INDIAN STATISTICAL INSTITUTE, KOLKATA

# Bayesian Fused Lasso Regression for Dynamic Binary Network using shrinkage prior

*Instructor:*
Dr. Sayantan Banerjee

*Student Name:*
Chirayata Kusari (JRF Statistics'2022)

# Contents

# 1 Bayesian Fused Lasso Regression for Dynamic Binary Networks

## 1.1 Bayesian inference: Latent variable approach

The reference paper used here is Betancourt, Rodríguez and Boyd,2017.
Following the similar notation used in the above mentioned paper, we get (dropping $i, j$ and $t$ for notional convenience and fixing $r$ at $r$)

$$L(\theta_r|\omega_r, \theta_{-r}) \propto (\frac{e^{\psi_r}}{1+e^{\psi_r}})^{y_r}(1-\frac{e^{\psi_r}}{1+e^{\psi_r}})^{1-y_r}$$

$$\propto \frac{e^{\psi_r y_r}}{(1+e^{\psi_r})}$$

$$\propto exp(\kappa_r\psi_r)exp(\psi_r^2\omega_r/2)PG_{\omega_r}(1,0)$$

$$\propto exp(-\psi_r^2\omega_r/2+\kappa_r\psi_r)$$

$$\propto exp(-\frac{\omega_r}{2}(\theta_r+C_r)^2+\kappa_r(\theta_r+C_r))$$

$$\propto exp(-\frac{\omega_r}{2}[(\frac{\kappa_r}{\omega_r}-C_r)-\theta_r]^2)$$

Hence, $\quad y^*_{i,j,t,r} = (\frac{\kappa_r}{\omega_r} - C_r) \sim N(\theta_r, \omega_r^{-1})$.
Now, using the fact the double exponential distribution can be expressed as a scale mixture of normals with exponential mixing density (Park and Casella 2008)

$$\frac{a}{2}\exp(-a|x|) = \int_0^\infty \frac{1}{\sqrt{2\pi\tau}}\exp\left(\frac{x^2}{2\tau}\right)\frac{a^2}{2}\exp\left(-\frac{a^2\tau}{2}\right)d\tau,$$

the proposed model can be expressed as a simple hierarchical extension of a dynamic linear model

$$y^*_{t,r} = \theta_{t,r} + \epsilon_{t,r}, \quad \epsilon_{t,r} \sim \text{Normal}\left(0, \omega_{t,r}^{-1}\right),$$
$$\theta_{t,r} = \theta_{t-1,r} + \varepsilon_{t,r}, \quad \varepsilon_{t,r} \sim \text{Normal}\left(0, \tau_{t,r}^2\right),$$

for $2 \leq t \leq T$, where $\theta_{1,r} \sim \text{Normal}\left(\hat{\theta}_{0,r}, \tau_{1,r}^2\right)$ and $\tau_{t,r}^2$ is exponentially distributed a priori with mean $\frac{2}{\lambda^2}$.

We rely on the dynamic linear model representation to update the parameters in a component-wise fashion using a forward filtering backward sampling (FFBS) algorithm (Carter and Kohn 1994; Frühwirth-Schnatter 1994). Furthermore, the latent parameters $\tau_{t,r}$ for $t = 0, \ldots, T-1$ are independent a posteriori and updated as

$$\left(1/\tau_{t,r}^2 \mid \mathbf{\Theta}_r, \lambda\right) \sim \text{IGau}\left(\sqrt{\frac{\lambda^2}{\left(\theta_{t,r} - \theta_{t-1,r}\right)^2}}, \lambda^2\right)$$

where IGau denotes the Inverse Gaussian distribution.

