1 Setup

Consider the normal means problem: for $i = 1, \ldots, n$, we observe $y_i \sim N(\beta_i, \sigma^2)$. Further suppose that we have a graph $G = (V, \mathcal{E})$ on n vertices with vertex set $V = \{1, 2, \ldots, n\}$ and edge set $\mathcal{E} \subset V \times V$. For each directed edge e from i to j we denote e^+ and e^- to be the sink (i) and source (j), respectively. We would like to encourage sparsity in the set $\{\beta_{e^+} - \beta_{e^-} : e \in \mathcal{E}\}$.

To this end, we propose the (possibly improper!) *spike-and-slab generalized LASSO* prior which is given by

$$\pi(\boldsymbol{\beta} \mid \boldsymbol{\gamma}) \propto \prod_{e \in \mathcal{E}} \exp\left\{-\left(\lambda_0(1-\gamma_e) + \lambda_1\gamma_e\right)|\beta_{e^+} - \beta_{e^-}|\right\} \tag{1}$$

where $\boldsymbol{\gamma} = \{\gamma_e : e \in \mathcal{E}\} \subset \{0,1\}^{|E|}$ is a collection of binary indicators. We model $\gamma_e \sim$ Bernoulli(θ) independently and complete our prior specification with $\theta \sim \text{Beta}(a, b)$.

We want to solve

$$(\hat{\beta}, \hat{\sigma}^2, \hat{\theta}) = \underset{\beta, \sigma^2, \theta}{\arg\min} \left\{ \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_i)^2 - \log \pi(\beta, \theta) - \log \pi(\sigma^2) \right\}$$

To this end, we use an EM algorithm in which we minimize the surrogate objective

$$\mathbb{E}_{\boldsymbol{\gamma}|\cdot}\left[\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \beta_i)^2 - \log \pi(\beta, \theta, \boldsymbol{\gamma}, \sigma^2) \mid y, \beta, \theta, \sigma^2\right]$$

where the expectation is taken over the conditional posterior of $\boldsymbol{\gamma}$ given all other quantities. This expectation is particularly easy to compute: the separability of the prior in (1) and the prior independence of the γ_e's renders the indicators independent *a posteriori* so that

$$\mathbb{E}[\gamma_e \mid \beta, \theta, \sigma^2, y] = p_e^\star := \frac{\lambda_1 \theta \mathrm{e}^{-\lambda_1|\beta_{e^+} - \beta_{e^-}|}}{\lambda_1 \theta \mathrm{e}^{-\lambda_1|\beta_{e^+} - \beta_{e^-}|} + \lambda_0(1-\theta)\mathrm{e}^{-\lambda_0|\beta_{e^+} - \beta_{e^-}|}}$$

Let $\lambda_e^\star = \lambda_0(1 - p_e^\star) + \lambda_1 p_e^\star$. In the M-step we minimize a surrogate objective that can be decomposed as $Q_1(\beta, \sigma^2) + Q_2(\theta)$ where

$$Q_1(\beta, \sigma^2) = \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \beta_i)^2 + \sum_{e \in \mathcal{E}} \lambda_e^\star |\beta_{e^+} - \beta_{e^-}| - \log \pi(\sigma^2)$$

$$Q_2(\theta) = -\left(a - 1 + \sum_{e \in \mathcal{E}} p_e^\star\right)\log\theta - \left(b + |\mathcal{E}| - 1 - \sum_{e \in \mathcal{E}} p_e^\star\right)\log(1-\theta)$$

Minimizing Q_2 can be done in close form – the optimal value is $(a+\sum_e p_e^\star)/(a+b+|\mathcal{E}-2)$. Minimizing Q_1 is somewhat more delicate. [skd]: If we place an inverse-gamma prior on σ^2, I'm not sure that the objective is convex as a function of β and σ^2. In case it isn't, we can resort to conditional updates – update β fixing σ^2 at its previous value and then updating σ^2 fixing β at its new value. If we iterate between these steps repeatedly until some convergence criterion is met, we have a full Expectation – Conditional Maximization algorithm. However, to keep the computational cost down, it's probably enough to do a single sweep over β and σ^2.

If we fix σ and decompose the graph into trails t then we need to solve

$$\arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2}\sum_{i=1}^{n}(y_i-\beta_i)^2 + \sum_t \sum_{e\in t} \sigma^2 \lambda_e^\star |\beta_{e^+}-\beta_{e^-}| \right\}$$

In trail t let the edges be $e_1,\ldots,e_{n_t}$ where $e_{k+1}^- = e_k^+$ (i.e. the end of the current edge is the start of the next edge along the trail). For each trail t we introduce a total of $2n_t$ slack variables $z_{2k-1,t} = \lambda_{e_k}^\star \beta_{e_k^-}$ and $z_{2k,t} = \lambda_{e_k}^\star \beta_{e_k^+}$. Then the minimization problem is

$$\arg\min_{\boldsymbol{\beta},z} \left\{ \frac{1}{2}\sum_{i=1}^{n}(y_i-\beta_i)^2 + \sigma^2 \sum_t \sum_{k=1}^{n_t} |z_{2k,t}-z_{2k}| \right\}$$

$$\text{subject to} \quad z_{2k-1,t} = \lambda_{e_{k,t}}^\star \beta_{e_{k,t}^+}$$
$$z_{2k,t} = \lambda_{e_{k,t}}^\star \beta_{e_{k,t}^-} \quad \text{for all } k,t$$

This can now be solved using ADMM.