The Canadian Journal of Statistics
Vol. xx, No. yy, 20??, Pages 1–??
La revue canadienne de statistique

Supplement to "Bayesian inference for high-dimensional linear regression under the mnet priors"

Aixin Tan[1]* and Jian Huang[1]

[1] *Department of Statistics and Actuarial Science, University of Iowa*

1. THE MNET PENALTY FUNCTION AND ITS SPECIAL CASES

We mentioned five penalty functions in Table 1 of the main text. They are the ridge, the lasso, the enet, the mcp and the mnet penalty functions. The first four are indeed special cases of the mnet penalty function, and Figure 1 provides one way to display their relationship.

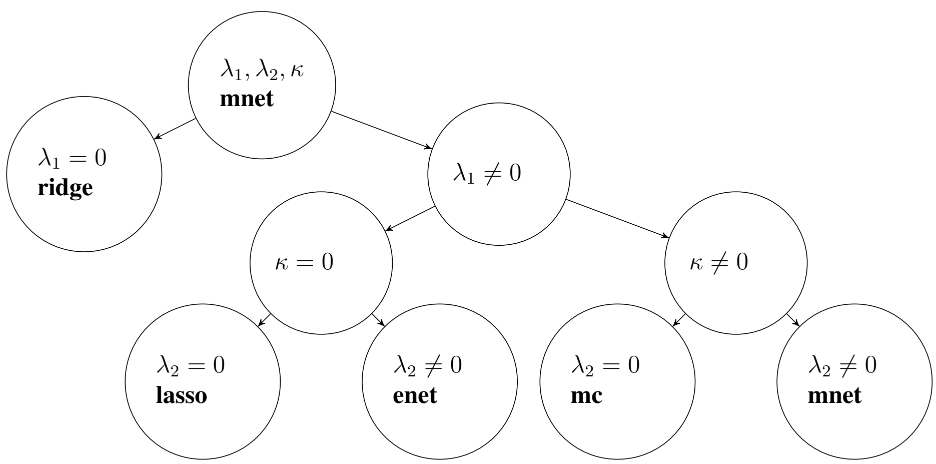

FIGURE 1: Relationship between the five penalty functions and the corresponding priors.

2. CORRESPONDENCE BETWEEN THE BAYESIAN MNET PRIOR AND ITS PENALIZED REGRESSION COUNTERPART

Besides the formal correspondence between the prior $f_{\boldsymbol{\lambda},\sigma^2}(\beta_j)$ and the penalty function $p(\cdot;\boldsymbol{\lambda})$,

$$f_{\boldsymbol{\lambda},\sigma^2}(\beta_j) = \sigma^{-1} c_{\boldsymbol{\lambda}} \exp\left\{ -p\left(\frac{\beta_j}{\sigma};\boldsymbol{\lambda}\right)\right\}, \tag{1}$$

---

* *Author to whom correspondence may be addressed.*
E-mail: aixin-tan@uiowa.edu

the Bayesian posterior and the penalized regression (PR) solution are related in the following way. When conditional on a fixed $(\sigma^2, \boldsymbol{\lambda})$, the posterior distribution of $\boldsymbol{\beta}$ is given by

$$\begin{aligned}\pi(\boldsymbol{\beta}|Y,\sigma^2,\boldsymbol{\lambda}) &\propto \exp\left\{-\frac{\|Y-X\boldsymbol{\beta}\|^2}{2\sigma^2} - p_{\text{mn}}\left(\frac{\boldsymbol{\beta}}{\sigma};\lambda_1,\lambda_2,\kappa\right)\right\} \\ &= \exp\left\{-\frac{n}{\sigma^2}\left[\frac{\|Y-X\boldsymbol{\beta}\|^2}{2n} + \frac{\sigma^2}{n}p_{\text{mn}}\left(\frac{\boldsymbol{\beta}}{\sigma};\lambda_1,\lambda_2,\kappa\right)\right]\right\} \\ &= \exp\left\{-\frac{n}{\sigma^2}\left[\frac{\|Y-X\boldsymbol{\beta}\|^2}{2n} + p_{\text{mn}}\left(\boldsymbol{\beta};\frac{\lambda_1\sigma}{n},\frac{\lambda_2}{n},\frac{\kappa}{n}\right)\right]\right\},\end{aligned}$$

where the last equality follows from the fact that, for any $a, b > 0$,

$$a\,p_{\text{mc}}\left(\frac{t}{b};\lambda_1,\kappa\right) = p_{\text{mc}}\left(t;\frac{a\lambda_1}{b},\frac{a\kappa}{b^2}\right) \quad \text{and} \quad a\,\frac{\lambda_2}{2}\left(\frac{t}{b}\right)^2 = \frac{a\lambda_2}{2b^2}t^2\,.$$