## 1.2 Data and prior structure for Horse shoe prior:

Now, instead of Laplace prior, we try to impose heavy-tailed Horse-shoe shrinkage prior on the consecutive difference of $\theta$. using the same idea in Banerjee and Shen,2022. Hence the hierarchical extension of the dynamic model will be -

$$\begin{gathered} y^*_{i,j,t,r} = \theta_{i,j,t,r} + \epsilon_{i,j,t,r}, \quad \epsilon_{i,j,t,r} \sim N(0, \omega_{i,j,t,r}) \\ \theta_{i,j,t,r} = \theta_{i,j,t-1,r} + \eta_{i,j,t,r} \end{gathered}$$

$$\begin{gathered} \eta_{i,j,t,r} | \tau, \lambda^2_{i,j,t,r}, \omega_{i,j,t,r} \sim N(0, \frac{\lambda^2_{i,j,t,r}\tau^2}{\omega_{i,j,t,r}}) \\ \lambda_{i,j,t,r} \sim C^+(0,1) \\ \tau \sim C^+(0,1) \\ \omega_{i,j,t,r} \sim PG(1,0) \end{gathered}$$

$\lambda_{t,r}$ and $\tau$ distributions can be further reduced to

$$\begin{gathered} \lambda^2_{i,j,t,r} | \nu_{i,j,t,r} \sim IG(\frac{1}{2}, \frac{1}{\nu_{i,j,t,r}}) \\ \tau^2 | \xi \sim IG(\frac{1}{2}, \frac{1}{\xi}) \\ \nu_{i,j,t,r}, \xi \sim IG(\frac{1}{2}, 1) \end{gathered}$$

## 1.3 Posterior computation- HS Prior:

We drop $i, j$ for convenience. Lets fix $r$ at $r$.

1. $\theta_{t,r} | - \sim N(\theta^*_t, \frac{1}{\omega^*_t})$, $\qquad 1 \leq t \leq T$

where,

$$\theta_1^* = \frac{\left(y_{1,r}^*\omega_{1,r} + \frac{\theta_{2,r}\omega_{2,r}}{\tau^2\lambda_{2,r}^2}\right)}{\left(\omega_{1,r} + \frac{\omega_{1,r}}{\tau^2\lambda_{1,r}^2} + \frac{\omega_{2,r}}{\tau^2\lambda_{2,r}^2}\right)}, \quad \omega_1^* = \left(\omega_{1,r} + \frac{\omega_{1,r}}{\tau^2\lambda_{1,r}^2} + \frac{\omega_{2,r}}{\tau^2\lambda_{2,r}^2}\right)$$

$$\theta_t^* = \frac{\left(y_{t,r}^*\omega_{t,r} + \frac{\theta_{t-1,r}\omega_{t,r}}{\tau^2\lambda_{t,r}^2} + \frac{\theta_{t+1,r}\omega_{t+1,r}}{\tau^2\lambda_{t+1,r}^2}\right)}{\left(\omega_{t,r} + \frac{\omega_{t,r}}{\tau^2\lambda_{t,r}^2} + \frac{\omega_{t+1,r}}{\tau^2\lambda_{t+1,r}^2}\right)} \quad \omega_t^* = \left(\omega_{t,r} + \frac{\omega_{t,r}}{\tau^2\lambda_{t,r}^2} + \frac{\omega_{t+1,r}}{\tau^2\lambda_{t+1,r}^2}\right) \qquad 2 \le t \le T$$

2. $\lambda_{t,r}^2|- \sim IG(1, \left(\frac{1}{\nu_{t,r}} + \frac{\eta_{t,r}^2\omega_{t,r}}{\tau^2}\right)) \qquad 1 \le t \le T$

where $\eta_{t,r}$ is defined as $\eta_{t,r} = (\theta_{t,r} - \theta_{t-1,r})$

3. $\tau^2|- \sim IG\left(\left(\frac{3n(n-1)Tr}{4} - 1\right), \sum\sum_{i<j} \sum_t \sum_r \left(\frac{1}{\xi} + \frac{\eta_{i,j,t,r}^2\omega_{i,j,t,r}}{\lambda_{i,j,t,r}^2}\right)\right)$

4. $\nu_{t,r}|- \sim IG\left(1, \frac{1+\lambda_{t,r}^2}{\lambda_{t,r}^2}\right) \qquad 1 \le t \le T$

