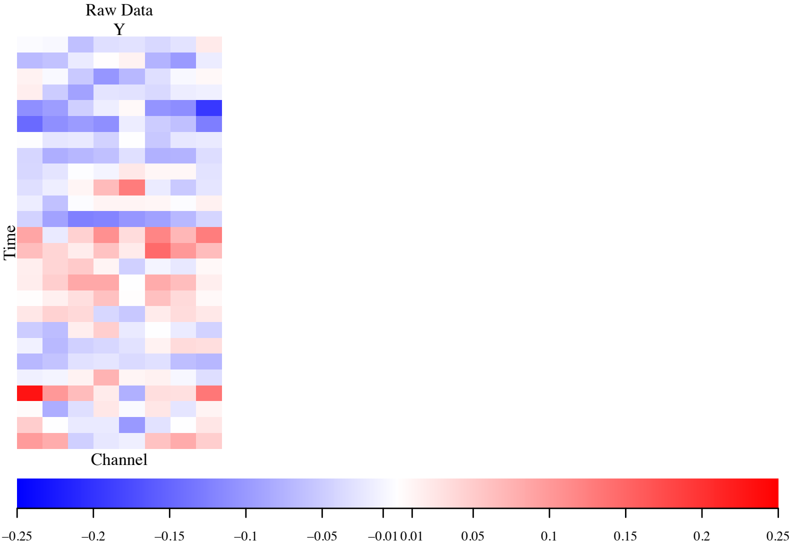
Approximate Estimates of Covariances



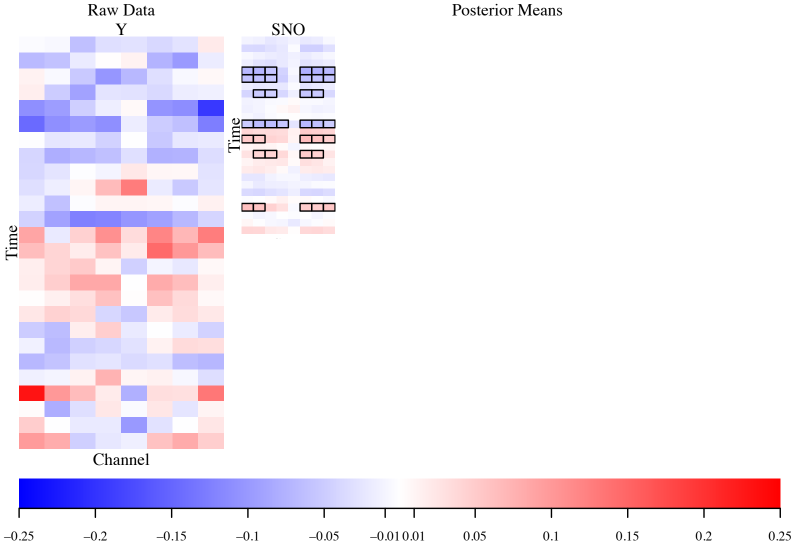
Approximate Estimates of Covariances



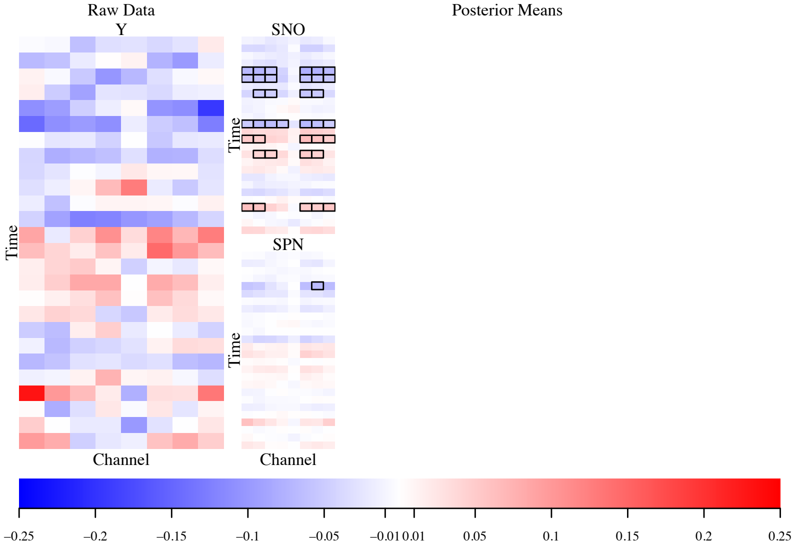
Posterior Mean Estimates



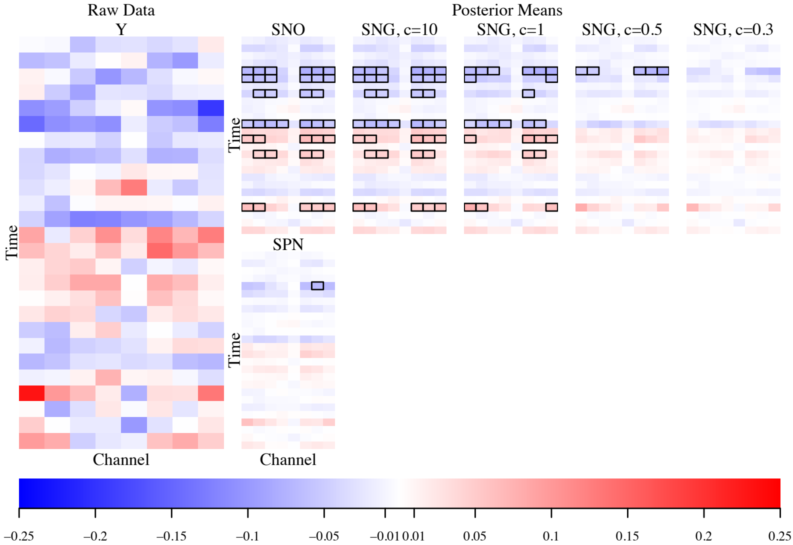