Hence, the mode of the above posterior density coincides with the solution to the mnet PR model with penalty parameter

$$(\lambda_1^*,\lambda_2^*,\kappa^*) = \left(\frac{\lambda_1\sigma}{n},\frac{\lambda_2}{n},\frac{\kappa}{n}\right)\,. \tag{2}$$

Recall that the mnet penalty function $p_{\text{mn}}(\cdot\,;\boldsymbol{\lambda})$ reduces to the normal, the lasso, and the enet penalty functions under special choices of $\boldsymbol{\lambda} = (\lambda_1,\lambda_2,\kappa)$. Naturally, we will refer to the corresponding priors in (1) in these special cases as the normal, the lasso, and the enet priors, respectively. Comparing to existing literature, the way in which σ^2 enters our definition of the lasso prior agrees with that of the double exponential prior of Park & Casella (2008). And our definition of the enet prior and the mnet prior can be considered a natural extension to this lasso prior. But our definition of the enet prior is different from that of Hans (2011) and Li & Lin (2010). Specifically, the conditional prior for β_j in their Bayesian enet model is $f_{\text{en}}(\beta_j|\lambda_1,\lambda_2,\sigma^2) \propto \exp\{-\frac{1}{2\sigma^2}(\lambda_1|\beta_j| + \lambda_2\beta_j^2)\}$, hence the mode of its conditional posterior density $\pi_{\text{en}}(\beta|Y,\sigma^2,\boldsymbol{\lambda})$ is the naive enet solution with penalty parameter $(\lambda_1^*,\lambda_2^*) = (\lambda_1/2n, \lambda_2/n)$, which unlike (2), is free of σ. Hence, their Bayesian enet model can not be re-parameterized to match ours. It is beyond the scope of this paper to study how the two versions of the Bayesian enet models compare, as we will focus on models allowed within the Bmnet framework defined in sec. 3 of the paper.

3. QUANTITIES NEEDED IN THE BLOCK GIBBS SAMPLER

Recall in sec. 4 of the main text, integrals I_1, I_2 and I_3 are needed in updating (β_j,γ_j) for $j = 1,\ldots,p$. We have

$$\begin{aligned}I_1 &= \int_0^{\frac{\lambda_1\sigma}{\kappa}} \exp\left\{-\frac{nt^2}{2\sigma^2} + \frac{X_j^T r_j t}{\sigma^2} - \frac{\lambda_1 t}{\sigma} - \frac{\lambda_2-\kappa}{2\sigma^2}t^2\right\}dt \\ &= \int_0^{\frac{\lambda_1\sigma}{\kappa}} \exp\left\{-\frac{A_1}{2}t^2 + B_1 t\right\}dt\,,\end{aligned}$$

where

$$A_1 = \frac{n+\lambda_2-\kappa}{\sigma^2} \quad \text{and} \quad B_1 = \frac{X_j^T r_j}{\sigma^2} - \frac{\lambda_1}{\sigma}.$$

Note that $A_1 > 0$ because κ/n corresponds to the κ parameter in the classical mcp or mnet penalty function, and is always set below 1. Therefore, I_1 can be expressed in terms of the pdf and the cdf of certain normal distributions:

$$I_1 = \exp\left\{\frac{B_1^2}{2A_1}\right\}\int_0^{\frac{\lambda_1\sigma}{\kappa}} \exp\left\{-\frac{A_1}{2}\left(t-\frac{B_1}{A_1}\right)^2\right\}dt = \Phi\left(\left(0,\frac{\lambda_1\sigma}{\kappa}\right);\frac{B_1}{A_1},\frac{1}{A_1}\right)\Big/\phi\left(0;\frac{B_1}{A_1},\frac{1}{A_1}\right),$$

where $\Phi((a,b);\mu,v)$ represents the probability that a Normal random variable with mean μ and variance v falls in the interval (a,b). Similarly, we have

$$\begin{aligned}
I_2 &= \int_{-\frac{\lambda_1\sigma}{\kappa}}^{0} \exp\left\{-\frac{nt^2}{2\sigma^2}+\frac{X_j^T r_j t}{\sigma^2}+\frac{\lambda_1 t}{\sigma}-\frac{\lambda_2-\kappa}{2\sigma^2}t^2\right\}dt \\
&= \int_{-\frac{\lambda_1\sigma}{\kappa}}^{0} \exp\left\{-\frac{A_2}{2}t^2+B_2 t\right\}dt \\
&= \Phi\left(\left(-\frac{\lambda_1\sigma}{\kappa},0\right);\frac{B_2}{A_2},\frac{1}{A_2}\right)\Big/\phi\left(0;\frac{B_2}{A_2},\frac{1}{A_2}\right),
\end{aligned}$$

where

$$A_2 = A_1 = \frac{n+\lambda_2-\kappa}{\sigma^2} \text{ and } B_2 = \frac{X_j^T r_j}{\sigma^2}+\frac{\lambda_1}{\sigma}.$$