5. $\xi|- \sim IG\left(1, \frac{(1+\tau^2)}{\tau^2}\right)$

6. $\omega_{t,r}|- \sim PG(1, (\theta_{t,r} + C_{t,r})) \qquad 1 \le t \le T$

## 1.4 Data and prior structure setup for t-shrinkage prior:

Similar to the above setup we get the following dynamic hierarchical model-

$$y_{i,j,t,r}^* = \theta_{i,j,t,r} + \epsilon_{i,j,t,r}, \quad \epsilon_{i,j,t,r} \sim N(0, \omega_{i,j,t,r})$$
$$\theta_{i,j,t,r} = \theta_{i,j,t-1,r} + \eta_{i,j,t,r}$$

$$\eta_{i,j,t,r}|\omega_{i,j,t,r} \quad \sim \quad t_v\left(\frac{m}{\omega_{i,j,t,r}}\right)$$
$$\omega_{i,j,t,r} \sim PG(1, 0)$$

The hierarchy is further reduced to

$$\begin{aligned} \eta_{i,j,t,r}|\omega_{i,j,t,r},\lambda_{i,j,t,r} &\sim N\Big(0,\lambda_{i,j,t,r}\cdot\frac{1}{\omega_{i,j,t,r}}\Big) \\ \lambda_{i,j,t,r} &\sim IG(a_\lambda,b_\lambda) \\ \omega_{i,j,t,r} &\sim PG(1,0) \end{aligned}$$

$a_\lambda, b_\lambda$ are known.

## 1.5 Posterior computation-t-shrinkage prior:

1. $\theta_{t,r}|- \sim N(\theta_t^{**}, \frac{1}{\omega_t^{**}})$, where

$$\theta_1^{**} = \frac{\Big(y_{1,r}^*\omega_{1,r} + \frac{\theta_{2,r}\omega_{2,r}}{\lambda_{2,r}}\Big)}{\Big(\omega_{1,r} + \frac{\omega_{1,r}}{\lambda_{1,r}} + \frac{\omega_{2,r}}{\lambda_{2,r}}\Big)} \qquad \omega_1^{**} = \Big(\omega_{1,r} + \frac{\omega_{1,r}}{\lambda_{1,r}} + \frac{\omega_{2,r}}{\lambda_{2,r}}\Big)$$

$$\theta_t^{**} = \frac{\Big(y_{t,r}^*\omega_{t,r} + \frac{\theta_{t-1,r}\omega_{t,r}}{\lambda_{t,r}} + \frac{\theta_{t+1,r}\omega_{t+1,r}}{\lambda_{t+1,r}}\Big)}{\Big(\omega_{t,r} + \frac{\omega_{t,r}}{\lambda_{t,r}} + \frac{\omega_{t+1,r}}{\lambda_{t+1,r}}\Big)} \qquad \omega_t^{**} = \Big(\omega_{t,r} + \frac{\omega_{t,r}}{\lambda_{t,r}} + \frac{\omega_{t+1,r}}{\lambda_{t+1,r}}\Big), \qquad 2 \le t \le T$$

2. $\lambda_{t,r}|- \sim IG\Big(a_\lambda + \frac{1}{2}, b_\lambda + \omega_{t,r}\eta_{t,r}^2\Big)$, where, $\eta_{t,r} = (\theta_{t,r} - \theta_{t-1,r})$, $1 \le t \le T$

3. $\omega_{t,r}|- \sim PG(1,(\theta_{t,r} + C_{t,r}))$, $1 \le t \le T$

## 1.6 Data and prior structure setup for Horse shoe like prior:

($i$ and $j$ being common to all are dropped for notional convenience)

$$\begin{aligned} y_{t,r}^* &= \theta_{t,r} + \epsilon_{t,r} \\ \epsilon_{t,r} &\sim N(0,\frac{1}{\omega_{t,r}}) \\ \omega_{t,r} &\sim PG(1,0) \end{aligned}$$

$$\begin{aligned} \theta_{t,r} &= \theta_{(t-1),r} + \eta_{t,r} \\ \eta_{t,r}|\nu_{t,r},a &\sim N(0,\frac{a}{2\nu_{t,r}}) \\ \pi(\nu_{t,r}) &\sim \frac{(1-exp(-\nu_{t,r}))}{2\pi^{\frac{1}{2}}\nu_{t,r}^{\frac{3}{2}}} \end{aligned}$$