Posterior Mean Estimates



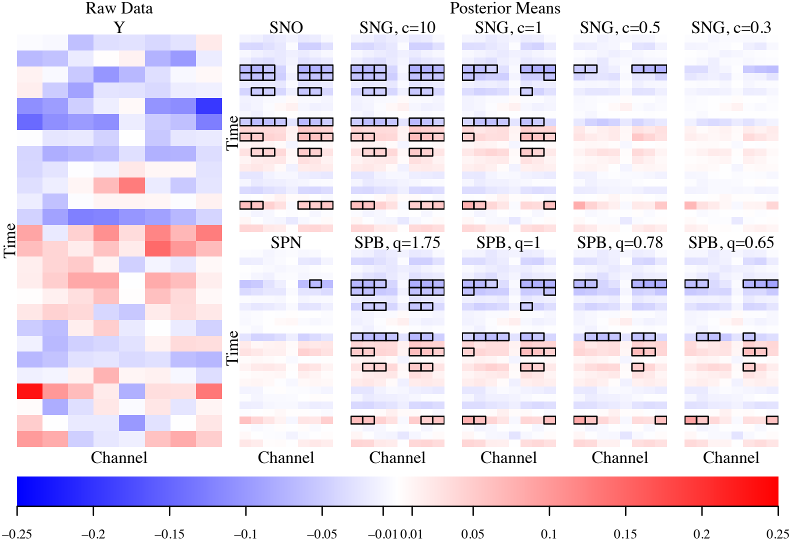
Posterior Mean Estimates



Posterior Mean Estimates



Posterior Mean Estimates



Comparison to “Ground Truth”

How do our estimates from 20 trials compare to estimates from all 240?