Finally, we have

$$\begin{aligned}
I_3 &= \int_{-\infty}^{-\frac{\lambda_1\sigma}{\kappa}} + \int_{\frac{\lambda_1\sigma}{\kappa}}^{\infty} \exp\left\{-\frac{nt^2}{2\sigma^2}+\frac{X_j^T r_j t}{\sigma^2}-\frac{\lambda_1^2}{2\kappa}-\frac{\lambda_2}{2\sigma^2}t^2\right\}dt \\
&= \exp\left\{-\frac{\lambda_1^2}{2\kappa}\right\}\int_{-\infty}^{-\frac{\lambda_1\sigma}{\kappa}} + \int_{\frac{\lambda_1\sigma}{\kappa}}^{\infty} \exp\left\{-\frac{n+\lambda_2}{2\sigma^2}t^2+\frac{X_j^T r_j}{\sigma^2}t\right\}dt \\
&= \exp\left\{-\frac{\lambda_1^2}{2\kappa}\right\}\int_{-\infty}^{-\frac{\lambda_1\sigma}{\kappa}} + \int_{\frac{\lambda_1\sigma}{\kappa}}^{\infty} \exp\left\{-\frac{A_3}{2}t^2+B_3 t\right\}dt \\
&= \left[\Phi\left(-\frac{\lambda_1\sigma}{\kappa};\frac{B_3}{A_3},\frac{1}{A_3}\right)+1-\Phi\left(\frac{\lambda_1\sigma}{\kappa};\frac{B_3}{A_3},\frac{1}{A_3}\right)\right]\Big/\phi\left(0;\frac{B_3}{A_3},\frac{1}{A_3}\right),
\end{aligned}$$

where

$$A_3 = \frac{n+\lambda_2}{\sigma^2} \text{ and } B_3 = \frac{X_j^T r_j}{\sigma^2}.$$

Further, the conditional distribution of β_j given the others can be expressed as a mixture of three truncated normal distributions,

$$\begin{aligned}
&\pi(\beta_j|\boldsymbol{\gamma},\beta_{(j)},Y) = I_1\cdot\mathrm{TN}\left(\frac{B_1}{A_1},\frac{1}{A_1};\left(0,\frac{\lambda_1\sigma}{\kappa}\right)\right)+ \\
&I_2\cdot\mathrm{TN}\left(\frac{B_2}{A_2},\frac{1}{A_2};\left(-\frac{\lambda_1\sigma}{\kappa},0\right)\right)+I_3\cdot\mathrm{TN}\left(\frac{B_3}{A_3},\frac{1}{A_3};\left(-\infty,-\frac{\lambda_1\sigma}{\kappa}\right)\cup\left(\frac{\lambda_1}{\kappa},\infty\right)\right),
\end{aligned}$$

where TN $(\mu, v; (a, b))$ stands for the truncated normal density with mean μ, standard deviation v, and support $[a, b]$.

4. SUMMARY STATISTICS FOR THE SIMULATION STUDIES IN SEC. 5.1.

This section complements the graphical comparisons of different methods in sec. 5.1 with numerical ones. The tables below provide the median of the prediction mean squared error (pmse), the false discoveries (FD), the false negatives (FN), and the number of variables selected (NVS) of several methods, under different combinations of correlation strength ρ and signal to noise ratio s. Recall from sec. 5 that, at each (ρ, s), 200 datasets are randomly generated, each contains $n = 100$ observations and $q = 150$ predictors, while the number of true predictors is 10. The bootstrap standard error is reported for the median pmse. Specifically, we draw $B = 1000$ samples with replacement, each of size 200, from the 200 pmse values obtained in the simulation study. We calculate their sample medians, $(m_1, \cdots, m_B)$, and report their standard deviation as an approximation to the standard error of the median pmse.