## 1.7 Posterior computation - HS like:

Fix some r at r (say)

1. $\theta_{t,r}|- \sim N\Big(\theta_{t,r}^{hs-like}, \frac{1}{\omega_{t,r}^{hs-like}}\Big)$

where, $\theta_{t,r}^{hs-like} = \frac{\frac{y^*_{t,r}\omega_{t,r}}{2}+\frac{\theta_{t-1,r}\nu_{t,r}}{a}+\frac{\theta_{t+1,r}\nu_{t+1,r}}{a}}{\frac{\omega_{t,r}}{2}+\frac{\nu_{t,r}}{a}+\frac{\nu_{t+1,r}}{a}}, \quad \omega_{t,r}^{hs-like} = (\omega_{t,r} + \frac{2\nu_{t,r}}{a} + \frac{2\nu_{t+1,r}}{a})$

2. $\nu_{t,r}|- \propto nu_{t,r}^{-1} exp\Big(-\nu_{t,r}(\frac{\eta^2_{t,r}}{a})\Big)\Big[1-e^{-\nu_{t,r}}\Big]$

3. $\omega_{t,r}|- \sim PG(1,(\theta_{t,r}+C_{t,r}))$

## 1.8 EM Algorithm using Horse shoe like prior:

log-likelihood can be written as,

$$lnL = \sum_i\sum_j\sum_t \frac{1}{2}ln(\omega_{t,r}) - \frac{1}{2}\omega_{t,r}(\theta_{t,r}-y^*_{t,r})^2 + \frac{1}{2}(ln(2\nu_{t,r})-ln(a)) - \frac{2\nu_{t,r}}{2a}(\theta_{t,r}-\theta_{t-1,r})^2$$
$$+ln(1-e^{-\nu_{t,r}}) - \frac{3}{2}ln(\nu_{t,r}) + ln(\pi(\omega_{t,r})$$

where $\pi(\omega_{t,r})$ is the density of PG(1,0).

posterior distribution of $\theta_{t,r}$ is $N(.,.)$
posterior distribution of $\nu_{t,r}$

$$\nu_{t,r} \propto \frac{\nu_{t,r}^{\frac{1}{2}}}{a^{\frac{1}{2}}}\cdot exp\Big(-\frac{2\nu_{t,r}}{2a}(\theta_{t,r}-\theta_{t-1,r})^2\Big)\cdot\Big(\frac{1-e^{-\nu_{t,r}}}{\nu_{t,r}^{\frac{3}{2}}}\Big)$$

**E-step**
m-th step iterate,

$$\hat{\theta_{t,r}}^{(m)} = \mathrm{E}(\theta_{t,r}|-) = \frac{\frac{y^*_{t,r}\hat{\omega_{t,r}}^{(m-1)}}{2}+\frac{\hat{\theta_{t-1,r}}^{(m-1)}\hat{\nu_{t,r}}^{(m-1)}}{a}+\frac{\hat{\theta_{t+1,r}}^{(m-1)}\hat{\nu_{t+1,r}}^{(m-1)}}{a}}{\frac{\hat{\omega_{t,r}}^{(m-1)}}{2}+\frac{\hat{\nu_{t,r}}^{(m-1)}}{a}+\frac{\hat{\nu_{t+1,r}}^{(m-1)}}{a}}$$

$$\hat{\nu_{t,r}}^{(m)} = E(\nu_{t,r}|-) = \int_0^\infty \frac{1}{2\sqrt{2}\pi a^{\frac{1}{2}}}\frac{1}{\cancel{\nu_{t,r}}}(1-exp(-\nu_{t,r}))exp(-\frac{\nu_{t,r}}{a}\eta^2_{t,r})\cancel{\nu_{t,r}}d\nu_{t,r}$$
$$= \frac{a^{\frac{1}{2}}}{2\pi}\Big(\frac{1}{\hat{\eta^2_{t,r}}^{(m-1)}} - \frac{1}{a+\hat{\eta^2_{t,r}}^{(m-1)}}\Big)$$