	SNO	SNG			
		$c = 10$	$c = 1$	$c = 0.5$	$c = 0.3$
TP%	18.18	18.18	13.64	0.00	0.00
FP%	17.68	17.68	10.98	3.05	0.00
	SPN	SPB			
		$q = 0.65$	$q = 0.78$	$q = 1$	$q = 1.75$
TP%	0.00	18.18	13.64	11.36	9.09
FP%	0.61	15.85	10.98	6.71	6.10

Conclusions and Next Steps

- ☺ Introducing structure to shrinkage priors has benefits
- ☺ Shrinkage priors w/o infinite spikes at zero have benefits

Property	**SNG**	**SPB**	**SPN**
Generalizes an Independent Shrinkage Prior	✓	✓	✓
Univariate Marginal is an Independent Shrinkage Prior	✓	✓	✓
Generalizes a Laplace Prior	$c = 1$	$q = 1$	
Generalizes a Normal Prior	$c \rightarrow \infty$	$q \rightarrow 2$	
Infinite Spike or Pole at Zero		$c \leq 1/2$	✓
Quadratic Scale Log Full Conditional			✓
Arbitrary Covariance Structure Achievable			✓

- How will these priors perform in other applications?
- Can we improve computation more?
- Are there other ways of introducing structure that lead to easier-to-compute modes?

Bruneel, B. (2018). De duikerklok en de p300-speller.

Finegold, M. and M. Drton (2011). Robust graphical modeling of gene networks using classical and alternative t-distributions. *Annals of Applied Statistics 5*, 1057–1080.

Forney, E., C. Anderson, P. Davies, W. Gavin, B. Taylor, and M. Roll (2013). A comparison of eeg systems for use in p300 spellers by users with motor impairments in real-world environments. *Proceedings of the Fifth International Brain-Computer Interface Meeting*.

Griffin, M. and P. D. Hoff (2024). Structured Shrinkage Priors. *Journal of Computational and Graphical Statistics 33*(1), 1–14.

Kalli, M. and Griffin (2014). Time-varying sparsity in dynamic regression models. *Journal of Econometrics 178*, 779–793.

Kowal, D. R., D. S. Matteson, and D. Ruppert (2017). Dynamic shrinkage processes. *ArXiv preprint. arXiv:1707.00763*, 1–45.

Kyung, M., J. Gill, M. Ghosh, and G. Casella (2010). Penalized regression, standard errors, and bayesian lassos. *Bayesian Analysis 5*, 369–412.

Murray, I., R. P. Adams, and D. J. C. Mackay (2010). Elliptical slice sampling. *Journal of Machine Learning Research: WCP 9*, 541–548.

Nishihara, R., I. Murray, and R. P. Adams (2014). Parallel mcmc with generalized elliptical slice sampling. *Journal of Machine Learning Research 15*, 2087–2112.

Perry, P. O. (2017). Fast moment-based estimation for hierarchical models. *Journal of the Royal Statistical Society. Series B 79*, 267–291.

Polson, N. G. and J. G. Scott (2010). Shrink globally, act locally: Bayesian sparsity and regularization.

Polson, N. G., J. G. Scott, and J. Windle (2013). Bayesian inference for logistic models using pólya-gamma latent variables. *Journal of the American Statistical Association 108*, 1339–1349.

van Gerven, M., B. Cseke, R. Oostenveld, and T. Heskes (2009). Bayesian source localization with the multivariate laplace prior. *Advances in Neural Information Processing Systems 22*, 1–9. Example of using¡br/¿¡br/¿It's sort of related to the normal product prior¡br/¿¡br/¿Perform approximate inference.

van Gerven, M. A. J., B. Cseke, F. P. de Lange, and T. Heskes (2010). Efficient bayesian multivariate fmri analysis using a sparsifying spatio-temporal prior. *Neuroimage 50*, 150–161. Very relevant!¡br/¿¡br/¿Approximate inference.

Wu, A., M. Park, O. Koyejo, and J. W. Pillow (2014). Sparse bayesian structure learning with dependent relevance determination prior. *Advances in Neural Information Processing Systems*, 1628–1636.

Zhao, S., C. Gao, S. Mukherjee, and B. E. Engelhardt (2016). Bayesian group factor analysis with structured sparsity. *Journal of Machine Learning Research 17*, 1–47.