TABLE 1: Under two simulation setups of $\rho = 0.3$ and $\rho = 0.9$, with signal to noise ratio fixed at $s = 2$, this table provides median of the prediction mean squared error (pmse), the false discoveries (FD), the false negatives (FN), and the number of variables selected (NVS) of several methods. For each setup, 200 datasets were randomly generated, each contains $n = 100$ observations and $q = 150$ potential predictors, while the number of true predictors is 10. Means of the 200 repetitions yield similar results, and hence are not shown.

	cvnorm	cvlasso	cvenet	cvenet.n	cvmcp	cvmnet	benet	bmnet-fx	bmnet-rd	nlp
$\rho = 0.3$										
pmse	6.23	**4.10**	4.30	4.38	5.15	4.59	4.46	4.69	**4.23**	5.65
std. err.	0.05	0.15	0.12	0.10	0.25	0.11	0.07	0.12	0.09	0.27
FD	140.00	20.00	25.00	30.00	10.00	22.00	7.00	3.00	3.00	2.00
FP	0.00	1.00	1.00	1.00	2.00	1.00	3.00	4.50	4.00	6.00
NVS	150.00	29.00	34.00	39.50	18.00	30.50	14.50	8.00	9.00	6.00
$\rho = 0.9$										
pmse	**1.25**	2.36	2.17	1.74	5.84	2.91	1.57	1.79	**1.48**	2.48
std. err.	0.07	0.06	0.09	0.07	0.09	0.25	0.04	0.04	0.03	0.10
FD	140.00	12.50	22.50	52.50	2.00	20.00	1.00	1.00	1.00	2.00
FP	0.00	8.00	7.00	4.00	9.00	7.00	10.00	10.00	10.00	10.00
NVS	150.00	15.00	25.00	57.50	2.00	22.50	2.00	1.00	1.00	2.00

TABLE 2: Medians over 200 replications of various statistics for different methods at $\rho = .3$ and $.9$. Signal to noise ratio is fixed at $s = 4$.

	cvnorm	cvlasso	cvenet	cvenet.n	cvmcp	cvmnet	benet	bmnet-fx	bmnet-rd	nlp
$\rho = 0.3$										
pmse	6.15	1.10	1.15	1.20	0.58	**0.47**	0.65	0.59	0.59	**0.41**
std. err.	0.03	0.04	0.03	0.04	0.07	0.03	0.02	0.02	0.02	0.02
FD	140.00	23.00	25.00	25.00	3.00	3.00	1.00	1.00	1.00	0.00
FP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NVS	150.00	33.00	35.00	35.00	13.00	13.00	11.00	11.00	11.00	10.00
$\rho = 0.9$										
pmse	1.29	1.24	1.09	1.01	3.05	1.22	**0.99**	1.09	**0.97**	1.81
std. err.	0.02	0.02	0.03	0.03	0.05	0.06	0.02	0.02	0.02	0.06
FD	140.00	18.00	25.00	52.00	3.00	25.50	9.50	2.00	4.00	2.00
FP	0.00	5.00	3.50	2.00	8.00	4.00	6.00	8.00	8.00	9.00
NVS	150.00	22.50	30.50	59.50	4.00	31.00	13.00	4.00	6.00	4.00

TABLE 3: Medians over 200 replications of various statistics for different methods at $\rho = .3$ and $.9$. Signal to noise ratio is fixed at $s = 8$.

	cvnorm	cvlasso	cvenet	cvenet.n	cvmcp	cvmnet	benet	bmnet-fx	bmnet-rd	nlp
$\rho = 0.3$										
pmse	6.14	0.29	0.29	0.32	**0.07**	**0.07**	**0.09**	**0.08**	**0.08**	**0.08**
std. err.	0.02	0.01	0.02	0.01	0.00	0.01	0.00	0.00	0.00	0.00
FD	140.00	22.00	24.00	24.50	0.00	0.00	0.00	0.00	0.00	0.00
FN	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NVS	150.00	32.00	34.00	34.50	10.00	10.00	10.00	10.00	10.00	10.00
$\rho = 0.9$										
pmse	1.30	**0.55**	**0.57**	0.61	1.31	0.72	0.61	0.59	0.60	0.74
std. err.	0.02	0.01	0.01	0.01	0.04	0.02	0.01	0.01	0.01	0.03
FD	140.00	20.00	24.00	33.00	4.00	14.00	21.00	5.00	8.00	2.00
FN	0.00	1.00	1.00	1.00	5.00	2.00	1.00	3.00	3.00	5.00
NVS	150.00	29.00	33.00	42.00	8.00	22.00	30.00	12.00	16.00	7.00

BIBLIOGRAPHY

Hans, C. (2011). Elastic net regression modeling with the orthant normal prior. *Journal of the American Statistical Association* 106 1383–1393.

Li, Q. & Lin, N. (2010). The Bayesian elastic net. *Bayesian Analysis* 5 151–170.

Park, T. & Casella, G. (2008). The Bayesian LASSO. *Journal of the American Statistical Association* 103 681–686.