Note that,

$$\sqrt{2\nu_{t,r}}\eta_{t,r}|a \sim N(0,a)$$

Thus,

$$\hat{a}^{(m)}|\eta_{i,j,t,r} = \tfrac{4}{n(n-1)TR} \quad \sum_{i<j}\sum\sum_{t}\sum_{r}\hat{\nu_{i,j,t,r}}^{(m-1)}\eta^2_{i,j,t,r} = \tfrac{2a^{\frac{3}{2}}}{N\pi}\sum_{i<j}\sum\sum_{t}\sum_{r}\tfrac{1}{a+\eta^2_{i,j,t,r}}$$

where $N = n(n-1)TR$

Then $(m+1)^{th}$ iterate of $a$ by expectation-conditional-maximization will be

$$\hat{a}^{(m+1)}|\boldsymbol{\eta}^{(m)} = \tfrac{2(\hat{a}^{(m)})^{\frac{3}{2}}}{N\pi}\sum_{i<j}\sum\sum_{t}\sum_{r}\tfrac{1}{\hat{a}^{(m)}+\hat{\eta^{(m)}_{i,j,t,r}}^2}$$

Similarly we have

$\hat{\omega_{t,r}}^{(m)} = E(\omega_{t,r}|-) = E(PG(1,(\theta_{t,r}+C_{t,r}))$

**M-step**

$$\frac{\partial lnL}{\partial\theta_{t,r}}$$

$$= -\omega_{i,j,t,r}(\theta_{i,j,t,r} - y^*_{i,j,t,r}) - \frac{2\nu_{i,j,t,r}}{a}(\theta_{i,j,t,r} - \theta_{i,j,t-1,r}) + \frac{2}{a}\nu_{i,j,t+1,r}(\theta_{i,j,t+1,r} - \theta_{i,j,t,r})$$

Equating with 0, we get the EM-estimate of $\theta_{t,r}$ -

$$\theta^{(m)}_{t,r\,EM} = \frac{\frac{y^*_{t,r}\hat{\omega_{t,r}}^{(m-1)}}{2} + \frac{\hat{\theta_{t-1,r}}^{(m-1)}\hat{\nu_{t,r}}^{(m-1)}}{a} + \frac{\hat{\theta_{t+1,r}}^{(m-1)}\hat{\nu_{t+1,r}}^{(m-1)}}{a}}{\frac{\hat{\omega_{t,r}}^{(m-1)}}{2} + \frac{\hat{\nu_{t,r}}^{(m-1)}}{a} + \frac{\hat{\nu_{t+1,r}}^{(m-1)}}{a}}$$

(putting the estimated values in E-step in the likelihood)

we will stop if

$$|\hat{\theta_{t,r}}^{(m)} - \hat{\theta_{t,r}}^{(m-1)}| < \epsilon,$$

$\epsilon$ is a predetermined small, positive quantity

# 2 Posterior Mode estimation:

MCMC Algorithm allows us to get full information about the posterior distribution but can be slow because no of observations grows rapidly with the square of the nodes. We try to implement local Linear approximation(for non-convex penalties) following the algorithm proposed in **'A Laplace Mixture Representation of the Horseshoe and Some Implications (Ksheera Sagar and Anindya Bhadra)'**. We present here an alternative procedure for MAP estimation under horse-shoe prior setup.( As we know that under the Gaussian Likelihood, the LLA algorithm is equivalent to an EM algorithm for MAP estimation under a prior, not necessarily proper.) Consider the following

$$y^*_{i,j,t,r} = \theta_{i,j,t,r} + \epsilon_{i,j,t,r}, \quad \epsilon_{i,j,t,r} \sim N(0, \frac{1}{\omega_{i,j,t,r}})$$

We specify the complete model along with the prior structure :

$$y^*_{i,j,t,r} = \theta_{i,j,t,r} + \epsilon_{i,j,t,r}, \quad \epsilon_{t,r} \sim N\Big(0, \frac{1}{\omega_{t,r}}\Big)$$
$$\theta_{t,r} = \theta_{t-1,r} + \eta_{t,r}$$

$$\eta_{t,r}|\lambda_{t,r}, \omega_{t,r}, \tau \sim N\Big(0, \frac{\lambda^2_{t,r}\tau^2}{\omega_{t,r}}\Big)$$
$$\lambda_{t,r} \sim C^+(0,1)$$
$$\tau \sim C^+(0,1)$$
$$\omega_{t,r} \sim PG(1,0)$$

Consider the penalized objective function :

$$\underset{\theta}{argmin}\Big(\sum_t (y^*_t - \theta_t)^2 + pen(|\eta|)\Big)$$
$$= \underset{\theta}{argmin}\Big(\sum_t (y^*_t - \theta_t)^2 + \sum_t \lambda^{(k)}_t |\eta_t|\Big)$$

Where $\lambda^{(k)} = pen'_{HS}(|\eta^{(k)}|)$.

In the paper **A Laplace Mixture Representation of the Horseshoe and Some Implications** Sagar and Bhadra have shown the following properties:

- HS density can be represented as a laplace scale mixture, which establishes the complete monotonicity of HS density.(via Bernstein-Widder theorem)
- Bochner's theorem ensures the derivative of the induced penalty function for Horseshoe [denoted by $pen_{HS}(x)$] is completely monotone. Note that, penalty function can be written as $pen_{HS}(x) = -log(p_{HS}(x))$

- According to *Zou and Li* for a penalty function $pen(|x|)$ such that $exp(-pen|x|)$ is completely monotone under Gaussian Likelihood, the LLA Algorithm is equivalent to EM Algorithm for MAP estimation under a prior(not necesssarily proper), with density $p(x) \propto exp(-pen(x))$

In out current setup, under a HS prior our induced penalty function is strongly concave and derivative of the penalty function is bounded.
If $\eta^{(k)}$ denotes the $k^{th}$ iterate,the optimization problem at the above reduces to:

$$\eta^{(k)} = \underset{\theta}{argmin}\Big(\sum_t (y_t^* - \theta_t)^2 + \sum_t pen^{'}(|\eta_{t,r}^{(k-1)}|)|\eta_{t,r}^{(k-1)}|\Big),$$

where $i, j, r$ *are fixed*

We start with an initial value of $\theta$ at $\theta_0^{(0)}$; and update the values at each iteration. The soft thresholding part in LLA, which is essentially a solution to the reduced Lasso equation gives us sparse estimates of $\eta$, hence we can recover the block-structure out of $\theta$ values.

**Laplace mixture representation of marginal Horseshoe density**

**Proposition 1** *The marginal Horseshoe density admits the following representation as a laplace mixture:*

$$p_{HS}(\eta) = \tfrac{2}{\pi^{\frac{3}{2}}\tau}\int_0^\infty exp\Big(-\tfrac{|\eta|}{\tau}u\Big)\cdot\phi(u)du,$$

*where* $\phi(t) = \mathbb{E}_\omega\Big(D_+(\frac{t}{\sqrt{2\omega}})\Big); \omega \sim p(\omega)(\equiv PG(0,1))$.
$Moreover, D_+(z) = exp(-z^2)\displaystyle\int_0^z e^{t^2}dt, z > 0,$ *is the Dawson function.*

*Proof:* From the hierarchy in the equations, the marginal will be

$$p_{HS}(\eta) = \int_\omega \int_{\lambda=0}^\infty \tfrac{1}{\sqrt{2\pi\frac{\lambda^2\tau^2}{\omega}}} exp\Big(-\tfrac{\tau^2}{\frac{2\lambda^2\tau^2}{\omega}}\Big)\cdot\tfrac{2}{\pi}\tfrac{1}{1+\lambda^2}\cdot p(\omega)d\omega d\lambda$$
$$= \tfrac{\sqrt{2}}{\pi^{\frac{3}{2}}\tau}\int_\omega \sqrt{\omega}\;\; p(\omega)\Big(\int_0^\infty \tfrac{1}{\lambda}exp\Big(-\tfrac{\omega\eta^2}{2\lambda^2\tau^2}\Big)\;\tfrac{1}{1+\lambda^2}\;\; d\lambda\Big)d\omega$$
$$= \tfrac{\sqrt{2}}{\pi^{\frac{3}{2}}\tau}\int_\omega \sqrt{\omega}\;\; p(\omega)\int_0^\infty \tfrac{1}{\lambda}exp\Big(-\tfrac{\omega\eta^2}{2\lambda^2\tau^2}\Big)\;\int_0^\infty exp\Big(-(\lambda^2+1)t\Big)dt\;\; d\lambda\;\; d\omega$$
$$= \tfrac{\sqrt{2}}{\pi^{\frac{3}{2}}\tau}\int_\omega \sqrt{\omega}\;\; p(\omega)\int_{\lambda=0}^\infty\int_{t=0}^\infty \tfrac{1}{\lambda}exp\Big(-\tfrac{\omega\eta^2}{2\lambda^2\tau^2} - \lambda^2 t\Big)\;\; e^{-t}dt\;\; d\lambda\;\; d\omega$$

(substitute $\lambda^2 = m$)

$= \frac{\sqrt{2}}{\pi^{\frac{3}{2}}\tau}\int_{\omega}\sqrt{\omega}\;\; p(\omega)\int_0^{\infty}\int_0^{\infty} exp\Big(-\frac{\omega\eta^2}{2m\tau^2}-mt\Big)\frac{1}{2m}\;\; e^{-t}dt\;\; dm\;\; d\omega$
$= \frac{\sqrt{2}}{\pi^{\frac{3}{2}}\tau}\int_{\omega}\sqrt{\omega}\;\; p(\omega)\int_0^{\infty} K_0\Big(\sqrt{2t\omega}\frac{|\eta|}{\tau}\Big)e^{-t}dt\;\; d\omega$

(Generalized inverse Gaussian integral; $K_0$= modified Bessel function of the second kind of order 0.)

$= \frac{\sqrt{2}}{\pi^{\frac{3}{2}}\tau}\int_{\omega}\sqrt{\omega}\;\; p(\omega)\int_0^{\infty}\int_0^{\infty} exp\Big(-\frac{|\eta|}{\tau}\sqrt{2t\omega}\;\; cosh\zeta\Big)e^{-t}d\zeta\;\; dt\;\; d\omega$
$= \frac{\sqrt{2}}{\pi^{\frac{3}{2}}\tau}\int_{\omega}\sqrt{\omega}\;\; p(\omega)\int_0^{\infty}\int_0^{\infty} exp\Big(-\frac{|\eta|}{\tau}\sqrt{2t\omega}\;\; cosh\zeta - t\Big)d\zeta\;\; dt\;\; d\omega$

Now consider the inside integral only,

$\int_0^{\infty}\int_0^{\infty} exp\Big(-\frac{|\eta|}{\tau}\sqrt{2t\omega}\;\; cosh\zeta - t\Big)d\zeta\;\; dt$

Let, $\sqrt{2t\omega}\;\; cosh\zeta = u$ and $\quad t = v$ with Jacobian

$$J = \begin{vmatrix} \frac{\partial\zeta}{\partial u} & \frac{\partial\zeta}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{\partial\zeta}{\partial u} & \frac{\partial\zeta}{\partial v} \\ 0 & 1 \end{vmatrix} = \left|\frac{\partial\zeta}{\partial u}\right| = \frac{1}{\sqrt{u^2-2v\omega}}$$

Thus the inside integral becomes,

$\int_0^{\infty}\int_0^{\frac{u^2}{2\omega}} exp\Big(-\frac{|\eta|}{\tau}u - v\Big)\;\; \frac{1}{\sqrt{u^2-2v\omega}}dv\;\; du$
$= \int_0^{\infty} exp\Big(-\frac{|\eta|}{\tau}u\Big)\int_0^{\frac{u^2}{2\omega}}\;\; \frac{exp(-v)}{\sqrt{u^2-2v\omega}}\;\; dv\;\; du$

substitute, $u^2-2v\omega = 2t^2$

$= \int_0^{\infty} exp\Big(-\frac{|\eta|}{\tau}u\Big)\int_0^{\frac{u}{\sqrt{2}}}\;\; \frac{\sqrt{2}}{\omega}\;\; e^{-\frac{u^2}{2\omega}}e^{\frac{t^2}{\omega}}\;\; dt\;\; du$
$= \frac{\sqrt{2}}{\omega}\int_0^{\infty} exp\Big(-\frac{|\eta|}{\tau}u\Big)\Bigg\{e^{-\frac{u^2}{2\omega}}\int_0^{\frac{u}{\sqrt{2}}}\;\; e^{\frac{t^2}{\omega}}\;\; dt\;\;\Bigg\}du$

Hence The final integral takes the form:

$= \frac{2}{\pi^{\frac{3}{2}}\tau}\int_{\omega}\;\; p(\omega)\;\; \frac{1}{\sqrt{\omega}}\int_0^{\infty} exp\Big(-\frac{|\eta|}{\tau}u\Big)\Bigg\{e^{-\frac{u^2}{2\omega}}\int_0^{\frac{u}{\sqrt{2}}}\;\; e^{\frac{t^2}{2\omega}}\;\; dt\Bigg\}\;\; du\;\; dw$

(Note that, $e^{-\frac{u^2}{2\omega}}\int_0^{\frac{u}{\sqrt{2}}}\;\; e^{\frac{t^2}{2\omega}}\;\; dt = \sqrt{\omega}\;\; D_+\Big(\frac{u}{\sqrt{2\omega}}\Big)$; where $D_+(\cdot)$ denotes the Dawson

Function.)

$$= \frac{2}{\pi^{\frac{3}{2}}\tau} \int_{\omega} p(\omega) \ \frac{1}{\cancel{\sqrt{\omega}}} \ \cancel{\sqrt{\omega}} \int_0^{\infty} exp\Big( - \frac{|\eta|}{\tau} u \Big) \ D_{+}\Big( \frac{u}{\sqrt{2\omega}} \Big) du \ d\omega$$
$$= \frac{2}{\pi^{\frac{3}{2}}\tau} \ \int_0^{\infty} \ exp\Big( - \frac{|\eta|}{\tau} u \Big) \ \Big( \int_{\omega} D_{+}\Big( \frac{u}{\sqrt{2\omega}} \Big) \ p(\omega) d\omega \Big) \ du$$
$$= \frac{2}{\pi^{\frac{3}{2}}\tau} \ \int_0^{\infty} \ exp\Big( - \frac{|\eta|}{\tau} u \Big) \ \phi(u) du$$

where $\phi(u) = \int_{\omega} D_{+}\Big( \frac{u}{\sqrt{2\omega}} \Big) \ p(\omega) d\omega = \mathbb{E}_{\omega}\Big[ D_{+}\Big( \frac{u}{\sqrt{2\omega}} \Big) \Big]$, $\forall u > 0$.

**Fast computation of $\lambda^{(k)}$:** $\lambda^{(k)}$ can be written as

$$\lambda^{(k)} = pen'_{HS}(|\eta^{(k)}|) = \frac{\int_0^{\infty} \ \frac{u}{\tau} \ \ exp\Big( -\frac{|\eta|}{\tau} u \Big) \ \ \phi(u) \ \ du}{\int_0^{\infty} \ \ exp\Big( -\frac{|\eta|}{\tau} u \Big) \ \ \phi(u) \ \ du}$$

The numerator and denominator in $\lambda^{(k)}$ can be approximated by Riemann sum using an approximation of Dawson function. Note that $\phi(u) = \mathbb{E}_{\omega}\Big[ D_{+}\Big( \frac{u}{\sqrt{2\omega}} \Big) \Big]$, $u > 0$ can be approximated by Monte-carlo integration for fixed positive $u$.